Journal of Experimental and Theoretical Physics, Vol. 92, No. 4, 2001, pp. 543–547.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 4, 2001, pp. 627–632.

GRAVITATION, ASTROPHYSICS

The Class of Exact Cosmological Models with a Scalar-Field Nonlinearity Induced by the Yang–Mills Field

V. K. Shchigolev* and M. V. Shchigolev
Ul'yanovsk State University, Ul'yanovsk, 432700 Russia
**e-mail: shchigol@sv.uven.ru*
Received October 10, 2000

Abstract—A new class of exact solutions of the system of Einstein–Yang–Mills equations and the equations of a nonlinear scalar field interacting with the Yang–Mills *SO*(3) field is obtained on the basis of the Friedman spatially homogeneous cosmological models. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The solutions of field equations for self-gravitating and self-acting fields play a special role in the physics of the young Universe in general and, in particular, in inflation problems. The cosmological models of inflation can be constructed on the basis of solving a self-consistent system of Einstein's equation and the equations for nonlinear (in the simplest version, scalar) fields. The nonlinearity of the scalar inflanton field is manifested in its self-action through a nonquadratic potential assuming various modifications in the hierarchy of inflation models [1]. It should be noted that the examples of exact solutions of the self-consistent system of Einstein's equations and the equations of a nonlinear scalar field are scarce in view of the complexity of the nonlinear system. In this respect, the method of fine tuning of the potential is quite productive in our opinion. This method, which was widely used in [2], makes it possible to reconstruct the field dependence of the potential and to analyze numerically and graphically the time dependence of the field and the potential in some cases.

On the other hand, the application of the nonlinear scalar field induced through the interaction with other fields in cosmology appears to us as a quite realistic approach to describing the generation of nonlinearity in the classical field theory. Unfortunately, in this case, one must solve a complex nonlinear system of equations not only for a scalar field, but also for the field inducing the nonlinearity. Exact solutions are difficult to obtain even for the simplest field configurations. The situation becomes even more complicated when self-gravitating fields are considered in cosmological or astrophysical applications of the solutions. Interesting solutions can be obtained under the assumption of the background nature of the cosmological gravitational field as, for example, was done in [3, 4]. However, such solutions are not directly connected with the problems of the cosmological scenario if only due to their stochastic nature. Nevertheless, the very concept of induced nonlinearity is very attractive for realization precizely in cosmology in general and, probably, in the cosmology of the young Universe especially. The exact solutions for self-gravitating Yang–Mills (YM) fields in cosmology were obtained in a number of earlier publications (see, for example, [5]). However, the application of these solutions for the induction of the scalar field nonlinearity in cosmology would lead to nonhomogeneous configurations of the scalar field and, accordingly, to nonhomogeneous cosmological models. At the same time, the nontrivial topology of the YM *SO*(3) fields opens new prospects in the search for exact solutions of the self-consistent system of the Einstein–Yang–Mills equations and the equations of a nonlinear scalar field [6]. In the present paper, such a system is solved in the Friedman model and the generalized Wu–Yang ansatz for the YM fields.

2. EQUATIONS IN THE MODEL OF GRAVITATING SCALAR AND GAUGE FIELDS

We will proceed from the Lagrangian of a self-gravitating scalar field φ whose nonlinearity is induced by the YM *SO*(3)-symmetry field [3]:

$$L = \frac{R}{2\kappa_0} + \frac{1}{2}\varphi_{,\alpha}\varphi^{,\alpha} - \frac{1}{16\pi}F^a_{\alpha\beta}F^{a\alpha\beta}\Psi(\varphi), \qquad (1)$$

where $F^a_{ij} = \partial_i W^a_j - \partial_j W^a_i + e\varepsilon_{abc}W^b_i W^c_j$ is the tensor of the YM field W^a_i and $\Psi(\varphi)$ is the function describing the interaction between the scalar field and the YM field. It should be noted that Rybakov *et al.* [4] used a similar Lagrangian for describing the interaction with an electromagnetic field and for an analysis of static configurations against the background of the Friedman–Robertson–Walker static model. In the same publication, a similar Lagrangian is substantiated by the existence of the decay $\pi \longrightarrow 2\gamma$, while Piccinelli *et al.* [7]

noted that for $\Psi(\varphi) = e^{-2\lambda\varphi}$, Lagrangian (1) contains the compacted Kaluza–Klein theory for $\lambda = \sqrt{3}$ and the theory of superstrings for $\lambda = 1$. The same type of interaction appears in the Brans–Dicke theory.

The equations for YM fields, the scalar field, and gravitation potentials can be obtained by varying the Lagrangian in the fields and the space–time metric. As a result, we obtain the system of self-consistent Einstein equations

$$G_\mu^\nu = \kappa_0 T_\mu^\nu, \tag{2}$$

where the energy–momentum tensor of the fields is given by

$$\begin{aligned} T_\mu^\nu &= \varphi_{,\mu}\varphi^{,\nu} - \frac{1}{4\pi}F^a_{\mu\beta}F^{a\nu\beta}\Psi(\varphi) \\ &- \delta_\mu^\nu\left[\frac{1}{2}\varphi_{,\alpha}\varphi^{,\alpha} - \frac{1}{16}F^a_{\alpha\beta}F^{a\alpha\beta}\Psi(\varphi)\right], \end{aligned} \tag{3}$$

the YM equation

$$D_\nu(\sqrt{-g}F^{a\nu\mu}\Psi(\varphi)) = 0, \tag{4}$$

and the equation for the scalar field

$$\frac{1}{\sqrt{-g}}\frac{\partial}{\partial x^\nu}\left(\sqrt{-g}g^{\nu\mu}\frac{\partial\varphi}{\partial x^\mu}\right) + \frac{1}{16\pi}F^a_{\alpha\beta}F^{a\alpha\beta}\Psi_\varphi = 0. \tag{5}$$

Symbol D_ν denotes the covariant derivative.

We assume that the space–time interval possesses spherical symmetry and present it in the form

$$\begin{aligned} ds^2 &= dt^2 - U(r,t)dr^2 - V(r,t)d\Omega^2, \\ d\Omega^2 &= d\theta^2 + \sin^2\theta d\phi^2. \end{aligned} \tag{6}$$

At the same time, the generalized Wu–Yang ansatz [8] for the YM fields can be written as

$$\begin{aligned} W_i^a &= \varepsilon_{iab}x^b\frac{K(r,t)-1}{er^2} + \left(\delta_i^a - \frac{x^a x_i}{r^2}\right)\frac{S(r,t)}{er}, \\ W_0^a &= x^a\frac{W(r,t)}{er}. \end{aligned} \tag{7}$$

Here and below, $i, j, k, \ldots = 1, 2, 3$ are the spatial indices; $a, b, \ldots = 1, 2, 3$ are isotopic indices; and K, S, and W are the field functions of the above arguments. Introducing the orthonormal reference frame [8]

$$\begin{aligned} \mathbf{n} &= (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta), \\ \mathbf{l} &= (\cos\theta\cos\varphi, \cos\theta\sin\varphi, -\sin\theta), \\ \mathbf{m} &= (-\sin\varphi, \cos\varphi, 0) \end{aligned}$$

and going over to spherical coordinates, we can write ansatz (7) in the form

$$\begin{aligned} \mathbf{W}_1 &= 0, \\ \mathbf{W}_2 &= e^{-1}\{(K-1)\mathbf{m} + S\mathbf{l}\}, \\ \mathbf{W}_3 &= e^{-1}\{-(K-1)\mathbf{l} + S\mathbf{m}\}\sin\theta, \\ \mathbf{W}_0 &= e^{-1}W\mathbf{n}. \end{aligned} \tag{8}$$

As a result of simple calculation, we obtain the following components for the YM tensor:

$$\begin{aligned} \mathbf{F}_{01} &= -\mathbf{F}_{10} = -e^{-1}W'\mathbf{n}, \\ \mathbf{F}_{02} &= -\mathbf{F}_{20} = e^{-1}((\dot{K} + WS)\mathbf{m} + (\dot{S} - WK)\mathbf{l}), \\ \mathbf{F}_{03} &= -\mathbf{F}_{30} = e^{-1}(-(\dot{K} + WS)\mathbf{l} + (\dot{S} - WK)\mathbf{m})\sin\theta, \\ \mathbf{F}_{12} &= -\mathbf{F}_{21} = e^{-1}(K'\mathbf{m} + S'\mathbf{l}), \\ \mathbf{F}_{23} &= -\mathbf{F}_{32} = e^{-1}\sin\theta(K^2 - 1 + S^2)\mathbf{n}, \\ \mathbf{F}_{13} &= -\mathbf{F}_{31} = e^{-1}\sin\theta(-K'\mathbf{l} + S'\mathbf{m}). \end{aligned} \tag{9}$$

Here and below, the following notation is introduced: $(\dot{\ }) \equiv \partial/\partial t$ and $(\,') \equiv \partial/\partial r$.

The Einstein equations (2) for the energy–momentum tensor (3) taking into account the ansatz (9) of the YM fields can be written in the form

$$\begin{aligned} G_0^0 &= \kappa_0\left\{\frac{\dot{\varphi}^2}{2} + \frac{\varphi'^2}{2}\right. \\ &+ \frac{1}{8\pi e^2}\left[\frac{W'^2}{U} + \frac{2[(\dot{K} + WS)^2 + (\dot{S} - WK)^2]}{V}\right. \\ &\left.\left. + \frac{2[K'^2 + S'^2]}{UV} + \frac{(K^2 - 1 + S^2)^2}{V^2}\right]\Psi\right\}, \\ G_1^1 &= \kappa_0\left\{-\frac{\dot{\varphi}^2}{2} - \frac{\varphi'^2}{2}\right. \\ &+ \frac{1}{8\pi e^2}\left[\frac{W'^2}{U} - \frac{2[(\dot{K} + WS)^2 + (\dot{S} - WK)^2]}{V}\right. \\ &\left.\left. - \frac{2[K'^2 + S'^2]}{UV} + \frac{(K^2 - 1 + S^2)^2}{V^2}\right]\Psi\right\}, \end{aligned} \tag{10}$$

$$G_2^2 = G_3^3 = \kappa_0\left\{-\frac{\dot\varphi^2}{2}+\frac{\varphi'^2}{2}\right.$$

$$\left.-\frac{1}{8\pi e^2}\left[\frac{W'^2}{U}+\frac{(K^2-1+S^2)^2}{V^2}\right]\Psi\right\},$$

$$G_1^0 = \kappa_0\left\{\dot\varphi\varphi'+\frac{1}{2\pi e^2}\right.$$

$$\left.\times\frac{K'(\dot K+WS)+S'(\dot S-WK)}{V}\Psi\right\},$$

where the expressions for the Einstein tensor components G_μ^ν in terms of the functions U and V of metric (6) are known. Using the components of tensor (9), we can write the equations for the YM fields (4) in the form

$$\frac{\partial}{\partial r}\left(\frac{VW'}{2\sqrt{U}}\Psi\right)$$

$$+\sqrt{U}((\dot S-WK)K-(\dot K+WS)S)\Psi = 0,$$

$$\frac{\partial}{\partial t}(\sqrt{U}(\dot S-WK)\Psi)-\frac{\partial}{\partial r}\left(\frac{S'}{\sqrt{U}}\Psi\right)$$

$$+\sqrt{U}\left(\frac{(K^2-1+S^2)S}{V}-(\dot K+WS)W\right)\Psi = 0, \qquad (11)$$

$$\frac{\partial}{\partial t}(\sqrt{U}(\dot K-WS)\Psi)-\frac{\partial}{\partial r}\left(\frac{K'}{\sqrt{U}}\Psi\right)$$

$$+\sqrt{U}\left(\frac{(K^2-1+S^2)K}{V}+(\dot S-WK)W\right)\Psi = 0,$$

while the equation for the scalar field can be written as

$$\frac{1}{V\sqrt{U}}\frac{\partial}{\partial t}(V\sqrt{U}\dot\varphi)-\frac{1}{V\sqrt{U}}\frac{\partial}{\partial r}\left(V\frac{\varphi'}{\sqrt{U}}\right)$$

$$-\frac{1}{8\pi e^2}\left[\frac{W'^2}{U}+\frac{2[(\dot K+WS)^2+(\dot S-WK)^2]}{V}\right. \qquad (12)$$

$$\left.-\frac{2[K'^2+S'^2]}{UV}-\frac{(K^2-1+S^2)^2}{V^2}\right]\Psi_\varphi = 0,$$

where $\Psi_\varphi = d\Psi/d\varphi$.

It should be noted that in the absence of the scalar field, the solution of the systems of Eqs. (10), (11) for a self-gravitating pure YM field was derived in [8] in the Tolman metric under the following simplifying assumptions concerning the field functions: $K = S = 0$, but $W \neq 0$. The latter inequality means that the YM field possesses an electric charge which is just responsible for the inhomogeneity of the cosmological model. If we proceed from the same assumptions as in [8], the complete self-consistent system of the Einstein–Yang–Mills equations and the equation of the nonlinear scalar field (12) cannot be solved. Another approach to the solution of these equations can be prompted by the results obtained by Rybakov *et al.* [4], who obtained the static solutions of similar field equations against the background of a static Universe. In the present work, we propose a new substitution for the YM field functions, which makes it possible to analyze the complete system of Eqs. (10)–(12) and to obtain exact solutions for this system.

3. CLASS OF MODELS WITH THE FRIEDMAN–ROBERTSON–WALKER METRIC

For a homogeneous and isotropic Universe, a linear Friedman–Robertson–Walker element can be presented in the form

$$ds^2 = dt^2 - a^2(t)(dr^2+\xi^2(r)d\Omega^2), \qquad (13)$$

where

$$\xi(r) = \begin{cases}\sin r, & k = +1,\\ r, & k = 0,\\ \sinh r, & k = -1,\end{cases}$$

and $k = 0, \pm1$ is the sign of the curvature of the three-dimensional hypersurface t = const. In other words, in metric (6), we must put $U = a^2(t)$ and $V = a^2(t)\xi^2(r)$. For the Friedman–Robertson–Walker interval (13), the nonzero components of the Einstein tensor are given by [9]

$$G_0^0 = \frac{3}{a^2}(\dot a^2+k),$$

$$G_1^1 = G_2^2 = G_3^3 = \frac{1}{a^2}(\dot a^2+2a\ddot a+k).$$

If we require that $W = 0$, $K = K(r)$, $S = S(r)$ and that the scalar field $\varphi = \varphi(t)$ is uniform, the system of Eqs. (11), (12) assumes the form

$$S''-\frac{(K^2-1+S^2)S}{\xi^2} = 0,$$

$$K''-\frac{(K^2-1+S^2)K}{\xi^2} = 0,$$

$$\frac{1}{a^3}\frac{\partial}{\partial t}(a^3\dot\varphi)+\frac{1}{8\pi e^2a^4} \qquad (14)$$

$$\times\left[\frac{2[(K'^2+S'^2)]}{\xi^2}+\frac{(K^2-1+S^2)^2}{\xi^4}\right]\Psi_\varphi = 0.$$

The first two equations have the same structure, which reflects the symmetry of the Lagrangian relative to transformations of functions K and S of the form

$$K = P(r)\cos\alpha, \quad S = P(r)\sin\alpha, \tag{15}$$

where α is an arbitrary constant. Consequently, we obtain the following equation for function $P(r)$:

$$P''(r) - \frac{(P^2(r)-1)P(r)}{\xi^2} = 0. \tag{16}$$

For $k = \pm 1$, i.e., for the closed and open models of the Universe, the solutions of Eq. (16) are the functions

$$P(r) = \xi'(r) = \begin{cases} \cos r, & k = +1 \\ \cosh r, & k = -1. \end{cases} \tag{17}$$

Using the obtained solutions of the YM equations, we can reduce the remaining independent Einstein equations to the following system:

$$\begin{aligned} \frac{3}{a^2}(\dot a^2 + k) &= \kappa_0\frac{\dot\varphi^2}{2} + \frac{3\kappa_0}{8\pi e^2}\frac{1}{a^4}\Psi(\varphi), \\ \frac{1}{a^2}(\dot a^2 + 2a\ddot a + k) &= -\kappa_0\frac{\dot\varphi^2}{2} - \frac{\kappa_0}{8\pi e^2}\frac{1}{a^4}\Psi(\varphi). \end{aligned} \tag{18}$$

The equation of the scalar field in this case has the form

$$\frac{1}{a^3}\frac{\partial}{\partial t}(a^3\dot\varphi) + \frac{3}{8\pi e^2}\frac{1}{a^4}\Psi_\varphi = 0. \tag{19}$$

It can easily be verified that the system of Eqs. (18) and (19) contains only two independent equations; consequently, in order to solve the system of equations, we must specify one of the functions $a(t)$, $\varphi(t)$, or $\Psi(\varphi)$. Defining the interaction function $\Psi(\varphi)$. It appears as the most natural, but not necessary, condition and depends on the specific problem to be solved.

It is worthwhile to note an interesting feature of the Einstein equations (18), which can be used to write the effective values of the energy density $\epsilon(t)$ and pressure $p(t)$ in the form

$$\begin{aligned} \epsilon(t) &= \frac{\dot\varphi^2}{2} + 3\frac{1}{8\pi e^2 a^4}\Psi(\varphi), \\ p(t) &= \frac{\dot\varphi^2}{2} + \frac{1}{8\pi e^2 a^4}\Psi(\varphi). \end{aligned}$$

It follows hence that $\epsilon - 3p = -\dot\varphi^2 \le 0$. The extremely stringent equation of state $\epsilon = p$ is naturally obtained when $\Psi = 0$. Thus, the effective pressure is determined by the inequality $\epsilon/3 \le p \le \epsilon$. By an appropriate selection of the interaction function $\Psi(\varphi)$, we can obtain the required equation of state from the above interval to solve a specific problem with a cosmological scenario. In our opinion, it is expedient to apply the model under investigation in the problems of quintessence in the way it was done in [10] on the basis of the possible magnetization of the Universe.

In order to obtain the exact solutions of system (18), (19), we can apply the method of fine tuning [11], in which the evolution of the scale factor $a(t)$ is specified. Using Eq. (18), we can find

$$\begin{aligned} \dot\varphi^2 &= -\frac{6}{\kappa_0}\left(\left(\frac{\dot a}{a}\right)^2 + \frac{\ddot a}{a} + \frac{k}{a^2}\right), \\ \Psi &= \frac{8\pi e^2}{\kappa_0}a^2(2\dot a^2 + a\ddot a + 2k). \end{aligned} \tag{20}$$

It follows hence that the solution for the real scalar field φ exists only if

$$\left(\frac{\dot a}{a}\right)^2 + \frac{\ddot a}{a} + \frac{k}{a^2} < 0. \tag{21}$$

This inequality indicates, among other things, that an accelerated expansion of the Universe with $\ddot a > 0$ is possible only in some cases, and for $k = +1$ in the given model, it is absent altogether. If we include an ideal liquid and the cosmological constant into the model, relation (21) and the conclusions concerning the possibility of the accelerated mode differ considerably from those presented above.

Let us consider a simple example of the solution for the nonlinear scalar field (20) by using the above method. Let us suppose that the scale factor $a(t) = a_0 t$, and the sign of curvature $k = -1$. Then the system of equations (20) leads to

$$\begin{aligned} \dot\varphi &= \pm\sqrt{\frac{6}{\kappa_0}(1-a_0^2)}\frac{1}{a_0 t}, \\ \Psi &= \frac{16\pi e^2}{\kappa_0}a_0^2(a_0^2-1)t^2. \end{aligned}$$

The former equations shows that the solution exists if $a_0 < 1$ and has the form

$$\varphi = \pm\sqrt{\frac{6}{\kappa_0}\left(\frac{1}{a_0^2} - 1\right)}\ln t + \varphi_0.$$

Eliminating time from the last two equations, we find the function of the interaction in the form

$$\Psi = \Psi_0 e^{\pm 2\lambda\varphi}, \tag{22}$$

where

$$\Psi_0 = \frac{16\pi e^2}{\kappa_0}a_0^2(a_0^2-1)e^{\mp 2\lambda\varphi_0},$$

and

$$\lambda = \sqrt{\frac{\kappa_0}{6}\frac{a_0^2}{(1-a_0^2)}}.$$

Let us consider another example, presuming the harmonic law of the evolution of the scale factor in the open ($k = -1$) model of the Universe:

$$a(t) = H_0^{-1}\sin(H_0 t).$$

The substitution of the scale factor into system (20) makes it possible to integrate the system, which leads to the following expressions:

$$\varphi = \pm\sqrt{\frac{12}{\kappa_0}}H_0 t + \varphi_0,$$

$$\Psi = -\frac{24\pi e^2}{\kappa_0 H_0^2}\sin^4(H_0 t).$$

Using these expressions, we can obtain the explicit form of the interaction function:

$$\Psi = \Psi_0\sin^4\left[\sqrt{\frac{\kappa_0}{12}}(\varphi - \varphi_0)\right], \qquad (23)$$

where $\Psi_0 = -24\pi e^2/\kappa_0 H_0^2$.

It should be noted that irrespective of the specific form of the interaction functions $\Psi(\varphi)$ and, hence, of the solutions of Eqs. (18), the nonzero components of the tensor for the YM fields, in accordance with formulas (9), (15), and (17), have the form

$$\mathbf{F}_{21} = -\mathbf{F}_{12} = ke^{-1}\xi(r)(\mathbf{m}\cos\alpha + \mathbf{l}\sin\alpha),$$

$$\mathbf{F}_{31} = -\mathbf{F}_{13} = ke^{-1}\xi(r)\sin\theta(\mathbf{m}\sin\alpha - \mathbf{l}\cos\alpha), \quad (24)$$

$$\mathbf{F}_{32} = -\mathbf{F}_{23} = ke^{-1}\xi^2(r)\mathbf{n}\sin\theta,$$

which means that the a YM field has only magnetic components. It is interesting to note that this circumstance has allowed us to obtain homogeneous solutions for a scalar field. Indeed, formulas (24) readily lead to the invariant of the YM field,

$$\mathbf{F}_{\mu\nu}\mathbf{F}^{\mu\nu} = 3e^{-2}a^{-4}(t),$$

which is just responsible for the dependence of the second term in the Eq. (18) for the scalar field, which is nonlinear in the field, on the time variable only.

4. CONCLUSION

Thus, it is found that the system of the Einstein–Yang–Mills equations and the equation of a nonlinear scalar field, which is obtained from Lagrangian (1), has solutions with a uniform scalar field interacting with a YM field of type (24) in the Friedman model of the Universe, limited only by condition (21). The system is reduced to two independent Eqs. (18) or (20), which should be additionally defined either by a specific interaction function or by indicating the required rate of the evolution of the scale factor. It is shown that the latter can be realized in all cases, and the simple examples (22) and (23) illustrate the possibility of determining the explicit dependence of the interaction function on the scalar field for certain modes of the expansion of the Universe.

ACKNOWLEDGMENTS

This work was financially supported by the Russian Foundation for Basic Research (project nos. 98-02-18040 and 00-01-00260).

REFERENCES

1. A. D. Linde, *Physics of Elementary Particles and Inflation Cosmology* (Nauka, Moscow, 1990).
2. S. V. Chervon, V. M. Zhuravlev, and V. K. Shchigolev, Phys. Lett. B **398**, 269 (1997).
3. K. A. Bronnikov, V. N. Melnikov, G. N. Shikin, and K. P. Stanukovich, Ann. Phys. **118**, 84 (1979).
4. Yu. P. Rybakov, B. Saha, and G. N. Shikin, gr-qc/96031031 (1996).
5. V. K. Shchigolev, S. V. Chervon, and O. V. Kudasova, Gen. Relativ. Gravit. **32** (1), 41 (2000).
6. V. K. Shchigolev and M. V. Shchigolev, in *Proceedings of the XII International School-Workshop on Theoretical and Mathematical Physics "Volga-12'2000"* (Kazan, 2000), p. 84.
7. G. Piccinelli, T. Matos, and M. Montesinos, gr-qc/9805032 (1998).
8. V. K. Shchigolev, V. M. Zhuravlev, and S. V. Chervon, Pis'ma Zh. Éksp. Teor. Fiz. **64**, 65 (1996) [JETP Lett. **64**, 71 (1996)].
9. S. Weinberg, *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity* (Wiley, New York, 1972; Mir, Moscow, 1975).
10. D. R. Matravers and C. G. Tsagas, astro-ph/0003181 (2000).
11. S. V. Chervon, *Nonlinear Fields in Gravitation Theory and Cosmology* (Izd. Ul'yanov. Gos. Univ., Ul'yanovsk, 1997).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 92, No. 4, 2001, pp. 548–551.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 4, 2001, pp. 633–637.

NUCLEI, PARTICLES, AND THEIR INTERACTION

Rare Kaon Decay $K^+ \longrightarrow \pi^+\nu\bar{\nu}$ in $SU(3)_C \otimes SU(3)_L \otimes U(1)_N$ Models¶

H. N. Long[a, b]*, L. P. Trung[c], and V. T. Van[b]

[a]*Asia Pacific Center for Theoretical Physics, Seoul 135-080, Korea*
[b]*Institute of Physics, P. O. Box 429, Bo Ho Hanoi 10000, Vietnam*
**e-mail: hnlong@iop.ncst.ac.vn*
[c]*HCM City Branch, Institute of Physics, Ho Chi Minh City, Vietnam*

Received November 1, 2000

Abstract—The rare kaon decay $K^+ \longrightarrow \pi^+\nu\bar{\nu}$ is considered in the framework of models based on the $SU(3)_C \otimes SU(3)_L \otimes U(1)_N$ (3–3–1) gauge group. In the 3–3–1 model with right-handed neutrinos, the lower bound of the Z' mass is derived at 3 TeV, and that in the minimal version, at 1.7 TeV. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Kaon is the lightest hadron having a nonzero strangeness quantum number. Due to the weak interactions, kaon decays into zero-strangeness states containing pions, photons, and/or leptons. The physics of kaons has played a major role in the development of particle physics. The concept of strangeness, with its implications for the quark model, the discovery of the *P* and *CP* violation, and the GIM mechanism have all emerged from the study of K mesons. Today, rare kaon decays remain a field of active investigations (see for example [1]). Flavor-changing neutral currents (FCNC) are completely suppressed at the tree level by the GIM mechanism in the standard model (SM). In the second or higher order interactions, this suppression is not complete because of different quark masses [2].

The first experimental evidence for atmospheric neutrino oscillations (and, consequently, a nonzero neutrino mass) observed at the SuperKamiokande Collaboration calls for an extension of the SM. Among the possible models, those based on the $SU(3)_C \otimes SU(3)_L \otimes U(1)_N$ (3–3–1) gauge group [3–6] contain a number of intriguing features. First, the models predict three families of quarks and leptons if the anomaly-free condition on $SU(3)_L \otimes U(1)_N$ and the QCD asymptotic freedom are imposed. Second, the Peccei–Quinn symmetry naturally occurs in these models [7]. The third interesting point is that one generation of quarks is treated differently from the other two. This could lead to a natural explanation for the unbalancingly heavy top quark. This family nonuniversality also leads to the FCNC by the Z' currents at the tree level [8, 9]. Finally, the 3–3–1 models predict new physics at a scale only slightly above the SM scale (several TeVs) [8–11].

In this work, we consider the implications of the main two 3–3–1 models for the rare $K^+ \longrightarrow \pi^+\nu\bar{\nu}$ decay; our aim is to obtain a bound on the Z' mass.

2. THE RARE KAON DECAY $K^+ \longrightarrow \pi^+\nu\bar{\nu}$ IN 3–3–1 MODELS

2.1. The Decay in the 3–3–1 Model with Right-Handed Neutrinos

We first recapitulate the basic elements of the model. The leptons in this model are arranged into triplets, with the third member being a right-handed neutrino [5, 6],

$$f_L^a = \begin{pmatrix} \nu_L^a \\ e_L^a \\ (\nu_L^c)^a \end{pmatrix} \sim (1, 3, -1/3), \quad e_R^a \sim (1, 1, -1), \quad (1)$$

where $a = 1, 2, 3$ is the family index.

The first two families of quarks are in antitriplets and the third one is in a triplet,

$$Q_{iL} = \begin{pmatrix} d_{iL} \\ -u_{iL} \\ D_{iL} \end{pmatrix} \sim (3, \bar{3}, 0), \quad (2)$$

$$u_{iR} \sim (3, 1, 2/3), \quad d_{iR} \sim (3, 1, -1/3),$$
$$D_{iR} \sim (3, 1, -1/3), \quad i = 1, 2,$$

$$Q_{3L} = \begin{pmatrix} u_{3L} \\ d_{3L} \\ T_L \end{pmatrix} \sim (3, 3, 1/3), \quad (3)$$

$$u_{3R} \sim (3, 1, 2/3), \quad d_{3R} \sim (3, 1, -1/3), \quad T_R \sim (3, 1, 2/3).$$

¶ This article was submitted by the authors in English.

The gauge bosons in this model are the photon (A), Z, Z', $W^{\pm}$, $Y^{\pm}$, and complex neutral bosons X^0 and X^{*0},

$$\sqrt{2}W_\mu^+ = W_\mu^1 - iW_\mu^2, \quad \sqrt{2}Y_\mu^- = W_\mu^6 - iW_\mu^7,$$
$$\sqrt{2}X_\mu^0 = W_\mu^4 - iW_\mu^5,$$
$$A_\mu = s_W W_\mu^3 + c_W\left(-\frac{t_W}{\sqrt{3}}W_\mu^8 + \sqrt{1-\frac{t_W^2}{3}}B_\mu\right), \tag{4}$$
$$Z_\mu = c_W W_\mu^3 - s_W\left(-\frac{t_W}{\sqrt{3}}W_\mu^8 + \sqrt{1-\frac{t_W^2}{3}}B_\mu\right),$$
$$Z'_\mu = \sqrt{1-\frac{t_W^2}{3}}W_\mu^8 + \frac{t_W}{\sqrt{3}}B_\mu,$$

where we use the notation $s_W \equiv \sin\theta_W$ and $t_W \equiv \tan\theta_W$. The physical states are a mixture of Z and Z',

$$Z_1 = Z\cos\phi - Z'\sin\phi,$$
$$Z_2 = Z\sin\phi + Z'\cos\phi,$$

where ϕ is the mixing angle.

The interactions between fermions and Z_1, Z_2 are given by

$$\mathcal{L}^{NC} = \frac{g}{2c_W}$$
$$\times\{\bar{f}\gamma^\mu[a_{1L}(f)(1-\gamma_5) + a_{1R}(f)(1+\gamma_5)]fZ_\mu^1 \tag{5}$$
$$+ \bar{f}\gamma^\mu[a_{2L}(f)(1-\gamma_5) + a_{2R}(f)(1+\gamma_5)]fZ_\mu^2\},$$

where

$$a_{1L,R}(f) = [T^3(f_{L,R}) - s_W^2 Q(f)]\cos\phi$$
$$-c_W^2\left[\frac{3N(f_{N,R})}{(3-4s_W^2)^{1/2}} - \frac{(3-4s_W^2)^{1/2}}{2c_W^2}Y(f_{L,R})\right]\sin\phi,$$
$$a_{2L,R}(f) \tag{6}$$
$$= c_W^2\left[\frac{3N(f_{L,R})}{(3-4s_W^2)^{1/2}} - \frac{(3-4s_W^2)^{1/2}}{2c_W^2}Y(f_{L,R})\right]\cos\phi$$
$$+ [T^3(f_{L,R}) - s_W^2 Q(f)]\sin\phi.$$

Here, $T^3(f)$ and $Q(f)$ are, respectively, the third component of the weak isospin and the charge of the fermion f. The mixing angle ϕ is constrained to be very small [6], $-2.8\times10^{-3} \le \phi \le 1.8\times10^{-4}$, and can therefore be neglected.

Because one family of left-handed quarks is treated differently from the other two, the N charges for left-handed quarks are also different (see Eq. (3)). Therefore, the FCNC Z' occurs through a mismatch between weak and mass eigenstates. We diagonalize the mass matrices by three biunitary transformations:

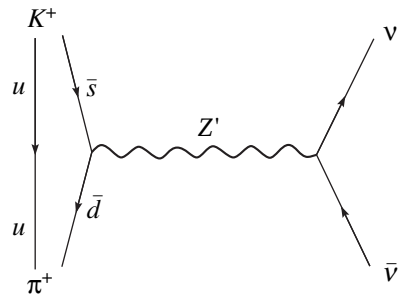


Fig. 1. Feynman diagram for $K^+ \longrightarrow \pi^+\nu\bar{\nu}$ in the 3–3–1 models.

$$U'_L = V_L^U U_L, \quad U'_R = V_R^U U_R,$$
$$D'_L = V_L^D D_L, \quad D'_R = V_R^D D_R, \tag{7}$$

where $U \equiv (u, c, t)^T$ and $D \equiv (d, s, b)^T$. The usual Cabibbo–Kobayashi–Maskawa matrix is given by

$$V_{CKM} = V_L^{U+}V_L^D. \tag{8}$$

Using unitarity of the V^D and V^U matrices, we obtain the flavor-changing neutral interactions [9]

$$\mathcal{L}_{ds}^{NC} = \frac{gc_W}{2\sqrt{3-4s_W^2}}[V_{Lid}^{D*}V_{Lis}^D]\bar{d}_L\gamma^\mu s_L Z'_\mu, \tag{9}$$

where i denotes the number of "different" quark families, i.e., the $SU(3)_L$ quark triplet. It was shown in [9] that i must be equal to 3; i.e., the third family of quarks must be different from the first two.

We consider the decay

$$K^+(p1) \longrightarrow \pi^+(p2)\nu(k1)\bar{\nu}(k2), \tag{10}$$

where the symbols in parentheses stand for the momenta of the particles. The one-loop effective SM Lagrangian for this process was calculated by Inami *et al.* [2]. Due to family nonuniversality in the 3–3–1 models, the decay can be mediated by Z' at the tree level. The Feynman diagram contributing to the above decay is depicted in Fig. 1.

The decay amplitude is given by

$$\mathcal{M}(K^+ \longrightarrow \pi^+\nu\bar{\nu}) = \frac{G_F}{\sqrt{2}}\frac{m_W^2}{M_{Z'}^2}V_{Lbd}^{D*}V_{Lbs}^D \tag{11}$$
$$\times\langle\pi^+(p2)|\bar{s}_L\gamma_\mu d_L|K^+(p1)\rangle\bar{\nu}_L(k1)\gamma^\mu\nu_L(k2),$$

where m_W and $M_{Z'}$ stand for the W and Z' boson masses, respectively.

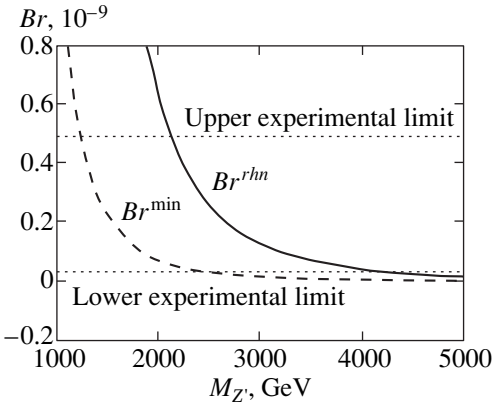


Fig. 2. Branching ratio (*Br*) as a function of $M_{Z'}$.

For our initial purpose, we present the well-measured semileptonic decay $K^+(p1) \longrightarrow \pi^0(p2)e^+(k1)\nu(k2)$. The tree-level amplitude for this process can be written as

$$\mathcal{M}(K^+ \longrightarrow \pi^0 e^+ \nu) = \frac{G_F}{\sqrt{2}} V_{us}^* \times \langle \pi^0(p2)|\bar{s}_L \gamma_\mu u_L|K^+(p1)\rangle \bar{\nu}_{eL}(k1)\gamma^\mu e_L(k2). \quad (12)$$

The isospin symmetry relates hadronic matrix elements in (11) to (12) to a very good precision [12],

$$\langle \pi^+(p2)|\bar{s}_L \gamma_\mu d_L|K^+(p1)\rangle = \sqrt{2}\langle \pi^0(p2)|\bar{s}_L \gamma_\mu u_L|K^+(p1)\rangle. \quad (13)$$

Neglecting differences in the phase space of two considered decays occurring because $m_{\pi^+} \neq m_{\pi^0}$ and $m_e \neq 0$, we sum over the three neutrino flavors and obtain

$$\frac{Br^{rhn}(K^+ \longrightarrow \pi^+ \nu\bar{\nu})}{Br(K^+ \longrightarrow \pi^0 e^+ \nu)} = 6\left(\frac{m_W^2}{M_{Z'}^2}\right)^2 \frac{\left|V_{Lbd}^{*D} V_{Lbs}^D\right|^2}{|V_{us}^*|^2}, \quad (14)$$

where the symbol *rhn* added to the branching ratio indicates the case under consideration. We now apply the simple Fritzsch [13] scheme as

$$V_{ij}^D \approx \left(\frac{m_i}{m_j}\right)^{1/2}, \quad i < j. \quad (15)$$

Inserting (15) in to (14), we obtain

$$Br^{rhn}(K^+ \longrightarrow \pi^+ \nu\bar{\nu}) = 6\left(\frac{m_W^2}{M_{Z'}^2}\right)^2 \frac{m_d m_s}{m_b^2 V_{us}^2} Br(K^+ \longrightarrow \pi^0 e^+ \nu). \quad (16)$$

In Fig. 2, we plot Br^{rhn} as a function of $M_{Z'}$, using the data [14]

$$m_W = 80.41 \text{ GeV}, \quad |V_{us}| = 0.2196, \quad m_d = 7 \text{ MeV}, \quad m_s = 115 \text{ MeV}, \quad m_b = 4.3 \text{ GeV}, \quad Br(K^+ \longrightarrow \pi^0 e^+ \nu) = 4.42 \times 10^{-2}. \quad (17)$$

The horizontal lines are the upper (4.9×10^{-10}) and the lower (0.3×10^{-10}) experimental data [15].

From this figure, we see that the lower bound on the Z' mass is in the range from 2.3 to 4.35 TeV. This bound is approximately twice as large as that derived from the mass difference of the kaon mixing system Δm_K [9]. We thus arrive at the previous conclusion again: for the Z' mass to be relatively low, the third family of quarks must be different from the other two.

2.2. The Decay in the Minimal 3–3–1 Model

This model treats the leptons as $SU(3)_L$ antitriplets [4, 10], with the third element being the antilepton (the name of this version comes from the fact that no new leptons are introduced):

$$f_L^a = \begin{pmatrix} e_L^a \\ -\nu_L^a \\ (e^c)^a \end{pmatrix} \sim (1, \bar{3}, 0). \quad (18)$$

Of the nine gauge bosons W^a ($a = 1, 2, \ldots, 8$) and B of $SU(3)_L$ and $U(1)_N$, four are light: the photon (A), Z, and $W^\pm$. The remaining five correspond to new heavy gauge bosons Z' and $Y^\pm$ and the doubly charged bileptons $X^{\pm\pm}$. They are expressed in terms of W^a and B as [10]

$$\sqrt{2}W_\mu^+ = W_\mu^1 - iW_\mu^2, \quad \sqrt{2}Y_\mu^+ = W_\mu^6 - iW_\mu^7, \quad \sqrt{2}X_\mu^{++} = W_\mu^4 - iW_\mu^5;$$

$$A_\mu = s_W W_\mu^3 + c_W(\sqrt{3}t_W W_\mu^8 + \sqrt{1-3t_W^2}B_\mu), \quad (19)$$

$$Z_\mu = c_W W_\mu^3 - s_W(\sqrt{3}t_W W_\mu^8 + \sqrt{1-3t_W^2}B_\mu),$$

$$Z'_\mu = -\sqrt{1-3t_W^2}W_\mu^8 + \sqrt{3}t_W B_\mu.$$

As before, the physical states are a mixture of Z and Z',

$$Z_1 = Z\cos\phi - Z'\sin\phi, \quad Z_2 = Z\sin\phi + Z'\cos\phi,$$

and the mixing angle ϕ is also constrained to be very small. We can therefore assume $\phi \approx 0$. Applying Eq. (4.4) in [10], we obtain the interactions among Z' and neutrinos,

$$a'_V(\nu) = -a'_A(\nu) = \frac{1}{2\sqrt{3}}\sqrt{1-4s_W^2}. \quad (20)$$

One necessary vertex, namely, the FCNC, is given in [8],

$$\mathcal{L}_{ds}^{NC} = \frac{gc_W}{2\sqrt{3(1-4s_W^2)}}[V_{Lid}^{D*}V_{Lis}^{D}]\bar{d}_L\gamma^\mu s_L Z'_\mu. \quad (21)$$

Combining Eqs. (20) and (21), we obtain the decay amplitude

$$\mathcal{M}^{\min}(K^+ \longrightarrow \pi^+\nu\bar{\nu}) = \frac{1}{3}\frac{G_F}{\sqrt{2}}\frac{m_W^2}{M_{Z'}^2}V_{Lbd}^{D*}V_{Lbs}^{D} \quad (22)$$
$$\times \langle \pi^+(p2)|\bar{s}_L\gamma_\mu d_L|K^+(p1)\rangle \bar{\nu}_L(k1)\gamma^\mu\nu_L(k2).$$

From Eq. (22), it is straightforward to obtain

$$\frac{Br^{\min}(K^+ \longrightarrow \pi^+\nu\bar{\nu})}{Br(K^+ \longrightarrow \pi^0 e^+\nu)} = \frac{2}{3}\left(\frac{m_W^2}{M_{Z'}^2}\right)^2\frac{\left|V_{Lbd}^{*D}V_{Lbs}^{D}\right|^2}{\left|V_{us}^*\right|^2}. \quad (23)$$

As in the previous section, we plot $Br^{\min}$ as a function of $M_{Z'}$ in Fig. 2. As a consequence, the lower bound on the Z' mass is in the range from 1.25 to 2.45 TeV. This bound is larger than the one derived from the mass difference of the kaon mixing system Δm_K (see Dumm *et al.* [8]). For the Z' mass to be relatively low, the third family of quarks must be different from the other two. It is worth mentioning that the branching ratio is not sensitive to the value for $\sin^2\theta_W$, while the expression for Δm_K in the minimal version is very sensitive due to the factor $(1-4s_W^2)^{-1}$.

3. CONCLUSIONS

We have considered the rare kaon decay $K^+ \longrightarrow \pi^+\nu\bar{\nu}$ in the 3–3–1 models at the tree level. It was shown that in the model involving right-handed neutrinos, the decay width is by about one order larger than in the minimal version. As a result, we obtained bounds on the Z' mass in the range from 2.3 to 4.3 TeV in the model with right-handed neutrinos and from 1.2 to 2.4 TeV in the minimal version. There is a point worth noting: these mass limits are in agreement with the recent analysis [16], showing that there are indications of Z' in electroweak precision data. We do hope that the new experimental data from the collaborations at BNL and Fermilab will bring new indications of the extra neutral gauge boson Z'—one of the best motivated extensions of the SM.

In this work, we considered only the CP conservating kaon decay $K^+ \longrightarrow \pi^+\nu\bar{\nu}$. Implications for the CP violating K and B decays are subjects of future studies.

One of the authors (H. N. L.) thanks the APCTP for financial support and hospitality extended to him. This work is supported in part by Vietnam National Research program on Natural Sciences under grant KT-04.1.2.

REFERENCES

1. J. Bijnens, G. Ecker, and J. Gasser, in *The DAPHNE Physics Handbook*, Ed. by L. Maiani, G. Pancheri, and N. Paver (Frascati, 1994), Vol. 1, p. 115; J. Bijnens, G. Colangelo, G. Ecker, *et al.*, Nucl. Phys. B **508**, 263 (1997); A. Buras, in *Probing of the Standard Model of Particles Interactions*, Ed. by F. David and R. Gupta (Elsevier, Amsterdam, 1998); hep-ph/9806471; G. Buchalla, Preprint CERN-TH/2000-057; hep-ph/0002207; G. D'Ambrosio, INFNNA-IV-2000/5; hep-ph/0002254; A. J. Buras, P. Gambino, M. Gorbahn, *et al.*, Nucl. Phys. B **592**, 55 (2001); hep-ph/0007313.
2. T. Inami and C. S. Lim, Prog. Theor. Phys. **65**, 297 (1981); A. J. Buras, Phys. Rev. Lett. **46**, 1354 (1981).
3. F. Pisano and V. Pleitez, Phys. Rev. D **46**, 410 (1992); R. Foot, O. F. Hernández, F. Pisano, and V. Pleitez, Phys. Rev. D **47**, 4158 (1993).
4. P. H. Frampton, Phys. Rev. Lett. **69**, 2889 (1992).
5. R. Foot, H. N. Long, and Tuan A. Tran, Phys. Rev. D **50**, R34 (1994).
6. H. N. Long, Phys. Rev. D **54**, 4691 (1996).
7. P. B. Pal, Phys. Rev. D **52**, 1659 (1995).
8. D. G. Dumm, F. Pisano, and V. Pleitez, Mod. Phys. Lett. A **9**, 1609 (1994); J. T. Liu and D. Ng, Phys. Rev. D **50**, 548 (1994).
9. H. N. Long and V. T. Van, J. Phys. G **25**, 2319 (1999).
10. D. Ng, Phys. Rev. D **49**, 4805 (1994).
11. P. Jain and S. D. Joglekar, Phys. Lett. B **407**, 151 (1997); H. Frampton and M. Harada, Phys. Rev. D **58**, 095013 (1998); H. N. Long and T. Inami, Phys. Rev. D **61**, 075002 (2000); M. B. Tully and G. C. Joshi, Phys. Lett. B **466**, 333 (1999); N. A. Ky, H. N. Long, and D. V. Soa, Phys. Lett. B **486**, 140 (2000).
12. W. Marciano and Z. Parsa, Phys. Rev. D **53**, R1 (1996).
13. H. Fritzsch, Phys. Lett. B **73B**, 317 (1978); Nucl. Phys. B **155**, 189 (1979).
14. Caso *et al.* (Particle Data Group), Eur. Phys. J. C **3**, 1 (1998).
15. S. Adler, M. S. Atiya, I.-H. Chiang, *et al.*, Phys. Rev. Lett. **84**, 3768 (2000).
16. J. Erler and P. Langacker, Phys. Rev. Lett. **84**, 212 (2000).

Journal of Experimental and Theoretical Physics, Vol. 92, No. 4, 2001, pp. 552–560.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 4, 2001, pp. 638–648.

ATOMS, SPECTRA, RADIATION

An Algorithm for Solving the Optical Problem for Stratified Anisotropic Media

S. P. Palto

Shubnikov Institute of Crystallography, Russian Academy of Sciences, Leninskiĭ pr. 59, Moscow, 117333 Russia
e-mail: palto@online.ru; lbf@ns.crys.ras.ru

Received November 2, 2000

Abstract—An algorithm for solving the Maxwell equations for propagation of light through anisotropic stratified media is considered. The algorithm uses the Berreman matrices of order 4×4. In contrast to the numerical methods suggested by Berreman, the new method is exact. The Sylvester theorem for calculating functions of a matrix and the Laguerre method for determining eigenvalues provide the basis for an algorithm with an efficiency comparable to that of the algorithms based on analytic solutions, which exist only in the case of uniaxial media. The method suggested in this paper allows for the analysis of complex optical systems where the effects of biaxiality, magnetic anisotropy, and optical activity play an important role. *© 2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Methods for analyzing the transmission and reflection in stratified anisotropic media were suggested by Teitler and Henvis in [1] and by Berreman in [2–4] almost simultaneously. The methods described in [1] and [3] are equivalent from the mathematical point of view, although the approaches are different. Berreman considers an optical medium with continuously varying parameters, which makes it possible to write the Maxwell equations in a differential matrix form. In contrast to the well-known approximate method of the Jones matrices [5], which are of order 2×2, Berreman's method uses 4×4 matrices. The increased order is the cost of the accuracy and generality. The Berreman matrix defines a linear transformation of the four tangential components of the electric and magnetic fields at the input of an optical system to the corresponding components at its output. This makes it possible to calculate simultaneously both the transmission and reflection of a light wave incident on a planar optical system at an arbitrary angle with account for interference effects of multiple reflection.

As soon as the Berreman method was published, it gained recognition and received widespread use, especially in designing liquid crystal optics, where reflection effects are crucial. Later, analytic expressions for the Berreman matrices were found for uniaxial optical media [6–8]. However, no general analytic solutions in the biaxial case are available. Thus, the most efficient methods for determining the Berreman matrices are approximate ones; such methods are studied in [4]. They are based on the expansion of a matrix in the finite Taylor series. Thus, even a homogeneous anisotropic medium must be divided into very thin sublayers for which Taylor's series is sufficiently accurate. It seems that it is the complexity of calculating the Berreman matrices that causes the development of the alternative method of Jones' matrices [9], although the latter method is basically approximate.

In this paper, we suggest an exact and efficient algorithm for determining the Berreman matrices. It is based on the Sylvester theorem [10], which represents a function of a matrix as a finite series, and on the Laguerre method for finding a complex root of a polynomial of an arbitrary degree [11]. The algorithm was applied to modeling electrooptics of liquid crystals used in designing liquid crystal displays. The method is general in the sense that its field of applicability is not restricted to modeling uniaxial optical media only. It is equally effective when the effects of biaxiality, magnetic anisotropy, and optical activity of the medium are taken into account.

2. THEORY AND METHOD

In this section, we reiterate the main results obtained by Berreman on the basis of transforming the Maxwell equations for linear media to the matrix form.

Let a plane monochromatic wave be incident in the plane *xz* on a plane-parallel plate at an arbitrary angle α to the normal (see Fig. 1). For brevity, Fig. 1 presents only the permittivity ellipsoid, although the magnetic susceptibility and optical activity can be anisotropic as well. We assume that the optical parameters of the plate smoothly depend on *z*, which makes it possible to use the differential formalism. According to [3], the tangential components of the electric and magnetic fields of the light wave are written in the matrix form as

$$\frac{\partial}{\partial z}\boldsymbol{\chi} = \frac{i\omega}{c}\boldsymbol{\Delta}\boldsymbol{\chi}, \tag{1}$$

1063-7761/01/9204-0552$21.00

where

$$\boldsymbol{\chi} = \begin{pmatrix} E_x \\ H_y \\ E_y \\ -H_x \end{pmatrix}, \quad \boldsymbol{\Delta} = \begin{pmatrix} \Delta_{11} & \Delta_{12} & \Delta_{13} & \Delta_{14} \\ \Delta_{21} & \Delta_{22} & \Delta_{23} & \Delta_{24} \\ \Delta_{31} & \Delta_{32} & \Delta_{33} & \Delta_{34} \\ \Delta_{41} & \Delta_{42} & \Delta_{43} & \Delta_{44} \end{pmatrix},$$

E_x, E_y, H_x, and H_y are the tangential components of the electric and magnetic fields, respectively. In the general case, the components of the matrix $\boldsymbol{\Delta}$ are expressed in terms of the permittivity and permeability tensors and the optical rotation tensor (see Appendix).

In many important cases, the magnetic anisotropy and optical activity can be neglected; then, the number of nonzero components of Δ is reduced to 10 [4]:

$$\Delta_{11} = -\frac{ck_x(\varepsilon_{13}\cos\phi - \varepsilon_{23}\sin\phi)}{\omega\varepsilon_{33}}, \tag{2.1}$$

$$\Delta_{12} = 1 - \frac{1}{\varepsilon_{33}}\left(\frac{ck_x}{\omega}\right)^2, \tag{2.2}$$

$$\Delta_{13} = -\frac{ck_x(\varepsilon_{13}\sin\phi + \varepsilon_{23}\cos\phi)}{\omega\varepsilon_{33}}, \tag{2.3}$$

$$\Delta_{21} = [\beta\cos 2\phi - \gamma\sin 2\phi] + \eta, \tag{2.4}$$

$$\Delta_{23} = \beta\sin 2\phi + \gamma\cos 2\phi, \tag{2.5}$$

$$\Delta_{43} = -[\beta\cos 2\phi - \gamma\sin 2\phi] + \eta - \left(\frac{ck_x}{\omega}\right)^2, \tag{2.6}$$

$$\Delta_{34} = 1, \tag{2.7}$$

$$\Delta_{22} = \Delta_{11}, \tag{2.8}$$

$$\Delta_{41} = \Delta_{23}, \tag{2.9}$$

$$\Delta_{42} = \Delta_{13}, \tag{2.10}$$

$$\Delta_{14} = \Delta_{24} = \Delta_{31} = \Delta_{32} = \Delta_{33} = \Delta_{44} = 0, \tag{2.11}$$

$$\beta = \frac{\varepsilon_{11} - \varepsilon_{22}}{2} - \frac{\varepsilon_{13}^2 - \varepsilon_{23}^2}{2\varepsilon_{33}}, \quad \gamma = \varepsilon_{12} - \frac{\varepsilon_{13}\varepsilon_{23}}{\varepsilon_{33}}, \quad \eta = \frac{\varepsilon_{11} + \varepsilon_{22}}{2} - \frac{\varepsilon_{13}^2 + \varepsilon_{23}^2}{2\varepsilon_{33}}, \tag{2.12}$$

$$\varepsilon_{11} = \varepsilon_a\cos^2\psi + \varepsilon_b\sin^2\psi, \tag{2.13}$$

$$\varepsilon_{12} = (\varepsilon_a - \varepsilon_b)\sin\psi\cos\psi\cos\theta, \tag{2.14}$$

$$\varepsilon_{13} = (\varepsilon_a - \varepsilon_b)\sin\psi\cos\psi\sin\theta, \tag{2.15}$$

$$\varepsilon_{22} = (\varepsilon_a\sin^2\psi + \varepsilon_b\cos^2\psi)\cos^2\theta + \varepsilon_c\sin^2\theta, \tag{2.16}$$

$$\varepsilon_{23} = (\varepsilon_a\sin^2\psi + \varepsilon_b\cos^2\psi - \varepsilon_c)\sin\theta\cos\theta, \tag{2.17}$$

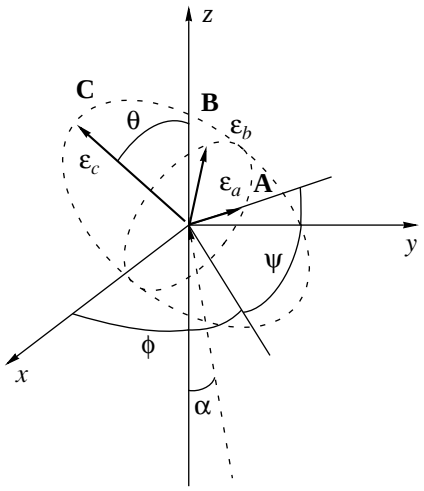


Fig. 1. Orientation of the principal axes, **A**, **B**, and **C**, of the permittivity ellipsoid: *xy* is the surface plane of the optical medium; *xz* is the incidence plane; α is the angle of incidence; θ, ϕ, and ψ are Euler's angles; and ε_a, ε_b, and ε_c are the principal values of the permittivity tensor components.

$$\varepsilon_{33} = (\varepsilon_a\sin^2\psi + \varepsilon_b\cos^2\psi)\sin^2\theta + \varepsilon_c\cos^2\theta. \tag{2.18}$$

In (2.1)–(2.18), k_x is the x component of the wave vector of the incident wave; ω is the angular frequency; θ, ϕ, and ψ are Euler's angles, which determine the orientation of the permittivity ellipsoid; and ε_a, ε_b, and ε_c are the principal values of the permittivity tensor components (see Fig. 1). We assume that the permittivity components are complex. Thus, the absorption anisotropy is taken into account.

In the case when the parameters of the optical plate of thickness h are independent of z, the integration of Eq. (1) yields

$$\boldsymbol{\chi}(h) = \exp(i\omega h\boldsymbol{\Delta}/c)\boldsymbol{\chi}(0) \equiv \mathbf{P}(h)\boldsymbol{\chi}(0), \tag{3}$$

where $\mathbf{P}(h)$ corresponds to the Berreman matrix for the homogeneous medium.

Thus, in the case of a homogeneous medium, calculation of the Berreman matrix is reduced to calculating the exponent of the matrix $\boldsymbol{\Delta}$. In the general case, when the parameters of the optical medium depend on z, the medium is divided into n layers such that the optical parameters can be assumed constant within each layer; then, integration of Eq. (1) is reduced to the multiplication of the corresponding matrices for each layer:

$$\boldsymbol{\chi}(h) = \prod_{i=1}^{n}\mathbf{P}(h_i)\boldsymbol{\chi}(0). \tag{4}$$

Equation (4) is also valid for complex optical media consisting of discrete optical elements (polaroids, phase plates, etc.). From the physical point of view, Eq. (4) is complete, since the tangential components of the electric and magnetic fields are continuous. Hence, even in the case of a stratified system of heterogeneous optical media,

no additional boundary conditions are required. Thus, even in the most general case, the numerical integration is reduced to calculating the exponent of the matrix $\mathbf{\Delta}$.

The well-known exact algorithm for calculating a function of a matrix is based on the similarity transformation

$$\mathbf{\Delta} = \mathbf{S}^{-1}\mathbf{D}\mathbf{S} \tag{5}$$

and on the calculation of the function of the elements of the diagonal matrix $\mathbf{D}$. However, this method requires that the eigenvalues of $\mathbf{\Delta}$, the matrix of eigenvectors $\mathbf{S}$, and the inverse matrix $\mathbf{S}^{-1}$ be calculated. Although the exponent of Eq. (5) has a clear physical meaning (it describes the propagation of the four characteristic waves in the forward and backward directions), this method is computationally rather costly. The method is more efficient only for uniaxial media, where analytic expressions for the eigenvalues and eigenvectors are available [8]. It seems likely that the complexity and inefficiency of the algorithms used to compute the matrices for the similarity transformation caused the development of approximate methods [3, 4]. For example, one of the methods suggested by Berreman is based on the representation of the exponent in the form of Taylor's series

$$\begin{aligned}\mathbf{P}(h) &= \exp(i\omega h\mathbf{\Delta}/c)\\ &\approx \mathbf{I} + \frac{i\omega h}{c}\mathbf{\Delta} - \frac{1}{2}\left(\frac{\omega h}{c}\right)^2\mathbf{\Delta}^2,\end{aligned} \tag{6}$$

where $\mathbf{I}$ is the identity matrix.

Berreman showed that a converging solution could be obtained if the second-order term in this expansion is taken into account. However, since Eq. (6) is an approximation, its application even for homogeneous portions of the medium requires that the medium be divided into layers much thinner than the wavelength. The procedure becomes inefficient when designing systems consisting of thick (hundreds of microns) homogeneous optical elements such as polaroids or phase plates. In [4], the numerical integration method was improved and applied to the analysis of the optics of a liquid-crystalline "twist cell." The improved procedure allowed one to use large values of h, but the method remained basically approximate and the step h was to be chosen with great caution.

In the present paper, we suggest an exact procedure. The procedure is based on the Sylvester theorem [10], which gives a formula for the function of a matrix $\mathbf{A}$ of order $n \times n$ provided that all its eigenvalues are different:

$$f(\mathbf{A}) = \sum_{k=1}^{n} f(\lambda_k)\frac{\prod_{i\neq k}(\mathbf{A} - \lambda_i\mathbf{I})}{\prod_{i\neq k}(\lambda_k - \lambda_i)}, \tag{7}$$

where λ_k are the eigenvalues of $\mathbf{A}$.

As applied to the case considered in this paper, formula (7) involves four terms:

$$\begin{aligned}\mathbf{P}(h) &= \exp(i\omega h\mathbf{\Delta}/c)\\ &\equiv \sum_{k=1}^{4}\left(\exp(i\omega h\lambda_k/c)\frac{\prod_{i\neq k}(\mathbf{\Delta} - \lambda_i\mathbf{I})}{\prod_{i\neq k}(\lambda_k - \lambda_i)}\right),\end{aligned} \tag{8}$$

which is almost equivalent to Eq. (6) in terms of the computational cost if the eigenvalues of $\mathbf{\Delta}$ are known. Since Eq. (8) is an exact representation of the function, the operation of dividing the homogeneous medium into a set of sublayers is no longer required. In contrast to the exact method based on the similarity transformation (5), the matrix of eigenvectors and its inverse are not calculated, which improves the efficiency of the method.

Thus, formula (8) reduces the problem to finding an efficient method for determining the eigenvalues of the matrix $\mathbf{\Delta}$. In the case of a uniaxial optical medium ($\varepsilon_b = \varepsilon_a$), analytic expressions for the eigenvalues are available [6–8]:

$$\lambda_{1,2} = \pm\left(\varepsilon_a - \left(\frac{ck_x}{\omega}\right)^2\right)^{1/2}, \tag{9.1}$$

$$\begin{aligned}\lambda_{3,4} = -\frac{1}{\varepsilon_{33}}\Bigg\{&\varepsilon\cos\theta\sin\phi\frac{ck_x}{\omega}\\ &\pm\left[\varepsilon_a\left(\varepsilon_c\varepsilon_{33} - (\varepsilon_c - \varepsilon\sin\theta\cos^2\phi)\left(\frac{ck_x}{\omega}\right)^2\right)\right]^{1/2}\Bigg\},\end{aligned} \tag{9.2}$$

where $\varepsilon = (\varepsilon_c - \varepsilon_a)\sin\theta$, and ε_{33} is determined by Eq. (2.18) with regard for the uniaxiality ($\varepsilon_a = \varepsilon_b$).

In the general case, the eigenvalues are the roots of the fourth-degree polynomial

$$\det(\mathbf{\Delta} - \lambda\mathbf{I}) = 0, \tag{10}$$

or the equivalent vector equation

$$\mathbf{\Lambda}\mathbf{A} = 0, \tag{11}$$

where

$$\mathbf{\Lambda} = (1\lambda\lambda^2\lambda^3\lambda^4), \quad \mathbf{A} = \begin{pmatrix} a_0\\ a_1\\ a_2\\ a_3\\ a_4\end{pmatrix}, \tag{12}$$

and a_0–a_4 are the complex coefficients of the polynomial. If the elements of $\mathbf{\Delta}$ are defined by Eqs. (2.1)–(2.11), then

the coefficients of the polynomial are given by the expressions

$$a_0 = 2\Delta_{11}\Delta_{13}\Delta_{23}\Delta_{34} + \Delta_{21}\Delta_{12}\Delta_{34}\Delta_{43} - \Delta_{11}^2\Delta_{34}\Delta_{43} - \Delta_{13}^2\Delta_{21}\Delta_{34} - \Delta_{23}^2\Delta_{12}\Delta_{34}, \quad (13.1)$$

$$a_1 = 2(\Delta_{11}\Delta_{34}\Delta_{43} - \Delta_{13}\Delta_{23}\Delta_{34}), \quad (13.2)$$

$$a_2 = \Delta_{11}^2 - \Delta_{34}\Delta_{43} - \Delta_{21}\Delta_{12}, \quad (13.3)$$

$$a_3 = -2\Delta_{11}, \quad (13.4)$$

$$a_4 = 1. \quad (13.5)$$

Another important thing is that our method uses the Laguerre method [11] for determining the roots of polynomials. The Laguerre method guarantees rapid convergence to one of the roots of the polynomial given an arbitrary initial approximation and provides a highly accurate result; the accuracy is actually limited only by the precision of the computer representation of floating-point numbers. It is clear that the Laguerre method can be used for our purposes if it is supplemented by a procedure for eliminating previously calculated roots.

For the sake of convenience, we introduce the operator **L**, which will be called the Laguerre operator, such that

$$\lambda_1 = \mathbf{L}(\mathbf{A}); \quad (14)$$

that is, its application to a vector **A** (the components of **A** are the coefficients of the polynomial under consideration, see Eq. (11)) gives the value of a root of the polynomial, which corresponds to an eigenvalue of the matrix **Δ**. Figure 2 presents a rather detailed flowchart of the algorithm based on the Laguerre operator.

Having calculated one of the roots, one can easily eliminate it and obtain a polynomial of a lower degree. The procedure of calculating all eigenvalues can be represented as a sequence of transformations

$$\begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \end{pmatrix} \longrightarrow \begin{pmatrix} b_0 = a_1 + b_1\lambda_1 \\ b_1 = a_2 + b_2\lambda_1 \\ b_2 = a_3 + b_3\lambda_1 \\ b_3 = a_4 \\ \lambda_1 = \mathbf{L}(\mathbf{A}) \end{pmatrix} \longrightarrow \begin{pmatrix} c_0 = b_1 + c_1\lambda_2 \\ c_1 = b_2 + c_2\lambda_2 \\ c_2 = b_3 \\ \lambda_2 = \mathbf{L}(\mathbf{B}) \\ \lambda_1 \end{pmatrix} \quad (15)$$

$$\longrightarrow \begin{pmatrix} d_0 = c_1 + d_1\lambda_3 \\ d_1 = c_2 \\ \lambda_3 = \mathbf{L}(\mathbf{C}) \\ \lambda_2 \\ \lambda_1 \end{pmatrix} \longrightarrow \begin{pmatrix} c_2 \\ \lambda_4 = -d_0/d_1 \\ \lambda_3 \\ \lambda_2 \\ \lambda_1 \end{pmatrix},$$

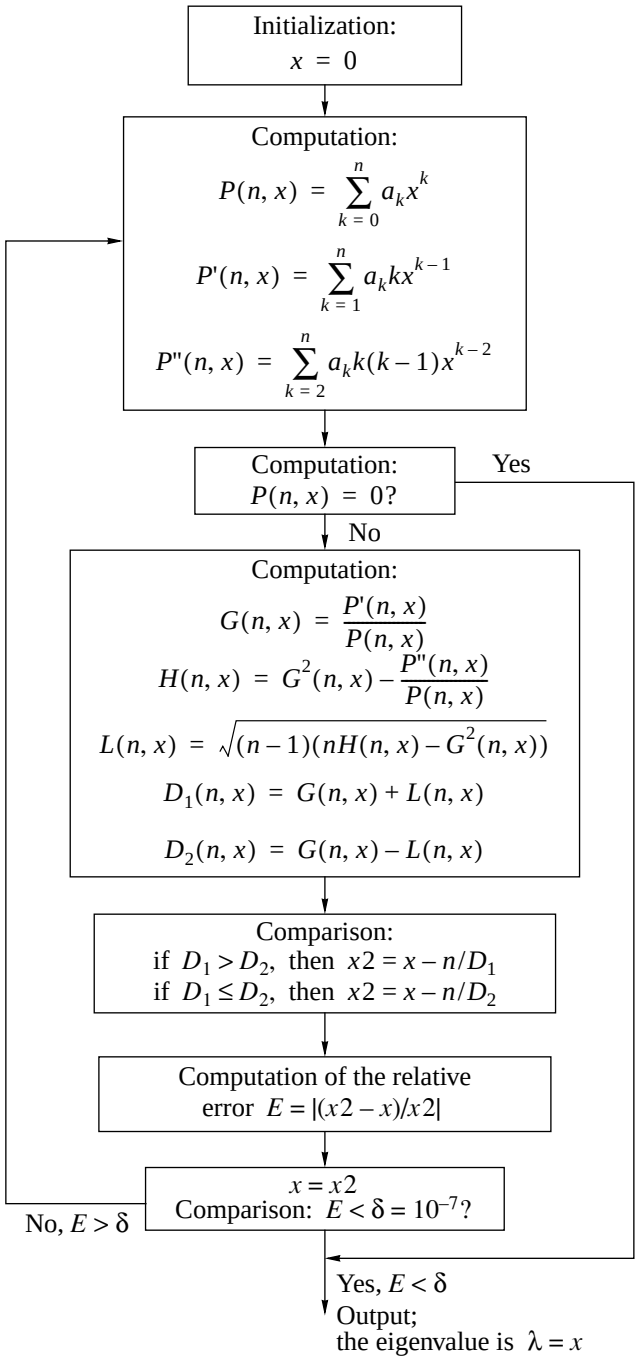


Fig. 2. Flowchart of the algorithm for determining an eigenvalue by the Laguerre method. The input parameters are the degree of the polynomial n and its coefficients $\{a_0, a_1, \ldots, a_n\}$.

where

$$\mathbf{B} = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} c_0 \\ c_1 \\ c_2 \end{pmatrix}. \tag{16}$$

Thus, applying the Laguerre operator three times in accordance with procedure (15) from left to right, we obtain all eigenvalues of $\mathbf{\Delta}$. Then, we substitute them into Eq. (8) to obtain the Berreman matrix. Besides the Laguerre operator, procedure (15) involves only the addition and multiplication operations, which, combined with the rapid convergence of the Laguerre method, guarantees high efficiency comparable with that of the method for determining eigenvalues based on the analytic expressions (9.1) and (9.2) in the case of uniaxial systems. For comparison, note that even an optimized analytic procedure (where the repeated calculation of the functions sin$x$ and others is eliminated) is only about 1.5 times quicker (the computation time of the Berreman matrix is about 100 μs on a Celeron 500–based computer).

It must be stressed that formula (8) assumes that all eigenvalues are different. In some cases, when the medium or geometry are symmetric (e.g., the medium is optically isotropic or the light is incident at exactly right angles) some eigenvalues can be multiple. In this case, an indeterminacy of the type 0/0 occurs, which can be evaluated using the L'Hospital rule. However, this can make the procedure more complicated and deteriorate its efficiency. An alternative possibility is to introduce a small perturbation that cannot affect the computation results. For example, in the case of an isotropic medium, the refraction indices can be made slightly different, for example, in the seventh decimal digit. This cannot noticeably affect the results even when the thickness of elements is hundreds of microns.[1] In the case of the normal incidence, it is sufficient to specify a very small deviation from the normal direction; for example, the deviation of 10^{-6} rad is unlikely to be detected in an experiment.

3. DETERMINING THE TRANSMISSION AND REFLECTION

In this section, we obtain formulas for determining the field in the reflected and transmitted waves. Although this problem was considered by Berreman, his original work contains an error in the corresponding expressions (formulas (86)–(90) in [3]).

As in [3], we assume that an anisotropic plate or an optical system consisting of discrete elements is confined between two nonabsorbing isotropic media with the refraction indices n_1 and n_2, respectively. A light wave is incident from the side of the medium with the index n_1 at the angle α_1 to the normal to the plate surface and goes to the second medium at the angle α_2. The field at the input of the system is determined by the superposition of the incident and the reflected waves, while only the transmitted wave is present at the output:

$$\mathbf{\chi}_T = \mathbf{P}(h)(\mathbf{\chi}_I + \mathbf{\chi}_R). \tag{17}$$

Here, $\mathbf{P}(h)$ is the Berreman matrix, which is determined by product (4) in the case of a heterogeneous medium or a system of optical elements; $\mathbf{\chi}_I$, $\mathbf{\chi}_T$, and $\mathbf{\chi}_R$ are the vectors of the incident, transmitted, and reflected waves, respectively.

In isotropic media, there is a definite relation between the components of the electric and magnetic fields. Thus,

$$\mathbf{\chi}_I = \begin{pmatrix} E_x \\ r_x E_x \\ E_y \\ r_y E_y \end{pmatrix}, \quad \mathbf{\chi}_R = \begin{pmatrix} R_x \\ -r_x R_x \\ R_y \\ -r_y R_y \end{pmatrix}, \quad \mathbf{\chi}_T = \begin{pmatrix} T_x \\ r_x^* T_x \\ T_y \\ r_y^* T_y \end{pmatrix}, \tag{18}$$

where

$$r_x = \frac{n_1}{\cos\alpha_1}, \quad r_y = n_1\cos\alpha_1, \quad r_x^* = \frac{n_2}{\cos\alpha_2}, \quad r_y^* = n_2\cos\alpha_2, \quad \frac{\sin\alpha_2}{\sin\alpha_1} = \frac{n_1}{n_2}. \tag{19}$$

Multiply both sides of Eq. (17) on the left by the inverse of the Berreman matrix $\mathbf{F} = \mathbf{P}^{-1}$ to obtain a system of linear equations, which is easily solved for the unknown components of the field R_x, R_y, T_x, and T_y:

$$T_y = \frac{2(r_y E_y - acE_x)}{bc + d}, \tag{20.1}$$

$$T_x = 2aE_x + bT_y, \tag{20.2}$$

$$R_x = (F_{11} + F_{12}r_x^*)T_x + (F_{13} + F_{14}r_y^*)T_y - E_x, \tag{20.3}$$

$$R_y = (F_{31} + F_{32}r_x^*)T_x + (F_{33} + F_{34}r_y^*)T_y - E_y, \tag{20.4}$$

where

$$a = \frac{r_x}{(F_{11} + F_{12}r_x^*)r_x + F_{22}r_x^* + F_{21}},$$
$$b = -\frac{(F_{13} + F_{14}r_y^*)r_x + F_{24}r_y^* + F_{23}}{(F_{11} + F_{12}r_x^*)r_x + F_{22}r_x^* + F_{21}},$$
$$c = (F_{31} + F_{32}r_x^*)r_y + F_{42}r_x^* + F_{41}, \tag{20.5}$$
$$d = (F_{33} + F_{34}r_y^*)r_y + F_{44}r_y^* + F_{43},$$

and F_{ij} ($i, j = 1, 2, 3, 4$) are the components of $\mathbf{F}$.

[1] We assume that the computations are performed with double precision.

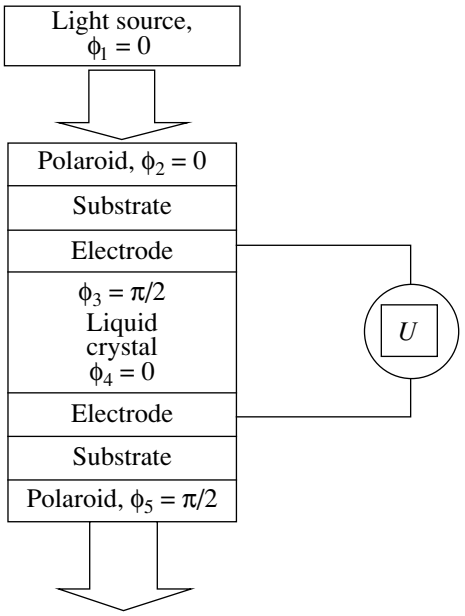


Fig. 3. A scheme of the liquid-crystal optical system to be modeled. *U* is the source of voltage applied to the liquid crystal.

The transmission and reflection coefficients, *T* and *R*, are determined as the ratio of the energy flux in the transmitted and the reflected wave, respectively, to the energy flux in the incident wave. In practice, the simplest case is most commonly encountered when the air acts as the first and the second medium ($n_1 = n_2 \approx 1$ and $\alpha_1 = \alpha_2 = \alpha$). Then

$$T = \frac{|T_x/\cos\alpha|^2 + |T_y|^2}{|E_x/\cos\alpha|^2 + |E_y|^2},$$
$$R = \frac{|R_x/\cos\alpha|^2 + |R_y|^2}{|E_x/\cos\alpha|^2 + |E_y|^2}. \quad (21)$$

We stress that all arithmetic operations are performed over complex numbers, and the square of the absolute value is calculated by an appropriate procedure.

4. APPLICATION OF THE METHOD FOR THE ANALYSIS OF THE "TWIST-EFFECT" OPTICS IN A LIQUID CRYSTAL TAKING INTO ACCOUNT POSSIBLE BIAXIALITY

By way of example, we give an analysis of the optics of a system depicted in Fig. 3. It includes a light source and a liquid crystal cell placed between two crossed polaroids. First, we analyze the static electrooptics assuming that the liquid crystal is uniaxial ($n_b = n_a \equiv n_\perp$). Then, we artificially introduce two types of biaxiality ($n_b < n_a$ and $n_b > n_a$) and consider how they affect the electrooptics and angular characteristics of the optical contrast when the electric field is switched.

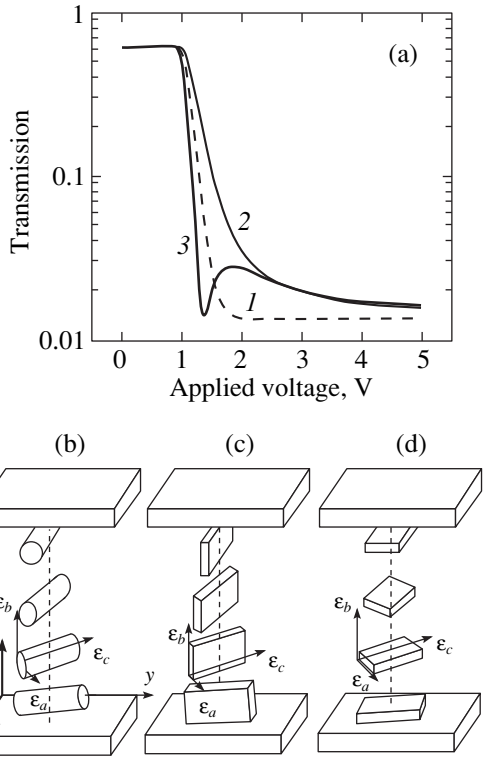


Fig. 4. Dependence of the optical transmission on the voltage applied. (a) (*1*) a uniaxial liquid crystal; (*2*) a positive biaxiality is introduced ($n_b = 1.58 > n_a = 1.54$); (*3*) a negative biaxiality is introduced ($n_b = 1.50 < n_a = 1.54$). It is assumed in the calculations that the principal axis **A** (Fig. 1) is oriented at the angle $\psi = 0$. The initial ($U = 0$) distribution of molecules of the liquid crystal: (b) uniaxial liquid crystal; (c) positive biaxiality; (d) negative biaxiality. Biaxial molecules of the liquid crystal are depicted by parallelepipeds. The light is polarized in the direction of the axis *x* and travels along the axis *z*.

Determine the parameters required for the analysis of the optical elements.

Polaroids. Polaroids are modeled by a uniaxial optical medium ($\varepsilon_a = \varepsilon_b$) with an anisotropy of the absorption coefficient ($k_c \gg k_a$). Thus, the principal values of the complex permittivity are expressed in terms of the refraction indices n_c, n_a and the absorption coefficients k_c, k_a as

$$\varepsilon_{c,a} = \left(n_{c,a} + \frac{i\lambda k_{c,a}}{4\pi}\right)^2, \quad (22)$$

where λ is the light wavelength and the subscripts *a* and *c* correspond to the principal axes of the ellipsoid **A** and **C**, respectively (see Fig. 1).[2]

[2] Sometimes, a different complex representation of the light wave is employed and, as a result, the second term in parentheses of Eq. (22) appears with the minus sign. In that representation, the signs at all imaginary quantities in this paper should be changed to opposite.

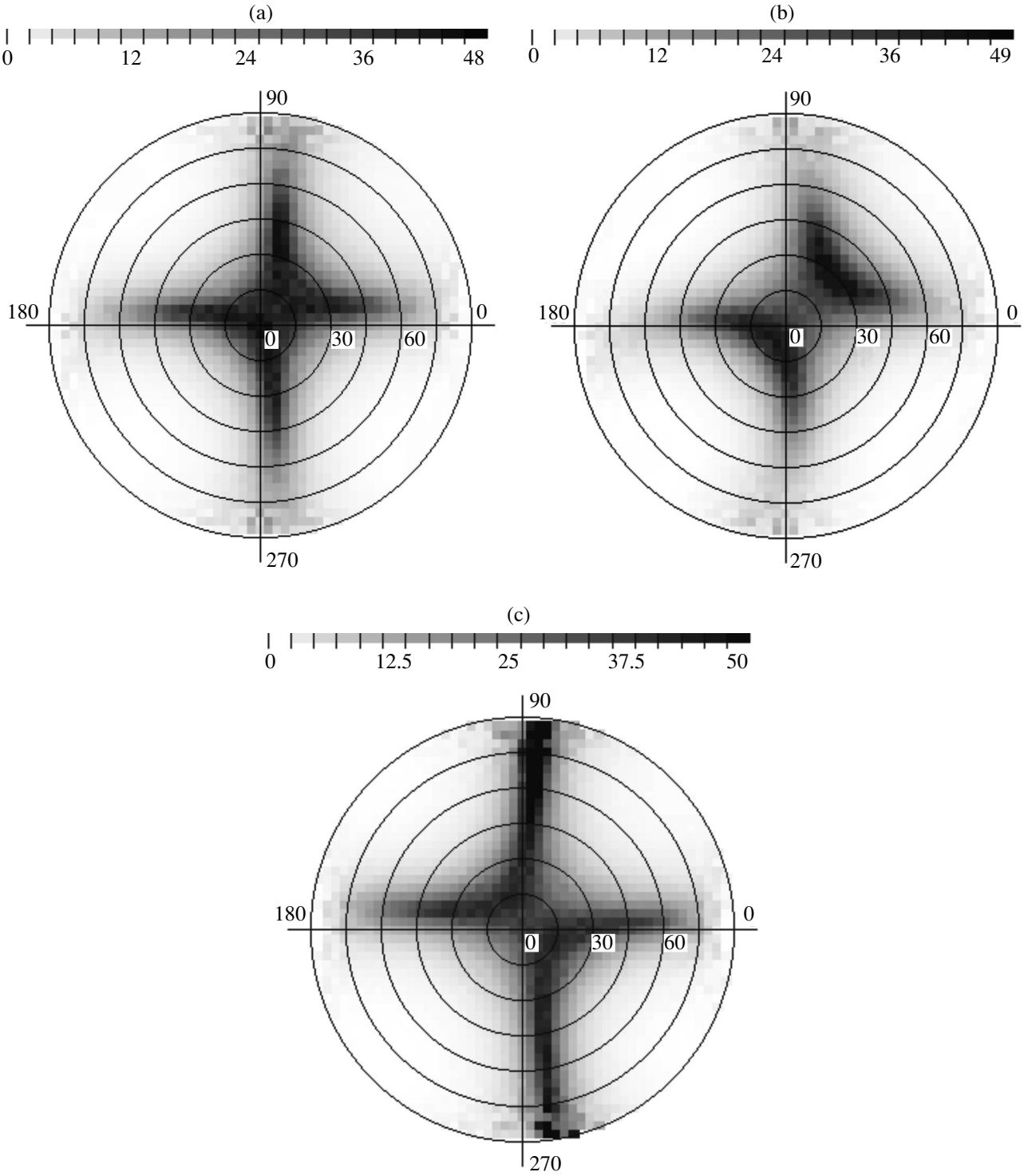


Fig. 5. Angular characteristics of the contrast ratio (see text) in the case of (a) a uniaxial liquid crystal; (b) negative biaxiality; and (c) positive biaxiality. The contrast ratio is represented by the intensity in the polar frame of reference. The incident angle of light is plotted on the radius. The angle plotted on the external circle corresponds to the azimuth angle, which determines the orientation of the plane of light incidence with respect to the axis *x* (see Fig. 4b).

Typical 200 μm polaroid films used in liquid-crystal displays can be modeled by the refractive index $n_a \approx n_c = 1.5$ and the absorption coefficients $k_a = 0.001\ \mu m^{-1}$ and $k_c = 0.02\ \mu m^{-1}$ at the wavelength 550 nm. The transmission coefficient of two polaroids arranged such that the weak absorption axes are parallel to the polarization of the incident light is ~0.6. If the polaroids are crossed, then the transmission coefficient is ~0.01. When the

Berreman matrices are calculated, it must be taken into account that the principal axis **C** lies in the plane of the polaroid film ($\theta = \pi/2$).

In the scheme shown in Fig. 3, the first polaroid is oriented such that its weak absorption axis is parallel to the polarization direction of the light source ($\phi_2 = \phi_1 = 0$) and is perpendicular to the orientation of the liquid crystal director on the first surface ($\phi_3 = \pi/2$). The second polaroid (the analyzer) is skewed with respect to the first one ($\phi_5 = \pi/2$).

Isotropic optical glasses. The substrates of a liquid-crystal cell are modeled by an isotropic medium. All the principal values of the permittivity ellipsoid are equal, and the absorption coefficient is assumed to be zero. For standard glasses, we can set $\varepsilon_a = \varepsilon_b = \varepsilon_c \approx 1.53^2 \approx 2.34$. In our model, we use 200-μm-thick glasses.

Electrodes. Electrodes are isotropic ($\varepsilon_a = \varepsilon_b = \varepsilon_c$), but the real part of the refraction index is greater than for glass, and the absorption coefficient is nonzero. In this model, we used 0.01-μm-thick electrodes with the real part of the refraction index equal to $n = 1.7$ and the absorption coefficient $k = 0.1\ \mu m^{-1}$.

Distribution of liquid crystal molecules in an electric field. The characteristics of the model liquid crystal are assumed to be as follows.

The coefficients of elasticity are $K_{11} = 6.4$ pN, $K_{22} = 3$ pN, and $K_{33} = 10$ pN.

Low-frequency permittivity indices are $\varepsilon_\perp = 6.7$ and $\varepsilon_\parallel = 19.7$.

The optical refraction indices at the wavelength 550 nm are $n_a \equiv n_\perp = 1.54$, $n_c \equiv n_\parallel = 1.72$, and the direction of the director of the liquid crystal molecules corresponds to the principal axis **C** of the permittivity ellipsoid (see Fig. 1).

The boundary conditions are as follows: the angle of deviation of the liquid crystal director from the normal is 89°, the zenithal binding energy is 0.2 mJ/m^2 the azimuth binding energy is 0.1 mJ/m^2, and the relative azimuth orientation of the director on the opposite surfaces is 90° (twist cell). At the first boundary, the director is oriented perpendicular to the polarizer; on the second boundary, it is perpendicular to the analyzer.

The thickness of the liquid crystal is 3.1 μm.

The distribution of the director in the electric field is calculated by solving the nonlinear differential equations of the force moment balance obtained as a result of the minimization of the Frank free energy [12].

Figure 4 shows the calculated dependence of the transmission of the entire optical system on the voltage applied to a liquid-crystal cell under the normally incident monochromatic light with the wavelength $\lambda = 550$ nm. Curve *1* corresponds to the uniaxial liquid crystal ($n_b = n_a = 1.54$). If a positive biaxiality is introduced ($n_b = 1.58 > n_a = 1.54$, $\psi = 0$), then the curve becomes less steep in the range 1–3 V. If a negative biaxiality is introduced ($n_b = 1.50 < n_a = 1.54$, $\psi = 0$), then two additional extrema occur (curve *3*) for voltages greater than 1 V. Note that such behavior of uniaxial liquid-crystal twist cells is possible only if the incident light is oblique. Hence, the shape of the dependence (curve *3*) can be used as a characteristic one for biaxial liquid crystals under the normal light incidence.

The presence of biaxiality is clearly visible in angular dependencies of the optical contrast (see Fig. 5). The contrast ratio was calculated as

$$\gamma = \frac{T_0 - T_1}{T_1}, \tag{23}$$

where T_0 corresponds to the transmission in the off (bright) state, and T_1 corresponds to the transmission in the on (dark) state at 4.5 V applied to the cell. In the case of a uniaxial liquid crystal (Fig. 5a) the well-known asymmetric contrast characteristic of the twist effect [12] is observed. Biaxiality drastically changes the angular characteristics. As is seen from Fig 5b, the negative biaxiality drastically deteriorates the angular characteristics, while the positive one (Fig. 5c) makes the view angle greater. Thus, the use of biaxial liquid crystals adds a new degree of freedom that can be used to improve angular characteristics of LCDs.

5. CONCLUSIONS

A new efficient method for analyzing the optics of heterogeneous anisotropic stratified media is described. The application of the method to the analysis of an optical system that includes a liquid-crystal twist cell with a biaxial liquid crystal is demonstrated.

ACKNOWLEDGMENTS

The author is grateful to Prof. L.M. Blinov and Prof. Ph. Martinot-Lagarde of the Laboratory of Solid-State Physics, Orsay, France, who initiated my work in the field of numerical modeling of liquid crystals.

This work was supported by the Russian Foundation for Basic Research, project no. 99-02-16484.

APPENDIX

In the most general case, the components of the matrix $\mathbf{\Delta}$ are expressed in terms of the components of the permittivity tensor

$$\boldsymbol{\varepsilon} = \|\varepsilon_{ij}\|, \quad i, j \in \{x, y, z\}, \tag{A.1}$$

the permeability tensor

$$\boldsymbol{\mu} = \|\mu_{ij}\|, \quad i, j \in \{x, y, z\}, \tag{A.2}$$

and the optical rotation tensor

$$\boldsymbol{\rho} = \|\rho_{ij}\|, \quad \boldsymbol{\rho}^* = \|\rho_{ij}^*\|, \quad i, j \in \{x, y, z\}, \tag{A.3}$$

as follows:

$$\Delta_{11} = \rho^*_{yx} + \left(\rho^*_{yz} + \frac{ck_x}{\omega}\right)a_1 + \mu_{yz}b_1,$$

$$\Delta_{12} = \mu_{yy} + \left(\rho^*_{yz} + \frac{ck_x}{\omega}\right)a_2 + \mu_{yz}b_2,$$

$$\Delta_{13} = \rho^*_{yy} + \left(\rho^*_{yz} + \frac{ck_x}{\omega}\right)a_3 + \mu_{yz}b_3,$$

$$\Delta_{14} = -\mu_{yx} - \left(\rho^*_{yz} + \frac{ck_x}{\omega}\right)a_4 - \mu_{yz}b_4,$$

$$\Delta_{21} = \varepsilon_{xx} + \varepsilon_{xz}a_1 + \rho_{xz}b_1,$$

$$\Delta_{22} = \rho_{xy} + \varepsilon_{xz}a_2 + \rho_{xz}b_2,$$

$$\Delta_{23} = \varepsilon_{xy} + \varepsilon_{xz}a_3 + \rho_{xz}b_3,$$

$$\Delta_{24} = -(\rho_{xx} + \varepsilon_{xz}a_4 + \rho_{xz}b_4),$$

$$\Delta_{31} = -(\rho^*_{xx} + \rho^*_{xz}a_1 + \mu_{xz}b_1),$$

$$\Delta_{32} = -(\mu_{xy} + \rho^*_{xz}a_2 + \mu_{xz}b_2),$$

$$\Delta_{33} = -(\rho^*_{xy} + \rho^*_{xz}a_3 + \mu_{xz}b_3),$$

$$\Delta_{34} = \mu_{xx} + \rho^*_{xz}a_4 + \mu_{xz}b_4,$$

$$\Delta_{41} = \varepsilon_{yx} + \varepsilon_{yz}a_1 + \left(\rho_{yz} - \frac{ck_x}{\omega}\right)b_1,$$

$$\Delta_{42} = \rho_{yy} + \varepsilon_{yz}a_2 + \left(\rho_{yz} - \frac{ck_x}{\omega}\right)b_2,$$

$$\Delta_{43} = \varepsilon_{yy} + \varepsilon_{yz}a_3 + \left(\rho_{yz} - \frac{ck_x}{\omega}\right)b_3,$$

$$\Delta_{44} = -\rho_{yx} - \varepsilon_{yz}a_4 - \left(\rho_{yz} - \frac{ck_x}{\omega}\right)b_4.$$

Here,

$$a_1 = (\rho^*_{zx}\rho_{zz} - \varepsilon_{zx}\mu_{zz})/(\varepsilon_{zz}\mu_{zz} - \rho_{zz}\rho^*_{zz}),$$

$$a_2 = [\mu_{zy}\rho_{zz} - (\rho_{zy} + ck_x/\omega)\mu_{zz}]/(\varepsilon_{zz}\mu_{zz} - \rho_{zz}\rho^*_{zz}),$$

$$a_3 = [(\rho^*_{zy} - ck_x/\omega)\rho_{zz} - \varepsilon_{zy}\mu_{zz}]/(\varepsilon_{zz}\mu_{zz} - \rho_{zz}\rho^*_{zz}),$$

$$a_4 = (\mu_{zx}\rho_{zz} - \rho_{zx}\mu_{zz})/(\varepsilon_{zz}\mu_{zz} - \rho_{zz}\rho^*_{zz}),$$

$$b_1 = (\rho^*_{zz}\varepsilon_{zx} - \varepsilon_{zz}\rho^*_{zx})/(\varepsilon_{zz}\mu_{zz} - \rho_{zz}\rho^*_{zz}),$$

$$b_2 = [(\rho_{zy} + ck_x/\omega)\rho^*_{zz} - \varepsilon_{zz}\mu_{zy}]/(\varepsilon_{zz}\mu_{zz} - \rho_{zz}\rho^*_{zz}),$$

$$b_3 = [\rho^*_{zz}\varepsilon_{zy} - (\rho^*_{zy} - ck_x/\omega)\varepsilon_{zz}]/(\varepsilon_{zz}\mu_{zz} - \rho_{zz}\rho^*_{zz}),$$

$$b_4 = (\rho^*_{zz}\rho_{zx} - \varepsilon_{zz}\mu_{zx})/(\varepsilon_{zz}\mu_{zz} - \rho_{zz}\rho^*_{zz}).$$

Note that all components are, generally, complex numbers, and tensors (A.1)–(A.3) are defined such that the equations of state that relate the induction of the electric (**D**) and magnetic (**B**) fields with their intensities are as follows:

$$\mathbf{D} = \boldsymbol{\varepsilon}\mathbf{E} + \boldsymbol{\rho}\mathbf{H}, \tag{A.4}$$

$$\mathbf{B} = \boldsymbol{\rho}^*\mathbf{E} + \boldsymbol{\mu}\mathbf{H}, \tag{A.5}$$

$$\mathbf{D} = (D_x, D_y, D_z)^T, \quad \mathbf{B} = (B_x, B_y, B_z)^T,$$
$$\mathbf{E} = (E_x, E_y, E_z)^T, \quad \mathbf{H} = (H_x, H_y, H_z)^T. \tag{A.6}$$

Note that, for example, in the presence of optical activity, the electric induction depends not only on the intensity of the electric field, but on the intensity of the magnetic field as well. Equations (A.4) and (A.5) are implied by the linear dependence between the six components of the electromagnetic field $(E_x, E_y, E_z, H_x, H_y, H_z)^T$ and $(D_x, D_y, D_z, B_x, B_y, B_z)^T$ determined by a 6×6 matrix. The second and the fourth quadrants of this matrix determine the permittivity and permeability tensors, and the first and third determine the optical activity tensors. A method for determining the optical activity tensors for certain particular cases can be found in [3].

REFERENCES

1. S. Teitler and B. W. Henvis, J. Opt. Soc. Am. **60**, 830 (1970).
2. D. W. Berreman and T. J. Scheffer, Phys. Rev. Lett. **25**, 577 (1970).
3. D. W. Berreman, J. Opt. Soc. Am. **62**, 502 (1972).
4. D. W. Berreman, J. Opt. Soc. Am. **63**, 1374 (1973).
5. R. C. Jones, J. Opt. Soc. Am. **31**, 488 (1941); A. Gerrard and J. M. Burch, *Introduction to Matrix Methods in Optics* (Wiley, New York, 1975; Mir, Moscow, 1978).
6. F. Cuypers and H. Pauwels, Liq. Cryst. **3**, 1157 (1988).
7. P. Allia, M. Arlone, C. Oldano, and L. Trossi, Mol. Cryst. Liq. Cryst. **179**, 253 (1990).
8. J. Papánek and Ph. Martinot-Lagarde, J. Phys. II **6**, 205 (1996).
9. P. Yeh, J. Opt. Soc. Am. **72**, 507 (1982).
10. G. A. Korn and T. M. Korn, *Mathematical Handbook for Scientists and Engineers* (McGraw-Hill, New York, 1961; Nauka, Moscow, 1973).
11. http://mathworld.wolfram.com/LaguerresMethod.htm; W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in FORTRAN: The Art of Scientific Computing* (Cambridge Univ. Press, Cambridge, 1992), pp. 365–366; A. Ralston and P. Ya. Rabinowitz, in *A First Course in Numerical Analysis* (McGraw-Hill, New York, 1978), Parag. 8.9–8.13.
12. L. M. Blinov and V. G. Chigrinov, *Electrooptic Effects in Liquid Crystal Materials* (Springer-Verlag, New York, 1994).

Translated by A. Klimontovich

Journal of Experimental and Theoretical Physics, Vol. 92, No. 4, 2001, pp. 561–575.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 4, 2001, pp. 649–664.

ATOMS, SPECTRA, RADIATION

Operator Representation of the Wave Function for a Charged Particle in an External Field and Its Applications

V. M. Katkov and V. M. Strakhovenko*

Budker Institute of Nuclear Physics, Siberian Division, Russian Academy of Sciences, pr. Akademika Lavrent'eva, Novosibirsk, 630090 Russia

**e-mail: v.m.strakhovenko@inp.nsk.su*

Received September 18, 2000

Abstract—A consistent derivation of the operator form for the solution of the wave equation for a charged particle in an arbitrary external electromagnetic field is presented. The expressions obtained can be used for solving any problems in quantum electrodynamics in external fields in the framework of the semiclassical operator method. The peculiarities of the application of this method are demonstrated for the small-angle elastic scattering of a high-energy photon in an arbitrary localized electric field. The problem is solved for the first time without presuming the central symmetry of the external field potential. *© 2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

It is convenient to exactly take into account the effect of an external field on processes in quantum electrodynamics (QED) in the Furry approximation. In this approximation, the series of the perturbation theory in the interaction of quantized fields of charge particles and photons has precisely the same form as for free particles if we replace the free wave functions (as well as Green's functions) by the solutions of the corresponding equations in an external field. The limitations on the application of this approach are due to the fact that the exact solutions of wave equations are known only in several special cases and are quite intricate functions, which hampers subsequent calculations.

These limitations can be removed by using the semiclassical operator method of the solution of QED problems in an external field, which was formulated for the first time in [1]. Later, this method was successfully used for describing the QED processes for many physically interesting configurations of the external field (see, for example, [2] and the literature cited therein). The starting point of the semiclassical operator method is the representation of the solution of the wave equation in the operator form suitable for any external field. Using this form, we obtain the expression for the square of the matrix element characterizing the process in the form of the average of the product of some operators. The next step is the appropriate transformation (regrouping) of the input operators followed by the evaluation of the average. The first two stages of the semiclassical operator method are quite universal when the external field may determine the conditions of applicability of the expressions obtained, but not their form. At the averaging stage, the type of the external field, e.g., the presence of a singularity in the case of a Coulomb potential, is essential. However, in this case also we are dealing with only two possible situations, one of which can be realized in localized potentials with a considerable inhomogeneity (e.g., in the Coulomb field for small impact parameters), and the other takes place in all remaining cases, for example, in a constant external field and even in the Coulomb field for large impact parameters.

In Section 2, the operator form of the solution of the wave equations for a charged particle in an arbitrary external electromagnetic field is derived consistently for the first time and the regrouping of the operators in matrix elements is considered. At each stage, the accuracy of the approximations is estimated. The averaging problem is considered in Section 3 using as an example the elastic scattering of a high-energy photon in a localized electric field. Until now, the problem has not been solved using the semiclassical operator method and is especially interesting from the viewpoint of this approach since both alternative schemes of averaging are realized in it. We have derived the expression for the amplitude of this process, which is valid in the entire range Δ of momentum transfers, which are small as compared to the photon energy ω. The central symmetry of the external field potential is not presumed in this case.

2. OPERATOR FORM OF THE SOLUTION OF WAVE EQUATIONS AND MATRIX ELEMENTS

The wave equation for particles with zero spin is the Klein–Gordon equation

$$\left(i\frac{\partial}{\partial t} - eA_0(\mathbf{r})\right)^2 \psi = \mathcal{H}^2\psi, \quad \mathcal{H} = \sqrt{\mathbf{P}^2 + m^2}, \quad (1)$$

where

$$\mathbf{P} = \mathbf{p} - e\mathbf{A}(\mathbf{r}), \quad \mathbf{p} = -i\frac{\partial}{\partial \mathbf{r}}, \quad A^\mu \equiv (A^0, \mathbf{A}),$$

A^μ being the 4-potential of the external electromagnetic field and e the particle charge. Here and below, we use a system of units in which $\hbar = c = 1$. For the time being, we assume that the 4-potential A^μ is independent of time. If the electric field $\mathbf{E} = -\partial A_0(\mathbf{r})/\partial \mathbf{r}$ is absent, Eq. (1) has two exact operator solutions

$$\psi^{(\pm)} = \exp(\mp i\mathcal{H}t)|0\rangle \tag{2}$$

corresponding to positive- and negative-frequency states. State $|0\rangle$ formally presents the wave function at instant $t = 0$. It should be emphasized that this state is absolutely arbitrary and is chosen in accordance with the problem to be solved. In the presence of an electric field, Eq. (1) can be solved by the method of successive approximations. For this purpose, we present Eq. (1) in the form $L\psi = 0$, where

$$L = \left(i\frac{\partial}{\partial t} + \mathcal{H} - V(\mathbf{r})\right) \times \left(i\frac{\partial}{\partial t} - \mathcal{H} - V(\mathbf{r})\right) + [\mathcal{H}, V(\mathbf{r})]$$

or

$$L = \left(i\frac{\partial}{\partial t} - \mathcal{H} - V(\mathbf{r})\right) \times \left(i\frac{\partial}{\partial t} + \mathcal{H} - V(\mathbf{r})\right) - [\mathcal{H}, V(\mathbf{r})].$$

Here, $V(\mathbf{r})$ denotes the potential energy of the particle: $V(\mathbf{r}) \equiv eA_0(\mathbf{r})$. If we now disregard commutator $[\mathcal{H}, V(\mathbf{r})]$ in the expression for L, we obtain as the zeroth approximation the following two solutions of Eq. (1) in the presence of an electric field:

$$\psi_0^{(\pm)} = \exp[\mp i(\mathcal{H} \pm V(\mathbf{r}))t]|0\rangle. \tag{3}$$

The next approximation will be sought in the form

$$\psi_1^{(\pm)}(t) = (1 + C^{(\pm)})\psi_0^{(\pm)}(t).$$

Equation $L\psi = 0$ gives $C^{(\pm)} = \pm V/(2\mathcal{H})$ in the first approximation, where we disregarded terms of the order of $[\mathcal{H}^{-1}, V]$. If we nevertheless use $\psi_0^{(\pm)}(t)$ in applications, the characteristic value of the correction of the order of V/ε (ε is the particle energy) determines the accuracy of solution (3). The specific estimate of this accuracy depends on the properties of potential $V(\mathbf{r})$ as well as on the nature of the process itself, which is described using the solution obtained. For example, in the case of elastic scattering of a photon with energy $\omega \gg m$ in the field of a nucleus having charge $Z|e|$ (e is the electron charge), the energy of virtual particles amounts to $\varepsilon \sim \omega$, and $|V|/\varepsilon \sim Z\alpha/(\rho\omega)$, where $\alpha = e^2 = 1/137$ is the fine-structure constant and ρ is the impact parameter. Since in this process the impact parameters satisfying the condition $\omega/m^2 \geq \rho \geq 1/(m + \Delta)$ are essential, we obtain $|V|/\varepsilon \leq Z\alpha(m + \Delta)/\omega$; i.e., $|V|/\varepsilon \ll 1$ for small-angle scattering ($\Delta \ll \omega$).

For particles with a spin of 1/2, we must solve the Dirac equation, which in the external field has the form

$$i\frac{\partial}{\partial t}\psi_D = H_D\psi_D, \quad H_D = \boldsymbol{\alpha}\cdot\mathbf{P} + \beta m + V(\mathbf{r}). \tag{4}$$

We will use the standard representation of matrices γ^μ: $\beta = \gamma^0$ and $\boldsymbol{\alpha} = \gamma^0\boldsymbol{\gamma}$. The operator solution of Eq. (4) is obvious,

$$\psi_D(t) = \exp(-iH_Dt)\psi_D(0), \tag{5}$$

but it does not suit us for several reasons. First, it would be interesting to obtain a solution which evolves in time with the classical relativistic Hamiltonian to within the spin terms, as was the case for scalar particles (see Eq. (3)). Second, it would be desirable to separate the positive- and negative-frequency states. For a free motion ($A^\mu = 0$), this can done by using the Foldy–Wouthuysen transformation (subscript *FW*, while subscript *D* denotes the initial (Dirac) representation), which removes the odd matrices $\boldsymbol{\alpha}$ mixing the upper and lower two-component spaces of the wave function from Hamiltonian H_D.

As for scalar particles, we will first consider the case of zero electric field; i.e., when $H_D = \boldsymbol{\alpha}\cdot\mathbf{P} + \beta m$. We carry out the transformation $\psi_D = U^{-1}\psi_{FW}$, where

$$U = R\left(1 + \frac{\boldsymbol{\gamma}\cdot\mathbf{P}}{\tilde{H} + m}\right), \quad \tilde{H} = ((\boldsymbol{\alpha}\cdot\mathbf{P})^2 + m^2)^{1/2}, \quad R = \left(\frac{\tilde{H} + m}{2\tilde{H}}\right)^{1/2}, \tag{6}$$

has the same form as under the Foldy–Wouthuysen transformation for free particles if we substitute $\mathbf{P}$ for $\mathbf{p}$ in the latter transformation. It should be noted that $[\boldsymbol{\gamma}\cdot\mathbf{P}, \tilde{H}] = 0$ since $(\boldsymbol{\alpha}\cdot\mathbf{P})^2 = -(\boldsymbol{\gamma}\cdot\mathbf{P})^2$. It can also be easily verified that U is a unitary operator: $U^+ = U^{-1}$. As a result of this transformation, Eq. (4) acquires the form

$$i\frac{\partial}{\partial t}\psi_{FW} = H_{FW}\psi_{FW}, \quad H_{FW} = UH_DU^{-1} = \beta\tilde{H}. \tag{7}$$

The solution of Eq. (7) has the form

$$\psi_{FW}(t) = \exp(-i\beta\tilde{H}t)\psi_{FW}(0). \tag{8}$$

Operator $\beta\tilde{H}$ is even; i.e., it does not mix the upper and lower two-component spinors of the wave functions.

Let us now choose for $\psi_{FW}(0)$ the eigenfunctions of matrix β:

$$\beta\psi_{FW}^{(\pm)}(0) = \pm\psi_{FW}^{(\pm)}(0),$$
$$\psi_{FW}^{(+)}(0) = \begin{pmatrix} \varphi \\ 0 \end{pmatrix}|0\rangle, \quad \psi_{FW}^{(-)}(0) = \begin{pmatrix} 0 \\ \chi \end{pmatrix}|0\rangle, \tag{9}$$

where state $|0\rangle$ has the same meaning as for scalar particles, and φ and χ are two-component spinors describing the polarization of particles at instant $t = 0$. The functions

$$\psi_{FW}^{(\pm)}(t) = \exp(-i\beta\tilde{H}t)\psi_{FW}^{(\pm)}(0) = \exp(\mp i\tilde{H}t)\psi_{FW}^{(\pm)}(0)$$

are obviously positive- (+) and negative- (–) frequency states; i.e., these are the states of electrons and positrons, respectively. Finally, for $\psi_D^{(\pm)} = U^{-1}\psi_{FW}^{(\pm)}$ in a magnetic field, we obtain

$$\psi_D^{(+)}(t) = \left(\frac{\tilde{H}+m}{2\tilde{H}}\right)^{1/2}\left(1 - \frac{\boldsymbol{\gamma}\cdot\mathbf{P}}{\tilde{H}+m}\right)\exp(-i\tilde{H}t)\begin{pmatrix} \varphi \\ 0 \end{pmatrix}|0\rangle,$$
$$\psi_D^{(-)}(t) = \left(\frac{\tilde{H}+m}{2\tilde{H}}\right)^{1/2}\left(1 - \frac{\boldsymbol{\gamma}\cdot\mathbf{P}}{\tilde{H}+m}\right)\exp(i\tilde{H}t)\begin{pmatrix} 0 \\ \chi \end{pmatrix}|0\rangle. \tag{10}$$

Expressions (10) are the exact operator solution of Eq. (4) for $V(\mathbf{r}) = 0$. In the Hamiltonian

$$\tilde{H} = ((\boldsymbol{\alpha}\cdot\mathbf{P})^2 + m^2)^{1/2} = (\mathbf{P}^2 + m^2 - e\mathbf{H}\cdot\boldsymbol{\Sigma})^{1/2},$$

where $\mathbf{H}(\mathbf{r})$ is the magnetic field and $\boldsymbol{\Sigma} = -\gamma^5\boldsymbol{\alpha}$, we can separate the spin term if $|e\mathbf{H}| \ll \varepsilon^2$. Then

$$\tilde{H} \approx \mathcal{H} - \frac{e\mathbf{H}(\mathbf{r})\cdot\boldsymbol{\Sigma}}{2\mathcal{H}}, \quad \mathcal{H} = \sqrt{\mathbf{P}^2 + m^2}. \tag{11}$$

In the expansion of Hamiltonian $\tilde{H}$, we have disregarded terms of the order of $[\mathcal{H}^{-1}, e\mathbf{H}\cdot\boldsymbol{\Sigma}]$, and the sequence of operators $\mathcal{H}^{-1}$ and $e\mathbf{H}\cdot\boldsymbol{\Sigma}$ in Eq. (11) is immaterial for this degree of accuracy. The relative value of the terms discarded as compared to $e\mathbf{H}\cdot\boldsymbol{\Sigma}\mathcal{H}^{-1}$ amounts to $v(a\varepsilon)^{-1}$, where a is the characteristic size of the nonuniformity of field $\mathbf{H}$ and v is the particle velocity. Note that the terms omitted while determining the correction to solution (3) have exactly the same magnitude relative to correction $C^{(\pm)}$ itself if a is treated as the size of the inhomogeneity of potential $V(\mathbf{r})$. Thus, if the conditions $|e\mathbf{H}| \ll \varepsilon^2$ and $a \gg v/\varepsilon$ are satisfied, we can replace $\tilde{H}$ in Eq. (10) by the approximate value (11). If $|e\mathbf{H}| \ll m\varepsilon$, we can additionally substitute $\tilde{H}$ for $\mathcal{H}$ in Eq. (10) everywhere except in the exponent $\exp(\pm i\tilde{H}t)$.

Let us now suppose that $V(\mathbf{r}) \neq 0$. As a result of the Foldy–Wouthuysen transformation with matrix U from Eq. (6), we obtain the following expression for the Hamiltonian $H_{FW}^{(1)} = UH_DU^{-1}$:

$$H_{FW}^{(1)} = \beta\tilde{H} + V(\mathbf{r}) + \frac{1}{2}[[R, V(\mathbf{r})], R^{-1}]$$
$$+ R\left(\frac{1}{2}[Q, [Q, V(\mathbf{r})]] + [Q, V(\mathbf{r})]\right)R, \tag{12}$$

where $Q = (\boldsymbol{\gamma}\cdot\mathbf{P})/(\tilde{H} + m)$ and operator R is defined in Eq. (6) so that $U = R(1 + Q)$. The only odd operator in $H_{FW}^{(1)}$ is the term $R[Q, V(\mathbf{r})]R$, which is linear in Q and generally small as compared to $V(\mathbf{r})$ like the remaining terms in Eq. (12) containing commutators. The value of the odd term in the Hamiltonian can be made even smaller by carrying out one more transformation: $H_{FW}^{(2)} = H_{FW}^{(1)} + [B, H_{FW}^{(1)}]$. Choosing

$$B = \frac{1}{2}\beta R[Q, V(\mathbf{r})]R\tilde{H}^{-1}$$

and disregarding the terms quadratic in potential, we obtain

$$H_{FW}^{(2)} = \beta\tilde{H} + V(\mathbf{r}) + \frac{1}{2}[[R, V(\mathbf{r})], R^{-1}]$$
$$+ \frac{1}{2}R([Q, [Q, V(\mathbf{r})]] + [[Q, V(\mathbf{r})], \tilde{H}]\tilde{H}^{-1})R. \tag{13}$$

The terms quadratic in potential, which have been omitted while deriving this relation, are of the order of $(\partial V(\mathbf{r})/\partial\mathbf{r})^2/\varepsilon^3$ (in a purely electric field). For example, in the Coulomb potential $V_C(\mathbf{r}) = -Z\alpha/r$, their magnitude in the units of the potential itself is $Z\alpha/(\varepsilon r)^3$. For $Z\alpha \sim 1$, this approximation is possible if $\varepsilon\rho \gg 1$, where ρ is the characteristic size (e.g., impact parameter) of the problem. The order of magnitude of the term $[[R, V(\mathbf{r})], R^{-1}]$ in Eq. (13) is $V(\mathbf{r})(mv/\varepsilon^2 r)^2$. An estimate of the odd term in Eq.(13) is $v^2V(\mathbf{r})/(\varepsilon r)^2$. The term $(1/2)R[Q, [Q, V(\mathbf{r})]]R$ can be written in the form

$$\frac{1}{2}R\left(\left[\frac{1}{\tilde{H}+m}, [V(\mathbf{r}), \tilde{H}]\right]\right.$$
$$\left. + \frac{1}{\tilde{H}+m}(\Delta V(\mathbf{r}) + 2e([\mathbf{P}\times\mathbf{E}]\cdot\boldsymbol{\Sigma}))\frac{1}{\tilde{H}+m}\right)R. \tag{14}$$

In this expression, the double commutator has the same order of magnitude as the odd term in Eq. (13). A formal estimate of the value of the term proportional to ΔV is $V(\mathbf{r})/(\varepsilon r)^2$. This term is small and will henceforth be omitted. It should be borne in mind, however, that while calculating the fine structure of the hydrogen atom with the help of Hamiltonian $H_{FW}^{(2)}$, we must retain this term since its smallness in the given problem is of the order of the sought correction to the energy levels. Finally, the term proportional to $([\mathbf{P}\times\mathbf{E}]\cdot\boldsymbol{\Sigma})$ in

Eq. (14) corresponds to the spin–orbit interaction and is of the order of $\nu V(\mathbf{r})/(\varepsilon r)$. Omitting now all terms in Eq. (13) which are of the order of $V(\mathbf{r})/(\varepsilon r)^2$ or smaller and expanding $\tilde{H}$ as it was done in Eq. (11), we obtain Hamiltonian H_{FW} in the form which will be used below:

$$H_{FW} = \beta \times \left[\mathscr{H} + \beta V(\mathbf{r}) - \frac{e}{2\mathscr{H}}\left(\left(\mathbf{H} + \beta\frac{[\mathbf{E}+\mathbf{P}]}{\mathscr{H}+m}\right)\cdot\mathbf{\Sigma}\right)\right]. \quad (15)$$

The condition $\varepsilon\rho \gg 1$ used while deriving Eq. (15) indicates (in the case of relativistic motion) that the wavelength $1/\varepsilon$ associated with a particle is much smaller than the characteristic scales of the problem such as the typical impact parameter or the inhomogeneity scale (cf. the estimate for the applicability of expansion (11)). This condition essentially coincides with the general condition of the applicability of the semiclassical approximation.

The equation for function $\psi_{FW}(t)$ has the form (7) with the Hamiltonian H_{FW} defined in Eq. (15), and its solution coincides with Eq. (8) if we substitute H_{FW} for $\beta\tilde{H}$ in it. As before, we choose for $\psi_{FW}(0)$ the eigenfunctions of operator β defined in Eq. (9). Then the positive- and negative-frequency states are given by

$$\psi_{FW}^{(\pm)}(t) = \exp(\mp iH^{(\pm)}t)\psi_{FW}^{(\pm)}(0),$$
$$H^{(\pm)} = \mathscr{H} \pm V(\mathbf{r}) - \frac{e}{2\mathscr{H}}\left(\left(\mathbf{H} \pm \frac{[\mathbf{E}\times\mathbf{P}]}{\mathscr{H}+m}\right)\cdot\mathbf{\Sigma}\right). \quad (16)$$

It should be noted that $H^{(+)}$ differs from $H^{(-)}$ in the sign reversal in potential $V(\mathbf{r})$. The resultant Foldy–Wouthuysen transformation has the form $U^{(1)} = (1 + B)U$. Since $B \sim |eE|/\varepsilon^2 \sim (\varepsilon r)^{-2}$ (the estimate obtained for a purely electric field) has the same smallness as the terms which have already been discarded while deriving Eq. (16), we can substitute U for $U^{(1)}$.

States $\psi_{FW}(t)$, as well as $\psi_D(t) = U^{-1}\psi_{FW}(t)$, are stationary if $\psi_{FW}(0)$ is an eigenfunction of the Hamiltonian: $H_{FW}\psi_{FW}(0) = \varepsilon\psi_{FW}(0)$. Among other things, we have verified that in the Coulomb potential, the function $\psi_D^{(+)}$ obtained from this equation (after the application of transformation U^{-1}) coincides with the Furry solution [3] (see [4, Eq. (39.10)]). The required accuracy in this case is ensured by retaining in Eq. (16) generally small spin terms. Naturally, these terms are important if we consider the time evolution of spin states. It will be shown below, however, that these terms can be neglected in the calculation of matrix elements if the spin varies insignificantly during the formation time of the process, which is usually observed in the ultrarelativistic case ($\varepsilon \gg m$). Besides, for $\varepsilon \gg m$ and $|e\mathbf{H}| \ll \varepsilon m$, we can simplify the transformation matrix U by expanding R and $\tilde{H}$. Finally, while describing radiative processes (the emission of a photon or the creation of a e^+e^- pair) involving ultrarelativistic particles with a spin 1/2, we will use the following operator representation for semiclassical wave functions:

$$|\psi^{(+)}\rangle = \frac{1}{\sqrt{2}}\left(1 - \frac{\boldsymbol{\gamma}\cdot\mathbf{P}}{\mathscr{H}+m}\right) \times \exp\{-i(\mathscr{H} + V(\mathbf{r}))t\}\begin{pmatrix}\varphi\\0\end{pmatrix}|0\rangle,$$
$$|\psi^{(-)}\rangle = \frac{1}{\sqrt{2}}\left(1 - \frac{\boldsymbol{\gamma}\cdot\mathbf{P}}{\mathscr{H}+m}\right) \times \exp\{i(\mathscr{H} - V(\mathbf{r}))t\}\begin{pmatrix}0\\\chi\end{pmatrix}|0\rangle. \quad (17)$$

Here, $|\psi^{\pm}\rangle$ indicates $\psi_D^{\pm}$ in the notation used above. In expressions (17), the representation of the wave function (coordinate, momentum, or other) has not been specified as yet. In order to go over to a specific representation, we must project $|\psi\rangle$ onto the corresponding eigenfunction. For example, in the coordinate representation, $\psi(\mathbf{r}, t) = \langle\mathbf{r}|\psi\rangle$, where $|\mathbf{r}\rangle$ is the eigenfunction of the coordinate operator.

Until now, we assumed that the 4-potential A^μ of the external field is independent of time. Otherwise, Hamiltonian H_D in the initial equation (4) is also a function of time, $H_D \equiv H_D(t)$, and instead of Eq. (5), we obtain the operator solution for Eq. (4) in the form

$$\psi_D(t) = [\vartheta(t)T_+ + \vartheta(-t)T_-] \times \exp\left\{-i\int_0^t ds H_D(s)\right\}\psi_D(0), \quad (18)$$

where $\vartheta(t) = 1$ for $t > 0$ and $\vartheta(t) = 0$ for $t < 0$. Symbol $T_\pm$ in Eq. (18) indicates that operators must be ordered in time. Namely, for T_+, the operators corresponding to earlier instants must be on the right of the operators corresponding to later instants (for T_-, the converse is true). Carrying out the Foldy–Wouthuysen transformation $\psi_D = U^{-1}\psi_{FW}$, where operator U is defined in Eq. (6) and may also be a function of time if the vector potential $\mathbf{A}$ depends on it, we arrive at Eq. (7) with Hamiltonian $H_{FW}^{(1)}(t)$ in the form

$$H_{FW}^{(1)}(t) = UH_D(t)U^{-1} + \frac{i}{2}\left(\frac{\partial U}{\partial t}U^+ - U\frac{\partial U^+}{\partial t}\right). \quad (19)$$

Term $UH_D(t)U^{-1}$ coincides with Eq. (12), where $V(\mathbf{r})$ can now be a function of time also: $V(\mathbf{r}) \longrightarrow V(\mathbf{r}, t)$.

Correct to the terms of the order of ε^{-2}, which will be omitted as before, the second term in Eq. (19) is given by

$$\begin{aligned}&\frac{i}{2}\left(\frac{\partial U}{\partial t}U^{+}-U\frac{\partial U^{+}}{\partial t}\right)\\ &\approx-\frac{e\left(\left[\mathbf{P}\times\frac{\partial\mathbf{A}}{\partial t}\right]\cdot\boldsymbol{\Sigma}\right)}{2\mathcal{H}(\mathcal{H}+m)}-\frac{ie}{2\mathcal{H}}(\boldsymbol{\gamma}\cdot\mathbf{A}).\end{aligned}\tag{20}$$

The term in Eq. (20) proportional to $\boldsymbol{\gamma}\cdot\mathbf{A}$ is an odd operator having the same order of magnitude as the odd operator in Eq. (12). Modifying operator B in the additional transformation $H_{FW}^{(2)}=(1+B)H_{FW}^{(1)}(1+B)^{-1}$, we can simultaneously decrease the magnitude of all odd terms in Eq. (19) down to values of the order of ε^{-2}, which are omitted in the final result. It should be emphasized that with the adopted accuracy, the role of the second transformation (transition from $H_{FW}^{(1)}$ to $H_{FW}^{(2)}$) is effectively reduced to the removal of odd operators from $H_{FW}^{(1)}$, while operator B itself can be omitted in the expression $\psi_D^{(+)}=U^{-1}(1+B)^{-1}\psi_{FW}(t)$ in view of its smallness. Finally, we obtain the operator solution of the Dirac equation in a time-dependent electromagnetic field in the semiclassical approximation:

$$\begin{aligned}\psi^{(\pm)}(t)&=U^{-1}[\vartheta(t)T_{+}+\vartheta(t)T_{-}]\\ &\times\exp\left\{\mp i\int_0^t ds\,H^{(\pm)}(s)\right\}\psi_{FW}^{(\pm)}(0).\end{aligned}\tag{21}$$

Here, operator U *is* defined in Eq. (6), functions $\psi_{FW}^{(\pm)}(0)$ in Eq. (9), and Hamiltonian $H^{(\pm)}(t)$ in Eq. (16), where the potential energy $V(\mathbf{r},t)$, the vector potential $\mathbf{A}(\mathbf{r},t)$, and hence fields $\mathbf{E}$ and $\mathbf{H}$ are functions of time. It should be noted that taking into account the contribution from the first term in formula (20), $H^{(\pm)}$ now contains vector $-\partial A_0/\partial\mathbf{r}-\partial\mathbf{A}/\partial t$ instead of $\mathbf{E}$, i.e., the correct expression for the electric field in the case of a time-dependent potential.

Using as examples the simplest processes, we will now write the corresponding matrix elements in the semiclassical operator method using the operator solutions of wave equations obtained above. We consider only particles with a spin of 1/2 since the crucial aspects in the application of this method for spinor and scalar particles are the same. In the Furry representation, a matrix element corresponding to the emission of a photon by an electron to within the normalization factor has the form

$$V_{if}^{\mathrm{rad}}=ie\int d^4x\,\bar{\psi}_f^{(+)}(x)\hat{e}^{*}\exp(ikx)\psi_i^{(+)}(x),\tag{22}$$

where e is the electron charge, k^{μ} and e^{μ} are the 4-vectors of the photon momentum and polarization, and $\hat{e}^{*}\equiv\gamma^{\mu}e_{\mu}^{*}$. We substitute into Eq. (22) the solution $\psi^{(+)}(x)=\psi^{(+)}(\mathbf{r},t)=\langle\mathbf{r}|\psi^{(+)}\rangle$, where $|\psi^{(+)}\rangle$ is defined by formula (17), and introduce $\exp(ikx)$ into $\langle\mathbf{r}|$ in

$$\begin{aligned}\exp(ikx)\psi^{(+)}(\mathbf{r},t)&=\exp(ikx)\langle\mathbf{r}|\psi^{(+)}\rangle\\ &=\langle\mathbf{r}|\exp(ikx)\psi^{(+)}\rangle.\end{aligned}$$

Then the integral over $\mathbf{r}$ is reduced to $\int d^3r|\mathbf{r}\rangle\langle\mathbf{r}|$ and is equal to the unit operator, which expresses the condition of the completeness of the system of states $|\mathbf{r}\rangle$. As a result, we obtain the following expression for V_{if}^{rad} at the gauge $e^0=0$:

$$\begin{aligned}V_{if}^{\mathrm{rad}}&=-i\frac{e}{2}\int_{-\infty}^{\infty}dt\langle f|(\varphi_f^{+},0)C(t)\begin{pmatrix}\varphi_i\\0\end{pmatrix}|i\rangle,\\ C(t)&=\exp(iH_et)\left(1+\frac{\boldsymbol{\gamma}\cdot\mathbf{P}}{\mathcal{H}+m}\right)(\boldsymbol{\alpha}\cdot\mathbf{e})\\ &\times\exp(ikx)\left(1-\frac{\boldsymbol{\gamma}\cdot\mathbf{P}}{\mathcal{H}+m}\right)\exp(-iH_et),\end{aligned}\tag{23}$$

where $H_e=\mathcal{H}+V(\mathbf{r})$, and the notation $|i\rangle$ is introduced in $|\psi_i^{(+)}\rangle$ and $|f\rangle$ in $|\psi_f^{(+)}\rangle$ for the state $|0\rangle$. Owing to the fact that the wave functions actually have two components, the nonzero contribution to Eq. (23) comes only from the even component of operator $C(t)$. Moving additionally operator $\exp(ikx)$ in $C(t)$ to the left, we find to the relativistic accuracy ($m/\varepsilon\ll1$) that

$$\begin{aligned}V_{if}^{\mathrm{rad}}&=-i\frac{e}{2}\int_{-\infty}^{\infty}dt\langle f|\exp(ikx)\exp(iH_e't)\\ &\times\varphi_f^{+}(a_r+i\mathbf{b}_r\cdot\boldsymbol{\sigma})\varphi_i\exp(-iH_et)|i\rangle,\\ a_r&=\left(2+\frac{\omega}{\mathcal{H}'}\right)\mathbf{e}^{*}\cdot\mathbf{v},\\ \mathbf{b}_r&=\frac{\omega}{\mathcal{H}'}\left[\mathbf{v}\left(1-\frac{m}{\mathcal{H}}\right)-\mathbf{n},\mathbf{e}^{*}\right].\end{aligned}\tag{24}$$

Here, $\boldsymbol{\sigma}$ are the Pauli matrices, $\mathcal{H}'\equiv\mathcal{H}(\mathbf{P}\longrightarrow\mathbf{P}-\mathbf{k})$, $H_e'=\mathcal{H}'+V(\mathbf{r})$, and $\mathbf{v}=\mathbf{P}/\mathcal{H}$. We have neglected the noncommutativity of the operators appearing in a_r and $\boldsymbol{b}_r$ since its inclusion would lead to corrections of the order of $|e\mathbf{H}|/\varepsilon^2$, and the terms of such an order of magnitude were omitted even in the solution of the wave equation. If we now move the operator $\exp(-iH_et)$ to the left to $\exp(iH_e't)$, the integrand assumes the form

$$\langle f|\exp(-i\mathbf{k}\cdot\mathbf{r})L_r(t)\varphi_f^{+}(a_r(t)+i\mathbf{b}_r(t)\cdot\boldsymbol{\sigma})\varphi_i|i\rangle,\tag{25}$$

where

$$L_r(t) = \exp\{i(H_e' + \omega)t\}\exp\{-iH_e t\},$$

and $a_r(t)$ and $\mathbf{b}_r(t)$ are Heisenberg operators: $a(t) = \exp(iH_e t)a\exp(-iH_e t)$. It should be noted that if we retain the spin terms in H_e (16), operators $\boldsymbol{\sigma}$ in Eq. (25) would also be transformed into Heisenberg operators. The expression for the square of the matrix element would contain the operators $\boldsymbol{\sigma}(t_1)$ and $\boldsymbol{\sigma}(t_2)$, which are functions of different times:

$$\sum_f \left|V_{if}^{\mathrm{rad}}\right|^2 = \frac{\alpha}{4}\int_{-\infty}^{\infty} dt_1 dt_2 \times \langle i|\varphi_i^+[a_r^+(t_2) - i\mathbf{b}_r^+(t_2)\cdot\boldsymbol{\sigma}(t_2)]L_r^+(t_2) \times L_r(t_1)[a_r(t_1) + i\mathbf{b}_r(t_1)\cdot\boldsymbol{\sigma}(t_1)]\varphi_i|i\rangle. \quad (26)$$

The correction emerging in the matrix element when the time dependence of the spin operators is taken into account has the relative value

$$\delta \sim \frac{e}{\varepsilon}\int_{t_1}^{t_2} dt[\mathbf{H}(t) + [\mathbf{E}(t)\times\mathbf{v}(t)]],$$

where we can roughly assume that the time dependence of the fields is taken on the classical trajectory; i.e., we have, for example $\mathbf{H}(t) = \mathbf{H}(\mathbf{r}_{cl}(t))$. The quantity δ has the meaning of the angle of rotation of the spin vector over the characteristic time of process formation. In accordance with the equations of motion, it coincides with the angle of rotation of the velocity vector during the same time. For instance (see, for example, [2]), for photons emitted with frequencies corresponding to the maximum intensity of magnetic bremsstrahlung in a uniform magnetic field, this angle is small and $\delta \sim m/\varepsilon \ll 1$. For a photon passing by a nucleus with an impact parameter ρ, we have $\delta \sim Z\alpha/(\rho\varepsilon) \ll 1$ for $\rho \gg 1/\varepsilon$.

The next step in the application of the semiclassical operator method is the combining of the two components constituting operator $L_r(t)$ (25) into one component. Differentiating $L_r(t)$ with respect to time, we obtain

$$\frac{dL_r(t)}{dt} = iL_r(t)B(t), \quad B(t) = \mathscr{H}(\mathbf{P}(t) - \mathbf{k}) + \omega - \mathscr{H}(\mathbf{P}(t)). \quad (27)$$

It should be noted that the potential energy $V(\mathbf{r})$ appears in Eq. (27) only implicitly through the time dependence of the Heisenberg operator $\mathbf{P}(t)$. Relations (27) lead to

$$L_r(t) = [\vartheta(t)T_- + \vartheta(-t)T_+]\exp\left\{i\int_0^t ds B(s)\right\}. \quad (28)$$

The meaning of the combination ("disentanglement") of the exponents in operator $L_r(t)$ is that "large" (high-frequency) operators appearing in each exponent cancel out in combination B. Indeed,

$$B(t) = \sqrt{(\mathbf{P}(t) - \mathbf{k})^2 + m^2} + \omega - \mathscr{H}(t) = \sqrt{(\mathscr{H}(t) - \omega)^2 + 2kP(t)} - (\mathscr{H}(t) - \omega) \approx \frac{kP(t)}{\mathscr{H}(t) - \omega} \approx \frac{\omega[m^2 + \mathbf{P}_\perp^2(t)]}{2\mathscr{H}(t)(\mathscr{H}(t) - \omega)}, \quad (29)$$

where

$$\mathbf{P}_\perp = \mathbf{P} - \mathbf{n}(\mathbf{n}\cdot\mathbf{P}), \quad \mathbf{n} = \mathbf{k}/\omega.$$

Expansion (29) is valid if operator $(kP)/(\mathscr{H}(t) - \omega)^2$ is small. In the ultrarelativistic case, its order of magnitude is $m^2/(\varepsilon - \omega)^2$; in this case, relation (29) is valid for $\varepsilon - \omega \gg m$. In such estimates, we replace operators by their characteristic values defined by the state $|i\rangle$ over which the averaging in Eq. (26) is carried out. For $\omega > \varepsilon$, there is no compensation in Eq. (29), and in relation (26) we are dealing with rapidly oscillating functions of time, whose integration leads to a negligibly small result. Product $L_r^+(t_2)L_r(t_1)$ in expression (26) can be transformed using the definition $L_r(t) = \exp\{i(H_e' + \omega)t\}\exp(-iH_e t)$ and solution (28):

$$L_r^+(t_2)L_r(t_1) = \exp(iH_e t)\exp(iH_e\tau/2) \times \exp\{-i(H_e' + \omega)\tau\}\exp(iH_e\tau/2)\exp(-iH_e t) = \exp(iH_e t)L_r^+(\tau/2)L_r(-\tau/2)\exp(-iH_e t) = [\vartheta(\tau)T_+ + \vartheta(-\tau)T_-]\exp\left\{i\int_{-\tau/2}^{\tau/2} ds B(s+t)\right\}, \quad (30)$$

where $\tau = t_2 - t_1$ and $t = (t_1 + t_2)/2$.

The method for calculating the average $\langle i|\ldots|i\rangle$ in expressions of type (26) depends on the value of commutators of the input operators. If all these commutators can be neglected, we choose for $|i\rangle$ the corresponding wave packets and simply replace the operators by their mean values in state $|i\rangle$. The packet width in this case must be much smaller than the mean value of an operator (e.g., of momentum) and much larger than the uncertainty associated with the discarded commutators. In other words, averaging in this case is reduced to the replacement of Heisenberg operators by the corresponding classical quantities, i.e., to the transition to classical trajectories. Since the argument of the exponent in Eq. (26) (the quantity $B(s + t)$ in Eq. (30)) is proportional to $m^2 + \mathbf{P}_\perp^2$, the contribution comes from $|\mathbf{P}_\perp| \le m$; this will be used in subsequent estimates. For this reason in particular, we can disregard to the relativistic accuracy in the transverse momentum in the expression for $\mathscr{H} = \sqrt{\mathbf{P}^2 + m^2}$. With this accuracy, we can single out three types of commutators in Eq. (26):

(1) the commutators of the components of $\mathbf{P}_\perp$ in the preexponential factor; (2) the commutators of these components with the exponential (with operator (30)); and (3) the commutators characterizing the accuracy of the transition in Eq. (30) from the chronological product to an ordinary exponential. It is remarkable that the order of magnitude of the ratio of each of these commutators to the product of operators themselves (which in fact characterizes the possibility of discarding the noncommutativity) is the same. In specific estimates, it is convenient to use the following expression for $\mathbf{P}_\perp(t)$, which is valid to the relativistic accuracy:

$$\mathbf{P}_\perp(t) = \mathbf{P}_\perp(0) + e\int_0^t ds[\mathbf{E}_\perp(\mathbf{r}(s)) + [\mathbf{n}\times\mathbf{H}(\mathbf{r}(s))]], \tag{31}$$

where $\mathbf{P}_\perp(0)$ is the Schrödinger operator of the momentum transverse to $\mathbf{n} = \mathbf{k}/\omega$. The conditions ensuring the transition to classical trajectories naturally depend on the form of the external field. For example, in a uniform magnetic field, the condition $|e\mathbf{n}\times\mathbf{H}|/m^2 \equiv |\mathbf{n}\times\mathbf{H}|/H_{cr} \ll 1$ must be satisfied, while in the Coulomb field, we must have $Z\alpha \ll (m\rho)^2$ or $\rho \gg 1/m$ (under the assumption that $Z\alpha \sim 1$). It should be noted at the very outset that for higher order processes, the argument of the exponential may contain, apart from mass, the total momentum transfer; the estimate of the characteristic value of $|\mathbf{P}_\perp|$ may change accordingly. We will return to this problem in the next section, where the method for calculating the average $\langle i|\ldots|i\rangle$ will be formulated for the case when a transition to classical trajectories is ruled out.

Proceeding in the same way as in the derivation of relations (24), we obtain the following expression for the matrix element corresponding to the creation of an e^+e^- pair by a photon having momentum k^μ and polarization e^μ (in the same normalization as in Eq. (22)):

$$V_{if}^{\text{pair}}(\mathbf{k},\mathbf{e}) = -e\int_{-\infty}^{\infty} dt V_{if}(\mathbf{k},\mathbf{e},t),$$
$$V_{if}(\mathbf{k},\mathbf{e},t) = \langle f|\varphi_f^+ R(\mathbf{k},\mathbf{e},t)\chi_i L_p(\mathbf{k},t)\exp(i\mathbf{k}\cdot\mathbf{r})|i\rangle. \tag{32}$$

Here, φ_f and χ_i are the two-component spinors from Eq. (17), states $|0\rangle$ in Eq. (17) are replaced by $|f\rangle$ for electrons and by $|i\rangle$ for positrons, and operator $R(\mathbf{k},\mathbf{e},t)$ with the relativistic accuracy has the form

$$R(\mathbf{k},\mathbf{e},t) = \exp(iH_e t)R(\mathbf{k},\mathbf{e})\exp(-iH_e t),$$
$$R(\mathbf{k},\mathbf{e}) = \frac{\omega}{2\mathcal{H}\mathcal{H}'} \times\left[([\mathbf{e}\times\mathbf{P}]\cdot\mathbf{n}) + i(\boldsymbol{\sigma}\cdot(\mathbf{e}m + \mathbf{n}(\mathbf{e}\cdot\mathbf{P})))\frac{2\mathcal{H}'-\omega}{\omega}\right], \tag{33}$$

where, as in Eq. (24), $H_e = \mathcal{H} + V(\mathbf{r})$ and $\mathcal{H}' = \mathcal{H}(\mathbf{P}\longrightarrow\mathbf{P}-\mathbf{k})$. Operator $L_p(\mathbf{k},t)$ in Eq. (32) is the product of two exponential functions,

$$L_p(\mathbf{k},t) = \exp(iH_e t)\exp\{i(\mathcal{H}' - V(\mathbf{r}) - \omega)t\},$$

which (cf. Eq. (28)) can be combined into one:

$$L_p(\mathbf{k},t) = [\vartheta(t)T_+ + \vartheta(-t)T_-]\exp\left\{i\int_0^t ds B_p(s)\right\},$$
$$B_p(s) = \exp(iH_e s)[\mathcal{H}+\mathcal{H}'-\omega]\exp(-iH_e s) = \mathcal{H}(s) + \mathcal{H}'(s) - \omega. \tag{34}$$

The compensation of "large" terms in $B_p(t)$ now takes place for positive values of operator $\omega - \mathcal{H}(t)$ ($\omega > \varepsilon$). Accordingly, an approximate expression for $B_p(t)$ in the compensation region, which is the only region contributing to Eq. (32), appears as Eq. (29) if we make the substitution $\mathcal{H}(t) - \omega \longrightarrow \omega - \mathcal{H}(t)$ in it:

$$B_p(t) \approx \frac{kP(t)}{\omega - \mathcal{H}(t)} \approx \frac{\omega[m^2 + \mathbf{P}_\perp^2(t)]}{2\mathcal{H}(t)(\omega - \mathcal{H}(t))}. \tag{35}$$

While deriving this expression, we assumed that $\varepsilon \gg m$ and $\omega - \varepsilon \gg m$.

Until now, we assumed that a photon is real ($k^2 = 0$). Higher order diagrams may contain vertices corresponding to the interaction with virtual ($k^2 \neq 0$) photons. Expressions (24) and (32) remain valid in this case also since we used the condition $k^2 = 0$ at a later stage (in the expansion of operators $B(t)$ and $B_p(t)$ in the compensation region). Let us refine this expansion for $k^2 \neq 0$. Considering that

$$\mathcal{H}' \equiv \sqrt{(\mathbf{P}-\mathbf{k})^2 + m^2} = \sqrt{(\mathcal{H}-\omega)^2 + 2kP + k^2},$$

we obtain

$$B(t) \approx \frac{kP(t) - k^2/2}{\mathcal{H}(t) - \omega} \approx \frac{\omega[m^2 + \mathbf{P}_\perp^2(t)]}{2\mathcal{H}(t)(\mathcal{H}(t)-\omega)} + \frac{k^2}{2\omega},$$
$$B_p(t) \approx \frac{kP(t) - k^2/2}{\omega - \mathcal{H}(t)} \approx \frac{\omega[m^2 + \mathbf{P}_\perp^2(t)]}{2\mathcal{H}(t)(\omega-\mathcal{H}(t))} - \frac{k^2}{2\omega}, \tag{36}$$

where $\omega = k^0$ and $k^2 = \omega^2 - \mathbf{k}^2$. Besides, higher order diagrams acquire new objects, viz., Green's functions for particles and photons. In the Furry representation, the photon propagator remains free, while Green's function for particles can be expressed in terms of the solutions of wave equations. For example, the electron Green's function $G(x_2, x_1)$ can be expressed in terms of the positive- and negative-frequency solutions $\psi_i^{(\pm)}(x)$

of the Dirac equation in the corresponding external field:

$$iG(x_2, x_1) = \vartheta(t_2 - t_1)\sum_i \psi_i^{(+)}(x_2)\bar{\psi}_i^{(+)}(x_1) - \vartheta(t_1 - t_2)\sum_i \psi_i^{(-)}(x_2)\bar{\psi}_i^{(-)}(x_1). \quad (37)$$

In this expression, we can use the operator form of the solutions of the Dirac equation (17) and then proceed in the same way as when calculating the matrix elements for lower order processes.

3. SCATTERING OF A HIGH-ENERGY PHOTON IN A LOCALIZED ELECTRIC FIELD

We will apply the term localized to an electric field having a maximum (probably a singularity) at a single point and decreasing with increasing distance from this point. Such fields are, for example, the field of a Coulomb center or the field of an solitary atom. We assume that the field is independent of time and is arbitrary in all other respects; in particular, potential $V(\mathbf{r})$ is not necessarily centrosymmetric. For high energies ($\omega \gg m$), the transition from the initial (momentum $\mathbf{k}_1$ and polarization $\mathbf{e}_1$) to the final photon ($\mathbf{k}_2$, $\mathbf{e}_2$) occurs through a virtual electron–positron pair which interacts with the external field. We consider here just this scattering mechanism. It was studied experimentally by some authors (see [5] and review [6]). This scattering (which is often referred to as Delbrück scattering) was studied comprehensively by Cheng and Wu [7–9] for a Coulomb center. Another representation for the amplitude of this process was obtained in [10] with the help of the semiclassical electron Green's function in the Coulomb field. A generalization of this function to the case of an arbitrary centrosymmetric field was made by Lee and Milstein [11] who also considered the Delbrück scattering by the screened Coulomb potential for a momentum transfers $\mathbf{\Delta} = \mathbf{k}_2 - \mathbf{k}_1$ much smaller than the electron mass, virtually under complete screening.

In all the theoretical publications mentioned above, small-angle scattering was considered, for which $\Delta \ll \omega$. This condition is required for the applicability of the semiclassical approximation since it ensures a large angular momentum: $l \sim \omega\rho \sim \omega/\Delta \gg 1$. The characteristic sizes of the problem under investigation (for a Coulomb field) are the Compton wavelength for an electron, $1/m$, and the distance (time) over which the process is formed, $\sim\omega/m^2$. The impact parameters of such an order of magnitude ($\rho \sim 1/\Delta$) correspond to $\Delta \sim m$ and $\Delta \sim m^2/\omega$. If the potential decreases at a higher rate than the Coulomb potential, the problem acquires one more parameter, viz., the range of the potential r_{sc}. For the potential of an atom, the condition $r_{sc} \gg 1/m$ holds in all cases, while the relation between r_{sc} and ω/m^2 may be arbitrary depending on ω. Clearly, the screening may affect the process starting from $\omega \sim m^2 r_{sc}$.

The amplitude of the process under investigation has the form

$$T(k_1, k_2) = 2i\alpha\int d^4x_1 d^4x_2 \mathrm{Tr}[G(x_2, x_1)\hat{e}_1 \times \exp(-ik_1x_1)G(x_1, x_2)\hat{e}_2^* \exp(ik_2x_2)], \quad (38)$$

where the function $G(x_2, x_1)$ is defined in Eq. (37). Substituting Eq. (37) into Eq. (38), taking into consideration that (see the discussion in [10]) for $\omega \gg m$, the contribution to Eq. (38) to within terms of the order of $(\omega/m)^2$ comes from the diagram in which the production of an e^+e^- pair by the initial photon precedes its annihilation to the final photon, we obtain, retaining only the terms proportional to $9(t_2 - t_1)$, and the following expression for $T(k_1, k_2)$:

$$\begin{aligned} T(k_1, k_2) &= 2i\alpha\sum_{i,f}\int d^4x_1 d^4x_2 \vartheta(t_2 - t_1) \\ &\times \bar{\psi}_f^{(+)}(x_1)\hat{e}_1 \exp(-ik_1x_1)\psi_i^{(-)}(x_1) \\ &\times \bar{\psi}_i^{(-)}(x_2)\hat{e}_2^* \exp(ik_2x_2)\psi_f^{(+)}(x_2) \\ &= 2i\alpha\sum_{i,f}\int_{-\infty}^{\infty} dt_1 \int_{-\infty}^{\infty} dt_2 \vartheta(t_2 - t_1) \\ &\times V_{if}(\mathbf{k}_1, \mathbf{e}_1, t_1)V_{if}^+(\mathbf{k}_2, \mathbf{e}_2, t_2). \end{aligned} \quad (39)$$

The factor $V_{if}(\mathbf{k}, \mathbf{e}, t)$ is defined in Eq. (32) and is the integrand (in t) in the matrix element corresponding to the production of an e^+e^- pair by a photon. In zero external field, photon scattering does not take place and, hence, we presume the subtraction of the quantity T for $A_\mu = 0$ in Eqs. (38), (39), and subsequent expressions for the amplitude. This subtraction will be carried out explicitly below. Carrying out the summation over $|i\rangle$ and over all spin states in Eq. (39) and taking advantage of the possibility of cyclic permutation for the operators in expression

$$\sum_f \langle f|\ldots|f\rangle,$$

we transform the integrand in Eq. (39) to

$$\begin{aligned} \sum_{i,f} V_{if}(\mathbf{k}_1, \mathbf{e}_1, t_1)V_{if}^+(\mathbf{k}_2, \mathbf{e}_2, t_2) &= \exp[i(\omega_2 - \omega_1)t] \\ \times \mathrm{Tr}\sum_f \langle f|R\left(\mathbf{k}_1, \mathbf{e}_1, -\frac{\tau}{2}\right)D(\mathbf{k}_1, \mathbf{k}_2, \tau)R^+\left(\mathbf{k}_2, \mathbf{e}_2, \frac{\tau}{2}\right)|f\rangle, \\ D(\mathbf{k}_1, \mathbf{k}_2, \tau) &= \exp\left\{-i(H_e - \omega_1)\frac{\tau}{2}\right\}\exp(i\mathbf{k}_1 \cdot \mathbf{r}) \end{aligned} \quad (40)$$

$$\times \exp\{-i(\mathcal{H} - V(\mathbf{r}))\tau\}\exp(-i\mathbf{k}_2\cdot\mathbf{r})$$
$$\times \exp\left\{-i(H_e - \omega_2)\frac{\tau}{2}\right\},$$

where operator $R(\mathbf{k}_1, \mathbf{e}_1, -\tau/2)$ is defined in Eq. (33), and we have made a transition from t_1, t_2 to $t = (t_1 + t_2)/2$ and $\tau = t_2 - t_1$. The integration with respect to t in Eq. (39) gives $2\pi\delta(\omega_2 - \omega_1)$. We can now put $\omega_1 = \omega_2 \equiv \omega$. Then operator $D(\mathbf{k}_1, \mathbf{k}_2, \tau)$ in Eq. (40) can be written in the form

$$D(\mathbf{k}_1, \mathbf{k}_2, \tau) = L_p\left(\mathbf{k}_1, -\frac{\tau}{2}\right)\exp(-i\boldsymbol{\Delta}\cdot\mathbf{r})L_p^{(+)}\left(\mathbf{k}_2, \frac{\tau}{2}\right)$$
$$= T_-\left[\exp\left\{i\int_0^{-\tau/2} B_p(s, \mathbf{k}_1)\right\}\exp(-i\boldsymbol{\Delta}\cdot\mathbf{r})\right. \quad (41)$$
$$\left.\times \exp\left\{-i\int_0^{\tau/2} B_p(s, \mathbf{k}_2)\right\}\right],$$

where we have used the definition of operator $L_p(\mathbf{k}, t)$ and its explicit expression (34) in terms of $B_p(s, \mathbf{k}) = \mathcal{H}(s) + \mathcal{H}'(s, \mathbf{k}) - \omega$. It should be recalled that the vector potential is now equal to zero ($\mathbf{A} = 0$) and

$$\mathcal{H} = \sqrt{\mathbf{p}^2 + m^2}, \quad \mathcal{H}(s) = \sqrt{\mathbf{p}^2(s) + m^2},$$
$$\mathcal{H}'(s, \mathbf{k}) = \sqrt{(\mathbf{p}(s) - \mathbf{k})^2 + m^2}.$$

Defining amplitude M through the relation $T(\mathbf{k}_1, \mathbf{k}_2) = M\delta(\omega_2 - \omega_1)$ and inserting

$$I = \int d^3r|\mathbf{r}\rangle\langle\mathbf{r}|$$

in front of operator $\exp(i\boldsymbol{\Delta}\cdot\mathbf{r})$, which allows us to carry out the summation over f, we obtain

$$M = 4\pi i\alpha\int d^3r\exp(-i\boldsymbol{\Delta}\cdot\mathbf{r})$$
$$\times \mathrm{Tr}\int_0^\infty d\tau\langle\mathbf{r}|T_-\exp\left\{-i\int_0^{\tau/2} ds B_p(s, \mathbf{k}_2)\right\}$$
$$\times R^+\left(\mathbf{k}_2, \mathbf{e}_2, \frac{\tau}{2}\right)R\left(\mathbf{k}_1, \mathbf{e}_1, -\frac{\tau}{2}\right) \quad (42)$$
$$\times T_-\exp\left\{i\int_0^{-\tau/2} ds B_p(s, k_1)\right\}|\mathbf{r}\rangle.$$

Amplitude M is normalized in the same way as in [5–11] so that the differential scattering cross section for a photon is $d\sigma/d\Omega = |M/(4\pi)|^2$. It should be noted that expression (42) has the form

$$M = \int d^3r\exp(-i\boldsymbol{\Delta}\cdot\mathbf{r})Q(\mathbf{r}, \boldsymbol{\Delta}).$$

It will be verified below that for $\Delta \ll m$, we can disregard the dependence of $Q(\mathbf{r}, \boldsymbol{\Delta})$ on the momentum transfer $\boldsymbol{\Delta}$. In this case, expression (42) assumes a form typical of the potential scattering amplitude obtained in the first Born approximation. The role of the potential of interaction of a photon with the external field is played here by the complex quantity

$$U_{ph}(\mathbf{r}) = -Q(\mathbf{r}, 0)/(2\omega).$$

In fact, this interaction is of the tensor type since $Q(\mathbf{r}, 0)$ is a contraction: $Q(\mathbf{r}, 0) = e_{1i}Q_{ij}(\mathbf{r}, 0)e_{2j}^*$.

Passing to the computation of amplitude M, we choose the axis of the cylindrical system of coordinates (z axis) along $\boldsymbol{\nu} = (\mathbf{n}_1 + \mathbf{n}_2)/|\mathbf{n}_1 + \mathbf{n}_2|$; then $\mathbf{r} = (\boldsymbol{\rho}, z)$, and $\boldsymbol{\rho}\cdot\boldsymbol{\nu} = 0$ and $\boldsymbol{\Delta}\cdot\mathbf{r} = \boldsymbol{\Delta}\cdot\boldsymbol{\rho}$ since $\boldsymbol{\Delta}\cdot\boldsymbol{\nu} = 0$. The coordinate and momentum operators in the interior of $\langle\mathbf{r}|\ldots|\mathbf{r}\rangle$ in expression (42) appear only in the Heisenberg momentum operators $\mathbf{p}(t)$:

$$\mathbf{p}(t) = \mathbf{p} - \int_0^t ds\frac{\partial V(\mathbf{r}(s))}{\partial\mathbf{r}} \equiv \mathbf{p} + \delta\mathbf{p}(t), \quad (43)$$

where (cf. Eq. (31)) $\mathbf{p} \equiv \mathbf{p}(0)$ is the Schrödinger momentum operator $\mathbf{p} = -i\partial/\partial\mathbf{r}$. For the z component of $\mathbf{p}(t)$, we obtain from relation (43) $p_z(t) = p_z + \delta p_z(t)$, where

$$\delta p_z(t) \approx \delta V(t) = V(\mathbf{r}) - V(\mathbf{r} + \boldsymbol{\nu}t).$$

Since the contribution to expression (42) comes from large values of $p_z \sim \varepsilon$, we can substitute p_z for $p_z(t)$ to within corrections of the order of $\delta V(t)/\varepsilon$, which we systematically disregard. The relative value of commutator $[p_z, p_\perp^i(t)]$ is of the same order of smallness ($\sim\delta V(t)/\varepsilon$). Thus, operator $p_z(t)$ coincides within the accepted accuracy with the free operator and commutes with all operators in expression (42). This allows us to rewrite this expression filling the interior of $\langle z|\ldots|z\rangle$ with the complete set of the eigenfunctions of operator p_z, i.e., with plane waves. Schematically, in expression (42) we have

$$\int_{-\infty}^\infty dz\langle z|f(p_z, z)|z\rangle = \int_{-\infty}^\infty\frac{dq}{2\pi}\int_{-\infty}^\infty dz f(q, z),$$

where the integration with respect to q must be carried out from 0 to ω since the compensation in the exponents (cancelling out in $B_p(s, \mathbf{k})$, see Eq. (35)) occurs precisely for this values of q. Introducing the notation $\varepsilon = \sqrt{q^2 + m^2}$ and going over to $x = \varepsilon/\omega$, we obtain

$$M = \frac{i\alpha}{2\omega}\mathrm{Tr}\int_0^1\frac{dx}{[x(1-x)]^2}\int d^2\rho\exp(-i\boldsymbol{\Delta}\cdot\boldsymbol{\rho})$$

$$\times\int_0^\infty d\tau\int_{-\infty}^{\infty} dz\langle\boldsymbol{\rho}|T_-\exp\left\{-i\int_0^{\tau/2} ds\, f_2(s)\right\} \tag{44}$$

$$\times R_2^+ R_1 T_-\exp\left\{-i\int_{-\tau/2}^{0} ds\, f_1(s)\right\}|\boldsymbol{\rho}\rangle,$$

where

$$f_{1,2}(s) = \frac{m^2 + (\mathbf{p}_\perp \pm x\boldsymbol{\Delta}/2)^2}{2\omega x(1-x)}.$$

As in [10, 11], we will consider the transitions from one state to another for photons with a definite helicity. In this case, all amplitudes can be expressed (see, for example, [10]) in terms of two independent quantities M_{++} (without a change in helicity) and M_{+-} (with a reversal in helicity). For spiral states, operators R_2 and R_1 in expression (44) have the form

$$R_1 = \left(\mathbf{e}_1\cdot\left(\mathbf{p}_\perp\left(-\frac{\tau}{2}\right)+\boldsymbol{\Delta}\frac{x}{2}\right)\right)$$

$$\times(\lambda_1 + (2x-1)\boldsymbol{\sigma}\cdot\mathbf{v}) - m\boldsymbol{\sigma}\cdot\mathbf{e}_1,$$

$$R_2 = \left(\mathbf{e}_2\cdot\left(\mathbf{p}_\perp\left(\frac{\tau}{2}\right)-\boldsymbol{\Delta}\frac{x}{2}\right)\right) \tag{45}$$

$$\times(\lambda_2 + (2x-1)\boldsymbol{\sigma}\cdot\mathbf{v}) - m\boldsymbol{\sigma}\cdot\mathbf{e}_2,$$

where $\lambda_{1,2}$ is the helicity of the initial and final photons, respectively. Since we are dealing with a small-angle problem ($\Delta \ll \omega$), we can put $\mathbf{e}_1\cdot\mathbf{v} = \mathbf{e}_2\cdot\mathbf{v} = 0$ in Eqs. (45). To the same degree of accuracy, the substitutions $\boldsymbol{\sigma}\cdot\mathbf{n}_1 \longrightarrow \boldsymbol{\sigma}\cdot\mathbf{v}$ and $\boldsymbol{\sigma}\cdot\mathbf{n}_2 \longrightarrow \boldsymbol{\sigma}\cdot\mathbf{v}$ have been carried out in R_1 and R_2, respectively.

In the integration over τ and z in Eq. (44), the contribution comes from $\tau \sim |z| \sim \omega/m^2$. The integral over ρ converges (for the difference $M(A_\mu) - M(0)$) on $\min(\Delta^{-1}, \omega/m^2, r_{sc})$, and for $\Delta \geq m$, the contribution comes from $\rho \leq 1/m$. For such values of momentum transfer, a transition to classical trajectories in the calculation of the average $\langle\boldsymbol{\rho}|\ldots|\boldsymbol{\rho}\rangle$ is ruled out since in this case fulfillment of the condition $\rho \gg 1/m$ is required (see Section 2). In accordance with the above analysis, we divide the domain of integration with respect to ρ in Eq. (44) into two parts: $M \equiv \tilde{M}_\mathrm{I} + \tilde{M}_\mathrm{II}$, such that the integration over ρ in $\tilde{M}_\mathrm{I}$ is carried out from zero to ρ_0; in $\tilde{M}_\mathrm{II}$, from ρ_0 to infinity. The value of the parameter ρ_0 is determined by the inequalities $1/m \ll \rho_0 \ll \min(\omega/m^2, r_{sc})$. The Heisenberg operator $\mathbf{r}(s)$ in expression (43) is given by

$$\mathbf{r}(s) = \mathbf{r} + \int_0^s dx\mathbf{v}(x),$$

where $\mathbf{v}(x)$ is the velocity operator. The motion in z is assumed to be free starting with Eq. (44): $z(s) \approx z + s$. For the coordinate transverse to $\mathbf{v}$, we have (in the first approximation)

$$\boldsymbol{\rho}(s) = \boldsymbol{\rho} + s\mathbf{v}_\perp$$

$$-\frac{1}{\varepsilon}\int_0^s dx(s-x)\frac{\partial}{\partial\boldsymbol{\rho}}V(\boldsymbol{\rho} + x\mathbf{v}_\perp + \mathbf{v}(z+x)).$$

The relative value of the integral term in this expression is V/ε and should be neglected. Thus, we can put $\mathbf{r}(s) = \mathbf{r} + s\mathbf{v}_\perp + (z+s)\mathbf{v}$ everywhere, which means that a rectilinear trajectory approximation is applicable in our problem. Then the expression for the transverse momentum operator reads

$$\mathbf{p}_\perp(t) = \mathbf{p}_\perp - \int_z^{z+t} ds\frac{\partial}{\partial\boldsymbol{\rho}}V(\boldsymbol{\rho} + (s-z)\mathbf{v}_\perp + s\mathbf{v}), \tag{46}$$

where the contribution to the integral over s for any ρ comes from an interval of the order of the impact parameter ρ itself. The term $(s-z)\mathbf{v}_\perp$ in Eq. (46) has the absolute value $|(s-z)\mathbf{v}_\perp| \sim |z\mathbf{v}_\perp| \sim 1/m$ and can be neglected in region II, where $\rho > \rho_0 \gg 1/m$. For $\rho < \rho_0$, we can replace the argument of the potential in Eq. (46) by $\boldsymbol{\rho} - z\mathbf{v}_\perp + s\mathbf{v}$ since $s \sim \rho \ll |z|$ in this region. We can now eliminate $\mathbf{v}_\perp$ in the argument of the potential by applying the following operation:

$$f(\boldsymbol{\rho} - z\mathbf{v}_\perp) = \exp\left(-\frac{i\mathbf{p}_\perp^2 z}{2\varepsilon}\right)f(\boldsymbol{\rho})\exp\left(\frac{i\mathbf{p}_\perp^2 z}{2\varepsilon}\right).$$

After this, operator $\mathbf{p}_\perp(t)$ in amplitude (44) for any ρ has the form

$$\mathbf{p}_\perp(t) = \mathbf{p}_\perp - \int_z^{z+t} ds\frac{\partial}{\partial\boldsymbol{\rho}}V(\boldsymbol{\rho} + s\mathbf{v}) \equiv \mathbf{p}_\perp + \boldsymbol{\delta}(t). \tag{47}$$

It should be noted that $|\boldsymbol{\delta}(t)| \sim |V(\boldsymbol{\rho})|$ and the characteristic values of $|\mathbf{p}_\perp|$ are $\max(m, \Delta)$, which follows ardeady from an analysis of the form of functions $f_{1,2}(s)$ in Eq. (44). According to this estimate, the value of $\boldsymbol{\delta}(t)$ in region II ($\rho > \rho_0$) is small ($|\boldsymbol{\delta}(t)| \ll m$) and the corresponding expansion should be carried out. As a result, the contribution of $\tilde{M}_\mathrm{II}$ to the total amplitude does not contain higher order corrections in the external field. Moreover, owing to the factor $\exp(-i\boldsymbol{\Delta}\cdot\boldsymbol{\rho})$ and condition $\rho_0 \gg 1/m$, this contribution is not negligibly small only for transferred momenta $\Delta \ll m$; this circumstance will be used in the subsequent analysis.

In region I, where $s \sim \rho \ll |z|$, the following approximate expression is valid for $\boldsymbol{\delta}(t)$ from relation (47):

$$\boldsymbol{\delta}(t) = \mathbf{Q}[\vartheta(-t)\vartheta(z)\vartheta(-t-z) - \vartheta(t)\vartheta(-z)\vartheta(t+z)],$$

$$\mathbf{Q} = \frac{\partial\chi(\boldsymbol{\rho})}{\partial\boldsymbol{\rho}}, \quad \chi(\boldsymbol{\rho}) = \int_{-\infty}^{\infty} ds V(\boldsymbol{\rho} + s\mathbf{v}). \tag{48}$$

In amplitude $\tilde{M}_I$, as in Eq. (44), the argument of the function $\delta(t)$ appearing in $f_{1,2}(s)$ as well as in $R_{1,2}$ belongs to the interval $|t| \le \tau/2$. For such values of t, the quantity $\delta(t)$ vanishes for $|z| \ge \tau/2$ in accordance with relations (48), and the entire expression is independent of the external field. Such contributions vanish when the amplitude is subtracted in zero field and can be discarded at the very outset. In other words, integration with respect to z in $\tilde{M}_I$ should be carried out between $-\tau/2$ and $\tau/2$. Thus, the integrand in $\tilde{M}_I$ (the integration is carried out over x, ρ, and τ as in Eq. (44)) assumes the form

$$\int_{-\tau/2}^{0} dz \langle \boldsymbol{\rho}| \exp\left\{\frac{iz}{2\omega x}(D_- - \mathbf{p}_\perp^2)\right\}$$

$$\times \exp\left\{\frac{-i(z+\tau/2)}{2\omega x(1-x)}\left[m^2 + \left(\mathbf{p}_\perp - \mathbf{Q} - x\frac{\boldsymbol{\Delta}}{2}\right)^2\right]\right\}$$

$$\times R_2^+(\mathbf{p}_\perp - \mathbf{Q}) R_1(\mathbf{p}_\perp) \exp\left\{\frac{i}{2\omega x}\left[z\mathbf{p}_\perp^2 - \frac{\tau}{2}D_+\right]\right\}|\boldsymbol{\rho}\rangle \quad (49)$$

$$+ \int_0^{\tau/2} dz \langle \boldsymbol{\rho}| \exp\left\{-\frac{i}{2\omega x}\left(z\mathbf{p}_\perp^2 + \frac{\tau}{2}D_-\right)\right\} R_2^+(\mathbf{p}_\perp) R_1(\mathbf{p}_\perp + \mathbf{Q})$$

$$\times \exp\left\{\frac{i(z-\tau/2)}{2\omega x(1-x)}\left[m^2 + \left(\mathbf{p}_\perp + \mathbf{Q} + x\frac{\boldsymbol{\Delta}}{2}\right)^2\right]\right\}$$

$$\times \exp\left\{\frac{iz}{2\omega x}[\mathbf{p}_\perp^2 - D_+]\right\}|\boldsymbol{\rho}\rangle,$$

$$D_\pm = \frac{m^2 + (\mathbf{p}_\perp \pm x\boldsymbol{\Delta}/2)^2}{1-x},$$

where $R_{1,2}(\mathbf{P})$ denotes the functions defined by formulas (45), in which $\mathbf{p}_\perp(\mp\tau/2)$ are replaced by $\mathbf{P}$. The factorization in relations (49) of the factors which are functions only of $\mathbf{p}_\perp$ and those depending on the operators $\mathbf{p}_\perp \pm \mathbf{Q}$ containing the external field is essential for the subsequent considerations. It should be emphasized that the existence of T ordering in Eq. (44) plays a key role for the factorization of the exponential operators appearing in this expression. For example, taking into account relations (48), for $-\tau/2 \le z \le 0$, we have

$$T_- \exp\left\{-i\int_0^{\tau/2} ds\, f_2(s)\right\}$$

$$= \exp\left\{\frac{iz}{2\omega x(1-x)}\left[m^2 + \left(\mathbf{p}_\perp - x\frac{\boldsymbol{\Delta}}{2}\right)^2\right]\right\} \quad (50)$$

$$\times \exp\left\{\frac{-i(z+\tau/2)}{2\omega x(1-x)}\left[m^2 + \left(\mathbf{p}_\perp - \mathbf{Q} - x\frac{\boldsymbol{\Delta}}{2}\right)^2\right]\right\}.$$

The eigenfunctions of operators $\mathbf{p}_\perp \pm \mathbf{Q}$ corresponding to eigenvalue $\mathbf{q}$ are $\exp\{i(\mathbf{q}\cdot\boldsymbol{\rho} \mp \chi(\boldsymbol{\rho}))\}$, where $\chi(\boldsymbol{\rho})$ is the eikonal phase defined in Eqs. (48). Using these functions, we can now calculate the average $\langle\boldsymbol{\rho}|\ldots|\boldsymbol{\rho}\rangle$ in Eqs. (49). Inserting

$$I = \int d^2\boldsymbol{\rho}_1 |\boldsymbol{\rho}_1\rangle\langle\boldsymbol{\rho}_1|$$

between the free operators and the operators depending on the field, we reduce the calculation of the average to determining the matrix elements of the following two types:

$$\langle\boldsymbol{\rho}_2| f(\mathbf{p}_\perp)|\boldsymbol{\rho}_1\rangle = \int \frac{d^2q}{(2\pi)^2} \exp\{i(\mathbf{q}\cdot(\boldsymbol{\rho}_2 - \boldsymbol{\rho}_1))\} f(\mathbf{q}),$$

$$\langle\boldsymbol{\rho}_2| f(\mathbf{p}_\perp \pm \mathbf{Q})|\boldsymbol{\rho}_1\rangle = \int \frac{d^2q}{(2\pi)^2} \quad (51)$$

$$\times \exp\{i(\mathbf{q}\cdot(\boldsymbol{\rho}_2 - \boldsymbol{\rho}_1))\} \exp\{\mp i[\chi(\boldsymbol{\rho}_2) - \chi(\boldsymbol{\rho}_1)]\} f(\mathbf{q}).$$

The c-number expression obtained as a result of averaging can easily be integrated over z and τ. This gives

$$\tilde{M}_I = -i\frac{\alpha\omega}{\pi^4}\int_0^{\rho_0} d^2\rho \exp(-i\boldsymbol{\Delta}\cdot\boldsymbol{\rho})$$

$$\times \int d^2\rho_1 \int_0^1 dx \exp\{i(2x-1)\boldsymbol{\Delta}\cdot\boldsymbol{\rho}_1\}$$

$$\times \int \frac{d^2q_1 d^2q_2}{(m^2+q_1^2)(m^2+q_2^2)} \exp\{2i\cdot((\mathbf{q}_2 - \mathbf{q}_1)\cdot\boldsymbol{\rho}_1)\} \quad (52)$$

$$\{\exp[i(\chi(\boldsymbol{\rho}+\boldsymbol{\rho}_1) - \chi(\boldsymbol{\rho}-\boldsymbol{\rho}_1))] - 1\}$$

$$\times \{m^2(\mathbf{e}_1\cdot\mathbf{e}_2^*) + [\lambda_1\lambda_2 + (2x-1)^2](\mathbf{e}_1\cdot\mathbf{q}_1)(\mathbf{e}_2^*\cdot\mathbf{q}_2)\}.$$

In this expression, the required subtraction has already been made so that $\tilde{M}_I$ (52) vanishes after "switching off" the external field. It should be noted that if we take the half-sum of expression (52) and of the expression obtained from it as a result of substitutions $x \longrightarrow 1-x$, $\mathbf{q}_{1,2} \longrightarrow -\mathbf{q}_{1,2}$, $\boldsymbol{\rho}_1 \longrightarrow -\boldsymbol{\rho}_1$, the term $\exp[i(\chi(\boldsymbol{\rho}+\boldsymbol{\rho}_1) - \chi(\boldsymbol{\rho}-\boldsymbol{\rho}_1))] - 1$ in Eq. (52) is transformed into $\cos[\chi(\boldsymbol{\rho}+\boldsymbol{\rho}_1) - \chi(\boldsymbol{\rho}-\boldsymbol{\rho}_1)] - 1$. It hence becomes obvious that the Furry theorem holds: the expansion of expression (52) into a series in perturbation theory contains only the even power of the potential. The integral with respect

to $\mathbf{q}_{1,2}$ in Eq. (52) can be expressed in terms of the Macdonald functions K_0 and K_1 using the relations

$$\int \frac{d^2q}{m^2+q^2}\exp(-2i\mathbf{q}\cdot\boldsymbol{\rho}) = 2\pi K_0(2m\rho),$$
$$\int \frac{d^2q}{m^2+q^2}\mathbf{q}\exp(-2i\mathbf{q}\cdot\boldsymbol{\rho}) = -2\pi i K_1(2m\rho)m\frac{\boldsymbol{\rho}}{\rho}. \quad (53)$$

We can also evaluate the elementary integral with respect to x, thus expressing $\tilde{M}_\mathrm{I}$ for an arbitrary potential in terms of the 4-fold integral:

$$\tilde{M}_\mathrm{I} = 4i\alpha\frac{\omega m^2}{\pi^2}\int_0^{\rho_0} d^2\rho\exp(-i\boldsymbol{\Delta}\cdot\boldsymbol{\rho})$$
$$\times\int d^2\rho_1\{1-\cos[\chi(\boldsymbol{\rho}+\boldsymbol{\rho}_1)-\chi(\boldsymbol{\rho}-\boldsymbol{\rho}_1)]\}$$
$$\times\left\{K_0^2(2m\rho_1)(\mathbf{e}_1\cdot\mathbf{e}_2^*)+\frac{(\mathbf{e}_1\cdot\boldsymbol{\rho}_1)(\mathbf{e}_2^*\cdot\boldsymbol{\rho}_1)}{\rho_1^2}\right. \quad (54)$$
$$\left.\times K_1^2(2m\rho_1)\left[\lambda_1\lambda_2-\frac{d^2}{dt^2}\right]\right\}\frac{\sin t}{t},$$

where $t=\boldsymbol{\Delta}\cdot\boldsymbol{\rho}_1$. Sometimes, a different integration sequence in Eq. (52) is more convenient for a comparison of our results with those obtained by other authors. For instance, the form of the scattering amplitude obtained in [12] for the Coulomb potential for $\Delta \gg m^2/\omega$ can be reproduced by presenting factors $(m^2+q_{1,2}^2)^{-1}$ in Eq. (52) in the form

$$\frac{1}{m^2+q_{1,2}^2} = i\int_0^\infty ds_{1,2}\exp\{-is_{1,2}(m^2+q_{1,2}^2)\}$$

and then integrating with respect to $\mathbf{q}_{1,2}$.

Let us now present $\tilde{M}_\mathrm{I}$ as $\tilde{M}_\mathrm{I}\equiv M_\mathrm{I}-\delta M_\mathrm{I}$, where we choose for M_I expression (52) or (54), in which the integration over ρ is extended to infinity, while in δM_I this integration is carried out from ρ_0 to infinity. It follows from relations (53) that the integral with respect to ρ_1 in Eq. (52) converges for $\rho_1\sim 1/m$. Then and the following expansion can be used for evaluating δM_I as $\rho\geq\rho_0\gg\rho_1$ in the entire domain:

$$1-\cos[\chi(\boldsymbol{\rho}+\boldsymbol{\rho}_1)-\chi(\boldsymbol{\rho}-\boldsymbol{\rho}_1)]\approx 2\left(\boldsymbol{\rho}_1\cdot\frac{\partial\chi(\boldsymbol{\rho})}{\partial\boldsymbol{\rho}}\right)^2.$$

Besides, the quantity δM_I (as well as $\tilde{M}_\mathrm{II}$, see the discussion following formula (47)) should be taken into account only for the transferred momenta $\Delta\ll m$. In this case, $|\boldsymbol{\Delta}\cdot\boldsymbol{\rho}_1|\ll 1$, and the quantity $\exp[i(2x-1)\boldsymbol{\Delta}\cdot\boldsymbol{\rho}_1]$ can be replaced by unity. After this, we first integrate over $\boldsymbol{\rho}_1$ with the result proportional to $\delta(\mathbf{q}_2-\mathbf{q}_1)$ and then over the remaining variables except for $\boldsymbol{\rho}$. Finally, for δM_I, we obtain

$$\delta M_\mathrm{I} = \frac{i\alpha\omega}{18\pi m^2}b_{ij}$$
$$\times\int_{\rho_0}^\infty d^2\rho\exp(-i\boldsymbol{\Delta}\cdot\boldsymbol{\rho})\frac{\partial\chi(\boldsymbol{\rho})}{\partial\rho_i}\frac{\partial\chi(\boldsymbol{\rho})}{\partial\rho_j}, \quad (55)$$

where $b_{ij}^{(++)}=7\delta_{ij}$ and $b_{ij}^{(+-)}=-2e_ie_j$ (the symbol (++) corresponds to amplitudes without a change in helicity and (+–) to those with helicity reversal). It should be emphasized that the quantity M_I no longer depends on the parameter ρ_0. For high values of transferred momentum (for $\Delta\gg m^2/\omega$ in the Coulomb potential), the entire amplitude is reduced to M_I.

Going over to the calculation of $\tilde{M}_\mathrm{II}$, we recall that the transferred momentum Δ in this quantity should be neglected as compared to mass m, and in the averaging $\langle\boldsymbol{\rho}|\ldots|\boldsymbol{\rho}\rangle$, all the interior operators can be regarded as commuting. Consequently, in $\tilde{M}_\mathrm{II}$ we have

$$\langle\boldsymbol{\rho}|f(\mathbf{p}_\perp,\boldsymbol{\rho})|\boldsymbol{\rho}\rangle = \int\frac{d^2q}{(2\pi)^2}f(\mathbf{q},\boldsymbol{\rho}),$$

and Eq. (44) leads to

$$\tilde{M}_\mathrm{II} = \frac{i\alpha}{\omega}\int\frac{d^2q}{(2\pi)^2}\int_{\rho_0}^\infty d^2\rho\exp(-i\boldsymbol{\Delta}\cdot\boldsymbol{\rho})$$
$$\times\int_0^1\frac{dx}{x^2(1-x)^2}\int_0^\infty d\tau\int_{-\infty}^\infty dz\exp\left\{-\frac{i\tau\Phi}{2\omega x(1-x)}\right\} \quad (56)$$
$$\times\{m^2(\mathbf{e}_1\cdot\mathbf{e}_2^*)+[\lambda_1\lambda_2+(2x-1)^2]$$
$$\times(\mathbf{e}_1\cdot(\mathbf{q}+\boldsymbol{\delta}(-\tau/2)))(\mathbf{e}_2^*\cdot(\mathbf{q}+\boldsymbol{\delta}(\tau/2)))\},$$

where $\boldsymbol{\delta}(t)$ is defined by Eq. (47) and

$$\Phi = m^2+\mathbf{q}^2+2(\mathbf{q}\cdot\langle\boldsymbol{\delta}\rangle)+\langle\boldsymbol{\delta}^2\rangle,$$
$$\langle\boldsymbol{\delta}\rangle = \frac{1}{\tau}\int_{-\tau/2}^{\tau/2}ds\boldsymbol{\delta}(s),\quad \langle\boldsymbol{\delta}^2\rangle = \frac{1}{\tau}\int_{-\tau/2}^{\tau/2}ds\boldsymbol{\delta}^2(s). \quad (57)$$

We now evaluate the Gaussian integral over $\mathbf{q}$ in Eq. (56), expand the exponential in $\boldsymbol{\delta}$ ($|\boldsymbol{\delta}|\ll m$), and subtract the value of $\tilde{M}_\mathrm{II}$ for zero external field. This gives

$$\tilde{M}_\mathrm{II} = -\frac{\alpha}{2\pi}\int_{\rho_0}^\infty d^2\rho\exp(-i\boldsymbol{\Delta}\cdot\boldsymbol{\rho})$$
$$\times\int_0^1 dx\int_0^\infty\frac{d\tau}{\tau}\exp(-i\mu\tau)\int_{-\infty}^\infty dzF, \quad (58)$$

where $\mu = m^2/[2\omega x(1-x)]$ and F has the form

$$F^{(++)} = \frac{x^2+(1-x)^2}{2x(1-x)}\left(\delta\left(\frac{\tau}{2}\right)-\delta\left(-\frac{\tau}{2}\right)\right)^2 + \frac{4}{\tau}\int_{-\tau/2}^{\tau/2} ds s\left(\frac{d\delta(s)}{ds}\cdot(\delta(s)-\langle\delta\rangle)\right), \quad (59)$$

$$F^{(+-)} = 4\left(\mathbf{e}\cdot\left(\delta\left(-\frac{\tau}{2}\right)-\langle\delta\rangle\right)\right)\left(\mathbf{e}\cdot\left(\delta\left(\frac{\tau}{2}\right)-\langle\delta\rangle\right)\right).$$

While deriving the expression for $F^{(++)}$, we integrated by parts with respect to variable τ in the term proportional to $m^2(\mathbf{e}_1\cdot\mathbf{e}_2^*)$ in Eq. (56). Substituting the Fourier representation of potential

$$V(\boldsymbol{\rho}) = \int\frac{d^2Q_\perp dQ_\parallel}{(2\pi)^3} \times \exp(iQ_\parallel z)\exp(i\mathbf{Q}_\perp\cdot\boldsymbol{\rho})\tilde V(\mathbf{Q}_\perp, Q_\parallel) \quad (60)$$

into expression (47) for $\boldsymbol{\delta}(t)$, we obtain from Eq. (58)

$$\tilde M_{II} = \frac{\alpha}{\pi^2}\int_{\rho_0}^{\infty} d^2\rho\exp(-i\boldsymbol{\Delta}\cdot\boldsymbol{\rho})\int_0^1 dx\int_{-\infty}^{\infty} dQ_\parallel \times\int\frac{d^2Q_{1\perp}}{(2\pi)^2}\int\frac{d^2Q_{2\perp}}{(2\pi)^2}\exp\{i(\boldsymbol{\rho}\cdot(\mathbf{Q}_{1\perp}+\mathbf{Q}_{2\perp}))\} \times\tilde V(\mathbf{Q}_{1\perp}, Q_\parallel)\tilde V(\mathbf{Q}_{2\perp}, -Q_\parallel)Q_{1\perp}^iQ_{2\perp}^j\int_0^1\frac{dss}{(Q_\parallel s)^2-\mu^2+i0}F_{ij}, \quad (61)$$

$$F_{ij}^{(++)} = \delta_{ij}\left[\frac{x^2+(1-x)^2}{4x(1-x)}+s(1-s)\right], \quad F_{ij}^{(+-)} = -s^2e_ie_j.$$

We present product $\tilde V(\mathbf{Q}_{1\perp}, Q_\parallel)\tilde V(\mathbf{Q}_{2\perp}, -Q_\parallel)$ in Eqs. (61) in the form

$$[\tilde V(\mathbf{Q}_{1\perp}, Q_\parallel)\tilde V(\mathbf{Q}_{2\perp}, -Q_\parallel) - \tilde V(\mathbf{Q}_{1\perp}, 0)\tilde V(\mathbf{Q}_{2\perp}, 0)] + \tilde V(\mathbf{Q}_{1\perp}, 0)\tilde V(\mathbf{Q}_{2\perp}, 0)$$

and, accordingly, $\tilde M_{II} \equiv M_{II} + \delta M_{II}$, where M_{II} is the integral of the difference in the brackets in the later expression and δM_{II} is the integral of $V(\mathbf{Q}_{1\perp}, 0)V(\mathbf{Q}_{2\perp}, 0)$. Since small values of ρ make zero contribution to M_{II}, the integration can be extended to the entire range of variation of ρ; i.e., ρ_0 can be replaced by zero. Then we integrate with respect to $\boldsymbol{\rho}$:

$$M_{II} = \frac{\alpha\omega}{\pi^4m^2}\int_0^1 dx\int_0^\infty dy\int d^2q \times[\tilde V(\mathbf{q}, \mu y)\tilde V(\boldsymbol{\Delta}-\mathbf{q}, -\mu y) - V(\mathbf{q}, 0)V(\boldsymbol{\Delta}-\mathbf{q}, 0)]B,$$

$$B^{(++)} = (\mathbf{q}\cdot(\boldsymbol{\Delta}-\mathbf{q}))\int_0^1\frac{dss}{s^2y^2-1+i0} \times\left[\frac{x^2+(1-x)^2}{4}+x(1-x)s(1-s)\right], \quad (62)$$

$$B^{(+-)} = (\mathbf{e}\cdot\mathbf{q})(\mathbf{e}\cdot(\mathbf{q}-\boldsymbol{\Delta}))x(1-x)\int_0^1\frac{dss^3}{s^2y^2-1+i0},$$

where we have made a transition to variable $y = Q_\parallel/\mu$. The integrals with respect to s in expressions (62) can be evaluated easily:

$$B^{(++)} = \frac{(\mathbf{q}\cdot(\boldsymbol{\Delta}-\mathbf{q}))}{2y^2}\left\{\frac{1}{4}(x^2+(1-x)^2)\ln(1-y^2-i0) + x(1-x)\left[1-\frac{1}{y^2}\ln(1-y^2-i0)+\frac{1}{y}\ln\left(\frac{1-y-i0}{1+y}\right)\right]\right\}, \quad (63)$$

$$B^{(+-)} = \frac{1}{2y^2}(\mathbf{e}\cdot\mathbf{q})(\mathbf{e}\cdot(\mathbf{q}-\boldsymbol{\Delta}))x(1-x) \times\left[1+\frac{1}{y^2}\ln(1-y^2-i0)\right].$$

Using definition (48) of the eikonal phase $\chi(\boldsymbol{\rho})$ and Eqs. (60) and (61), we write δM_{II} in the form

$$\delta M_{II} = -\frac{\alpha}{\pi^2}\int_{\rho_0}^{\infty} d^2\rho\exp(-i\boldsymbol{\Delta}\cdot\boldsymbol{\rho})\frac{\partial\chi(\boldsymbol{\rho})}{\partial\rho_i}\frac{\partial\chi(\boldsymbol{\rho})}{\partial\rho_j} \times\int_0^1 dx\int_{-\infty}^{\infty} dQ_\parallel\int\frac{dss}{(Q_\parallel s)^2-\mu^2+i0}F_{ij}, \quad (64)$$

where quantities F_{ij} are defined in Eqs. (61). The integration in Eq. (64) can be easily performed with respect to $Q_\parallel$ and then with respect to x and s, and we verify that $\delta M_{II} = \delta M_I$, where δM_I is the quantity defined by Eq. (55). Thus, the total amplitude is given by

$$M = \tilde M_I + \tilde M_{II} = M_I - \delta M_I + M_{II} + \delta M_{II} = M_I + M_{II}$$

and does not depend on ρ_0.

Thus, for any $\Delta \ll \omega$, the photon scattering amplitude with $\omega \gg m$ for an arbitrary localized potential is given by the sum of expressions (62) and (52) (or (54)); in the latter equation, the integration with respect to $\boldsymbol{\rho}$ is carried out over the entire domain. It should be emphasized that until now, the expression for this amplitude, which is valid for any $\Delta \ll \omega$ without any additional limitations on the parameters of the problem, was known only for the Coulomb potential.

By way of example, we will now determine the contribution to amplitude M_{II} in the Moliére potential [13], for which

$$\tilde{V}(\mathbf{Q}) = -4\pi Z\alpha\sum_{n=1}^{3}\frac{a_n}{\mathbf{Q}^2+b_n^2},$$
$$a_1 = 0.1,\quad a_2 = 0.55,\quad a_3 = 0.35,\quad b_n = \beta_0\beta_n, \tag{65}$$
$$\beta_1 = 6,\quad \beta_2 = 1.2,\quad \beta_3 = 0.3,\quad \beta_0 = mZ^{1/3}/121.$$

It should be noted that the Fourier transform of the Coulomb potential can be obtained from relations (65) by putting $b_n = 0$ and $a_n = 1/3$ in them. We will henceforth assume that the unit vector $\mathbf{e}_1$ in the circular polarization vector $\mathbf{e} = (\mathbf{e}_1 + i\mathbf{e}_2)/\sqrt{2}$ lies in the scattering plane ($\mathbf{e}_1 \parallel \boldsymbol{\Delta}$). In this case, the amplitudes $M_\parallel$ and $M_\perp$ with the linear polarization vector lying in the scattering plane and perpendicular to this plane, which were used in [8], turn out to be connected with the helicity amplitudes through the following simple relations:

$$M_\perp = M_{++} - M_{+-},\quad M_\parallel = M_{++} + M_{+-}.$$

Substituting relations (65) into (62) and using the parametrization

$$(cd)^{-1} = \int_0^1 dt[ct + d(1-t)]^{-2},$$

we evaluate the integrals with respect to $\mathbf{q}$ and y in relations (62). Then the contribution of M_{II} to the total amplitude takes the form

$$M_{II}^{Mol} = \frac{4i\alpha(Z\alpha)^2\omega}{m^2}\int_0^1 dx\int_0^1 ds\int_0^1 dt\sum_{k,n=1}^{3} a_k a_n B_{Mol},$$
$$B_{Mol}^{(++)} = \left[x(1-x)[s^2+(1-s)^2]-\frac{1}{2}\right]\times\left[Q(t,s)+2\ln\left(1+\frac{i\mu}{sf(t)}\right)\right], \tag{66}$$
$$B_{Mol}^{(+-)} = s^2x(1-x)Q(t,s),$$

where $\mu = m^2/[2\omega x(1-x)]$ as before and the following notation is introduced:

$$f^2(t) = tb_k^2 + (1-t)b_n^2 + \Delta^2 t(1-t),$$
$$Q(t,s) = \frac{\Delta^2 t(1-t)}{f(t)}\left[\frac{s}{i\mu+sf(t)}-\frac{1}{f(t)}\right]. \tag{67}$$

For the Coulomb potential, we obtain M_{II}^C in form (66), where the sum

$$\sum_{k,n} a_k a_n B_{Mol}$$

is replaced by the quantity B_C, for which we have

$$B_C^{(++)} = \left[x(1-x)[s^2+(1-s)^2]-\frac{1}{2}\right]\times[g(s,t)-1-2\ln g(s,t)],$$
$$B_C^{(+-)} = s^2x(1-x)[g(s,t)-1], \tag{68}$$
$$g(s,t) = \left[1+\frac{i\lambda}{2s\sqrt{t(1-t)}}\right]^{-1},$$

where $\lambda = 2\mu/\Delta = m^2/[\omega\Delta x(1-x)]$. Note that for $\Delta \sim m$, the quantity M_{II}^{Mol} coincides with M_{II}^C to within small terms of the order of b_n^2/Δ^2. In turn, for $\Delta \sim m$, quantity B_C is small (of the order of m/ω), which explicitly confirms the above conclusion that M_{II} must be taken into account only for $\Delta \ll m$. Carrying out the substitution $t \longrightarrow (1+\sqrt{1-t^2})/2$, integrating by parts the term proportional to the logarithm in $B_C^{(++)}$ with respect to t, and integrating the terms which do not contain $g(s, t)$ with respect to x and s, we obtain

$$M_{II}^C = \frac{4i\alpha(Z\alpha)^2\omega}{m^2}\left[\int_0^1 dx\int_0^1 ds\int_0^1 dt F(s,t,x) + R_{II}\right],$$
$$F^{(++)}(s,t,x) = \left[x(1-x)[s^2+(1-s)^2]-\frac{1}{2}\right]\times\frac{s(2-t^2)}{\sqrt{1-t^2}(i\lambda+st)},$$
$$F^{(+-)}(s,t,x) = \frac{s^3t^2x(1-x)}{\sqrt{1-t^2}(i\lambda+st)}, \tag{69}$$
$$R_{II}^{(++)} = -\frac{7}{9}\left[1+\ln\left(\frac{im^2}{2\omega\Delta}\right)\right]-2,$$
$$R_{II}^{(+-)} = -\frac{1}{18}.$$

In order to obtain the total amplitude of the photon scattering in the field of a Coulomb center for $\Delta \ll m$, we must sum up (69) with the corresponding asymptotic form of the contribution of M_I (52). Since the expression obtained by us for M_I^C coincides with the result reported in [12], we can directly use formula (17) from [12]. Finally, we obtain

$$M^C(\Delta \ll m) = \frac{4i\alpha(Z\alpha)^2\omega}{m^2}\left[\int_0^1 dx\int_0^1 ds\int_0^1 dt F(s,t,x) + R\right],$$

$$R^{(++)} = \frac{7}{9} \times \left[\ln\left(\frac{2\omega}{m}\right) - \frac{i\pi}{2} - \frac{109}{42} - C - \mathrm{Re}\Psi(1 + iZ\alpha)\right], \quad (70)$$

$$R^{(+-)} = 0,$$

where functions $F(s, t, x)$ are defined in Eq. (69), $C = 0.577\ldots$ is the Euler constant, and $\Psi(y) = d\ln\Gamma(y)/dy$ is the logarithmic derivative of the gamma function. Expression (70) coincides with formula (8.1) from [8] if we correct the obvious misprints in this formula (eliminate the erroneous factor 4 in front of the integral in (8.1) and change the sign of the Coulomb correction $C + \mathrm{Re}\Psi(1 + iZ\alpha)$ in the nonintegral term).

Thus, the expression for the amplitude of the small-angle ($\Delta \ll \omega$) elastic scattering of a high-energy ($\omega \gg m$) photon that we obtained with the help of the semiclassical operator method for an arbitrary localized potential reproduces, among other things, all the available results in this field. The developed method of averaging the product of noncommuting operators makes it possible to apply the semiclassical operator method for studying various processes in QED in a localized external field.

ACKNOWLEDGMENTS

This work was partly supported by the Russian Foundation for Basic Research (project no. 00-02-18007).

REFERENCES

1. V. N. Baier and V. M. Katkov, Zh. Éksp. Teor. Fiz. **53**, 1478 (1967) [Sov. Phys. JETP **26**, 854 (1968)].
2. V. N. Baier, V. M. Katkov, and V. M. Strakhovenko, *Electromagnetic Processes at High Energies in Oriented Single Crystals* (World Scientific, Singapore, 1998).
3. W. Furry, Phys. Rev. **46**, 391 (1934).
4. V. B. Berestetskii, E. M. Lifshitz, and L. P. Pitaevskiĭ, *Quantum Electrodynamics* (Nauka, Moscow, 1980; Pergamon, Oxford, 1982).
5. Sh. Zh. Akhmadaliev, G. Ya. Kezerashvili, S. G. Klimenko, *et al.*, Phys. Rev. C **58**, 2844 (1998).
6. A. I. Milstein and M. Schumacher, Phys. Rep. **243**, 183 (1994).
7. M. Cheng and T. T. Wu, Phys. Rev. **182**, 1873 (1969).
8. M. Cheng and T. T. Wu, Phys. Rev. D **2**, 2444 (1970).
9. M. Cheng and T. T. Wu, Phys. Rev. D **5**, 3077 (1972).
10. A. I. Milstein and V. M. Strakhovenko, Zh. Éksp. Teor. Fiz. **85**, 14 (1983) [Sov. Phys. JETP **58**, 8 (1983)].
11. R. N. Lee and A. I. Milstein, Zh. Éksp. Teor. Fiz. **107**, 1393 (1995) [JETP **80**, 777 (1995)].
12. R. N. Lee, A. I. Milstein, and V. M. Strakhovenko, Zh. Éksp. Teor. Fiz. **116**, 78 (1999) [JETP **89**, 41 (1999)].
13. G. Z. Moliére, Z. Naturforsch. A **2**, 133 (1947).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 92, No. 4, 2001, pp. 576–593.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 4, 2001, pp. 665–684.

PLASMA, GASES

Simulation of Turbulence

L. N. Pyatnitsky
Institute of High Temperatures, Russian Academy of Sciences, Moscow, 127412 Russia
e-mail: piat7@mail.ru
Received September 4, 2000

Abstract—The linear theory of propagation of a spherical wave layer is used to treat the principles of simulating turbulent process and the criteria of its similarity. The mechanism of turbulence is given for a flat channel defined by two walls and for a square tube. We use this channel as an example to demonstrate the effect of the wave layer properties on velocity pulsations, as well as the effect of the characteristics of a sequence of disturbances on the structure of pulsations. The model is used to describe the structure of a plasma channel developed by a Bessel beam of laser radiation. The Mathematica-4 language is used for simulation. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

A turbulent flow is observed when the Reynolds criterion Re exceeds a certain critical value. Formally, this similarity criterion is derived from the equations of motion for liquid. Its physical meaning may be readily clarified using a simple example of laminar flow of viscous liquid along the x axis in a flat channel of height $z = d$ formed by two parallel walls of which one coincides with the xy plane.

A flux of momentum of liquid in the middle of the channel has the density $Q = \rho U_0^2$, where U_0 is the velocity and ρ is the density of liquid. The friction force develops a momentum flux in the direction of the z axis and forms the distribution of velocity $U = U_0[1 - (2z/d - 1)^2]$. The density q of this flux is defined by the xz component of the viscous stress tensor and is equal at the wall to

$$q = \sigma'_{xz} = \eta\left|\frac{\partial U}{\partial z}\right|,$$

where η is the coefficient of viscosity. Hence, $q = 4\eta U_0/d$. We will now take the ratio Q/q to derive, accurate within the numerical factor, the customary expression for the Reynolds criterion,

$$\frac{Q}{q} = \frac{\rho U_0 d}{\eta} = \mathrm{Re}.$$

The momentum transfer is always associated with the energy transfer; therefore, the quantities Q and q may be taken to mean the respective energy fluxes, and the velocity U_0 may be replaced by its effective value. These replacements change nothing but the numerical factor in the final result.

Therefore, the Reynolds criterion characterizes the fraction of momentum (or energy) flux that is lost by liquid due to friction or, to be more precise, its reciprocal. However, it appears impossible to recover a concrete mechanism from this characteristic. Indeed, under conditions of steady-state flow, the hydrodynamic parameters both in the foregoing estimates and in the solution of hydrodynamic equations are time-independent integrated quantities. On the contrary, pulsations prove to be explicit functions of time irrespective of the mode of flow, and, as regards pulsations, the steady state implies only dynamic equilibrium. At the same time, there is no doubt that it is the momentum flux q that sustains pulsations of parameters under conditions of turbulent flow. Approximately based on this reasoning, Lorentz [1] came up with an idea of the need for an energy-related estimation of the dynamics of pulsations whose behavior did not necessarily have to be described by equations pertaining to the flow proper. It has been demonstrated in [2] that the turbulence may be interpreted on the basis of linear propagation of disturbances, and a mechanism of the dynamics of pulsations of parameters under conditions of turbulent flow was suggested. In short, this mechanism reduces to the following.

(1) The stagnation of liquid causes a local increase in density (or pressure) at the flow boundary, which may be represented in the form of a sequence of weak disturbances of characteristic scale a that arise sporadically at the wall and propagate in the flow in all directions at the velocity of sound c. At a distance $r \gg a$, a disturbance, irrespective of its previous form, takes the form of a spherical layer of thickness $2a$.

(2) At $a \ll d$ (d is the transverse dimension of the channel) within this layer, $(r - a)/c < t < (r + a)/c$, and the hydrodynamic parameters pulsate in accordance with the distribution $P(r)$ of density ρ in the primary disturbance. For example, pulsations of velocity may be represented by the expression of the type

$$u = \frac{cP}{2\rho}\frac{r - ct}{ct}.$$

Pulsations in the layer decay with the distance $r = ct$, but they retain the form and properties of the function

P of the primary disturbance, which, generally speaking, may depend on the time *t*. Then, the linear dimension of disturbance is defined by the larger of the quantities *a* or *ct*.

(3) A collection of direct and reflected (from the channel walls) wave layers forms a space-time structure of pulsations of hydrodynamic parameters, which is formed on the basis of the linear superposition of waves.

This approach to the problem of turbulence is valid both for compressible liquid and for media such as a low-temperature plasma, in which acoustic waves exist. This approach may also be generalized to the case of waves of finite amplitude. To start with, however, it is natural to investigate the model of turbulence and identify its main regularities in a linear approximation; the more so that, under these conditions as well, pulsations of parameters of the medium arise, which are characterized by a wide spectrum of spatial and time frequencies, and their amplitude, phase, and direction prove to be irregular. This irregularity is usually interpreted as the stochasticity of pulsations. In reality, the turbulence is quite a determinate phenomenon. In turbulence, pulsations may be expressed in terms of functions that are relatively simple but depend on a number of parameters, for example, on the relative position of initial disturbances. Therefore, a collection of pulsations is similar to abstract automations, and its description falls in the category of cybernetic problems which are usually solved by imitation simulation methods. These methods assume the possibility of calculating the actual values of functions of the process, although, as in any cybernetic problem, it is permitted to add individual stochastic elements as an argument to the description of the process, along with determinate functions.

Our formulation of the problem calls for appropriate means for its solution. The desired capabilities are possessed by the Mathematica-4 interactive language of system simulation adapted for the actual description of parameters, functions, and correlations of arbitrary systems. In this language, the sequence of operations involved in solving the problem is written in the form of customary formulas, no abbreviations and acronyms are used, and, as a rule, no special comments are required. In addition, the language offers additional procedures for servicing computer experiments with easy control of the graphic representation of data including animation. The most important fragments of the model will be discussed further in this paper.

The paper deals with the mechanism of initiation of velocity pulsations and vortices and the similarity conditions for turbulence, as well as the principles of constructing the model of this phenomenon using the examples of a flat channel defined by two walls and of a square tube. A study is made into the effect of the characteristics of individual disturbances and their sequence on the properties and structure of pulsations of hydrodynamic parameters. A description is given of a model of experimentally observed structure of a plasma channel developed by a Bessel beam of laser radiation.

2. THE FACTORS AFFECTING THE PROPERTIES OF TURBULENCE, AND THE SIMILARITY CONDITIONS

The momentum (and energy) transfer due to viscosity is realized by way of propagation of spherical disturbance waves. Their superposition defines the properties of pulsations of hydrodynamic parameters in the channel. The wave amplitude during propagation decreases proportionally with a/r. The decay is largely due to the increase in the volume of the spherical layer, and the dissipation of energy plays no significant part at this stage. As it propagates and is reflected from the walls, the layer fills an increasingly larger volume of the channel, this leading to a reduction of the scale of the spatial structure of pulsations. And when the structure size becomes comparable with the molecular free path, the energy of disturbance of parameters simply passes over to the region of small-scale pulsations and causes a multiple increase in the importance of energy dissipation. This is how the known postulate of Kolmogoroff concerning energy transfer to small-scale turbulence finds a natural explanation.

On the flow boundary, the processes of interaction and the superposition of waves exhibit a more complex behavior, and we will discuss their singularities in more detail. It is known that the *z* component of velocity pulsations disappears on the wall, $u_z = 0$, and the pulsations of density and pressure decrease accordingly. In addition, one can readily see that, at the wall, the interaction between the incident and reflected waves results in a double increase in the amplitude of pulsations with a simultaneous double decrease in their frequency. In the direction of the *z* axis, these effects show up in a zone of thickness *a*. Along the wall, this zone has the form of a ring moving at a velocity of $c/\sin\theta$ (θ is the polar angle). The intensity of the processes occurring in this ring may be estimated proceeding from the constancy of the area of its wave front, $2\pi da$, and the *z* component of density pulsations is proportional to $\cos^2\theta$.

A third, side wave with a conical wave front takes part in the interaction at the wall, in addition to the incident and reflected spherical waves. The front has the form of the surface of a truncated right cone. The base of the cone is provided by the circle at the intersection of the refracted wave with the wall plane, and the small circle is formed by a line along which the conical surface is in contact with the reflected wave. The angle χ_w at the cone base is preassigned by the relation $\sin\chi_w = c/c_w$, where c_w is the velocity of sound in the wall, and the front moves at an angle of several degrees to the wall. The wave amplitude is low, of the order of ρ/ρ_w (ρ_w is the density of the wall), and decays very rapidly (as r^{-2}). However, as is clear from the configuration of reflection, this amplitude is formed by the peripheral part of the incident wave whose central part within the angles

$\theta < \theta_s = \pi/2 - \chi_w$ does not participate in the formation of the amplitude. In the immediate vicinity of the wall, this gives rise to a local nonuniformity of scale $\chi_w d$.

Another factor affecting the development and properties of turbulence is associated with the motion of liquid and formation of vortices. In treating this effect, we will use the notation and conditions of the example given in the Introduction and the expression for wave velocity $\mathbf{V} = \partial\omega/\partial\mathbf{k}$, where $\mathbf{k}$ is the wave vector, and $\omega(\mathbf{k})$ is the dispersion law. For waves in a liquid at rest, a simple dispersion relation $\omega = ck$ is valid. The motion of liquid may be included by adding to it an appropriate term, after which it takes the form

$$\omega = ck + \mathbf{U} \cdot \mathbf{k}. \quad (1)$$

We will use the concept of an "acoustic ray" of geometrical acoustics to assess the wave configuration. This ray is a line, the unit vector $\mathbf{s}$ of the tangent to which at point $\mathbf{r}$ coincides with the direction of propagation of the wave (and pulsations of parameters) at this point. The change of this direction depends on the ray curvature defined by the equation [3]

$$\frac{d\mathbf{s}}{dl} = \frac{1}{c}[\mathrm{rot}\,\mathbf{U} \times \mathbf{s}], \quad (2)$$

where dl is an element of the path traveled by the ray. When the liquid is stationary, $\mathbf{U} = 0$, the wave velocity in accordance with Eq. (1) is $\mathbf{V} = c\mathbf{k}/k$, i.e., the rays from the center of disturbance propagate at the velocity of sound along straight lines $\mathbf{s} = \mathbf{r}/r$, and the pattern retains complete spherical symmetry. If the liquid moves, but $\mathbf{U}$ is independent of the coordinates, the vector $\mathbf{s}$ is a constant quantity, and we derive for the wave velocity $\mathbf{V} = c\mathbf{k}/k + \mathbf{U}$. Then, in a moving coordinate system, the pattern of ray propagation is fully identical to the previous case, and for a fixed observer the wave proper is completely carried away by the flow without changing shape, and it is only the density of the energy of pulsations in the wave that depends on the ray direction (Doppler effect).

We will now turn to the case when the velocity of liquid in the channel cross section is profiled. We will assume that the flow is directed along the x axis and the velocity depends on only one variable, $U(z)$, and increases with the distance from the wall, so that, at $z = 0$, we have $U = 0$ and $dU/dz > 0$. One can see in Eq. (2) that in the $U(z)$ field the ray bends and the curvature increases with the velocity gradient. One must expect the maximum curvature at the wall, in the region of the highest gradient values. However, it is not so much the curvature that is of interest, but the total angle of turn of the ray, which may be estimated using Eq. (2). We express the components of the vector $\mathbf{s}$ in terms of the direction angles θ and φ, introduce the angle $\chi = \pi/2 - \theta$, and use the subscript 0 to indicate the initial orientations of the rays to derive for the component s_z, after integration of Eq. (2), ($U \ll c$),

$$\sin^2\chi \approx \sin^2\chi_0 - \frac{2U(z^*)}{c}\cos\chi_0\cos\varphi_0. \quad (3)$$

In accordance with this formula, the slope χ of the ray increases for the azimuth angles $|\varphi_0| > \pi/2$, while for $|\varphi_0| < \pi/2$ this slope decreases. In other words, in the gradient field the ray turns in a clockwise direction: it moves away from the wall when moving against the flow and approaches the wall when moving with the flow. Then, a fairly small slope χ_0 will be found, for which the ray that emerged from the point $z = 0$ will move away from the wall to a distance z^* or less to form a vortex. Therefore, in the $U(z)$ field, the direction of pulsations varies, and vortices arise, the condition of existence of which is given by the equality $\chi = 0$. In view of Eq. (3), this equality gives

$$\sin^2\chi_0 \approx \frac{2U(z^*)}{c}\cos\chi_0\cos\varphi_0. \quad (4)$$

This relation defines the value of z^* for the angle χ_0 with the preassigned profile of velocity $U(z)$. Note that the main parameter governing the vortex formation is the Mach number U/c. As is clear from Eqs. (2) and (4), the gradient and, accordingly, the velocity in the region of vortex formation must be fairly high. In the case of the slowly increasing function $U(z)$ over a distance $z \sim d$, no vortex may form at all, as in the case of $U = \mathrm{const}$.

By way of example, we will treat the power dependence of velocity in a channel on the ratio z/h, where h is the half-height, so that $z/h < 1$. We will represent this dependence and condition (4) in the form

$$U(z) = \frac{U_0}{c}\left(\frac{z}{h}\right)^{1/m}, \quad \frac{z^*}{h} = \left(\frac{c}{2U_0}\frac{\sin^2\chi_0}{\cos\varphi_0}\right)^m. \quad (5)$$

The region of most curvature of the velocity profile adjoins the wall. Because m is usually within the range of two to 10, the velocity reaches half its maximum value even at $z/h \sim 0.2$ to 0.02. As follows from formulas (5), in this region, for $m < 10$ and $U_0/c < 0.2$ and for angles $\chi_0 < 20°$ to 30°, the value of z^*/h is in the range from 0.01 to 0.2; i.e., formulas (5) further enable one to estimate the depth a of the zone of initiation of vortices. The configuration of this zone along the wall has the shape of a ring analogous to that described above, with the difference that the intensity of the processes in this case depends also on the azimuthal direction φ_0.

Substantial velocity gradients imply that gradients and other hydrodynamic parameters are present in this zone. Their level is defined by the dynamic equilibrium between the processes of stagnation and emission of waves from the stagnation region. The former processes lead to an increase in the gradients, and the latter processes lead to equalization of the parameters. Against this background, the interaction between sound waves in the vicinity of the wall and the dependence of the effects of this interaction on the gradient profile, on

the angle of incidence, and on other factors identified above are responsible for the emergence of local nonstationary nonuniformities of parameters along the wall surface. These nonuniformities derive energy from the liquid upon its stagnation and serve as the sources of new disturbances. Their reproduction closes the cycle of the process of maintaining the dynamic equilibrium of pulsations. This process is determinate and defined by the Mach number, channel geometry, and properties of the liquid. At the same time, similar to the superposition of waves in the main flow, it is a multiparameter one and develops in a wide frequency band. Therefore, the pattern of initiation of disturbances appears to be irregular and resembling a chaotic pattern both in space and in time (when performing observation at a fixed point). Note that the formation of wall disturbances begins even during the period of build-up of flow, when close-to-spherical waves of starting momentum, which set the liquid in motion, propagate in the liquid and are reflected from the walls; i.e., disturbances and nonuniformity of parameters at the walls arise from the very start.

Therefore, the profile of parameters in the channel cross section and the structure of pulsations are related to one another by momentum transfer, whose mechanism is realized via sound wave emission. It is natural to expect that the real value of momentum flux q is of importance, in which the losses due to propagation of disturbances are included. Of these losses, the absorption of the wave energy by the liquid has no serious effect on the formation of pulsations, as was noted above. However, upon reflection from the wall, the losses may be significant, because a part of the energy goes away with the refracted and side waves and is transmitted to the wall in the form of heat. The other part, possibly a more substantial one, is lost upon emergence of vortices.

Generally speaking, the reflection coefficient is a function of the angle of incidence. However, the momentum flux q is among the averaged parameters of flow, and integrally the losses upon each reflection may be estimated by the constant coefficient α. Then, with n reflections, the real momentum flux q will decrease by a factor of $(1-\alpha)^n$, and the flow crisis will occur at higher values of the Reynolds number Re_{cr}. On the contrary, the wall roughness intensifies the flux q, and the turbulence arises at lower values of Re_{cr}, i.e., at a lower velocity. Because the parameters α and n vary in a wide range, depending on the concrete conditions [2], the losses upon reflection do not remain constant, and this factor must be included when estimating the flow mode.

The explicit time dependence of pulsations calls for the introduction of an additional criterion for describing the turbulence. The process of wave propagation is of importance in the formation of the structure of disturbances and emergence of vortices. In this case, as for any motion, the scale of displacement Δ and the corresponding time interval τ are important. Such processes are usually characterized [3] by the dimensionless Strouhal criterion $S = Ut/\Delta$. In this case, however, we deal with the propagation of sound, and the equality $\Delta/t = c$ is valid. Therefore, $S = U/c = M$, and the Strouhal number may be replaced by the Mach number; as was shown above, it is this latter number on which the processes of structurization and vortex formation in the channel depend. Consequently, in analyzing the turbulence, the similarity must be set with the aid of two dimensionless criteria, those of Reynolds and Mach,

$$\mathrm{Re}^* = \frac{\mathrm{Re}}{(1-\alpha)^n} = \frac{\rho U_0 d}{\nu(1-\alpha)^n}, \quad M = \frac{U_0}{c}, \qquad (6)$$

where ν is the kinematic viscosity.

The foregoing treatment is based on the linear theory of propagation of wave packets of spherical shape. A comparison of the processes accompanying their propagation in a narrow zone on the flow boundary and in the main part of flow, in the flow core, reveals important differences in the behavior of interaction and in the results of superposition. This gives grounds to isolate a narrow zone of width a at the wall into a special region, i.e., the boundary layer. Outside of this layer, the velocity gradient is low, and the function $U(z)$ increases slowly. Then, in a first approximation, this variation may be ignored, and, in estimating the superposition of waves in the flow core, one can use the condition $U = \mathrm{const}$ simplifying the problem, where the constant may be zero.

3. PROPAGATION AND STRUCTURE OF DISTURBANCES IN THE CHANNEL

The structure of disturbances in the channel depends on the shape of the bounding surface. In typical simple options, one can regard as a channel a space above a plane (the case of a flat plate), a plane gap formed by two parallel walls, and the internal space of a tube. The structure of disturbances developed by spherical waves emanating from the wall surface is obvious, and its determination is trivial. The situation is somewhat more complicated in the case of channels in which reflected waves affect the structure. We will first dwell on a channel shaped as a gap of height d between two parallel walls.

We will bring the xy plane into coincidence with the lower wall and direct the z axis toward the upper wall. It is natural to take the channel height d as the scale defining the processes in the channel. Then, the position of a point in the channel will be expressed in conventional units of length corresponding to dimensionless coordinates for which the previous designations x, y, and z are used. For timing the process, we use the dimensionless unit of time τ during which a wave passes a distance d at a velocity c so that $\tau = ct/d$. A disturbance of dimensionless radius a, which emerged at some moment of time at an arbitrary point on the channel walls, will be characterized by the parameters p_τ and $p_\mathbf{r} = \{p_x, p_y, p_z\}$ which, for a series of disturbances, will represent sequences. The parameter p_τ may be both positive and negative, and the parameter p_z takes the values of 0 (on the lower wall) or 1 (on the upper wall).

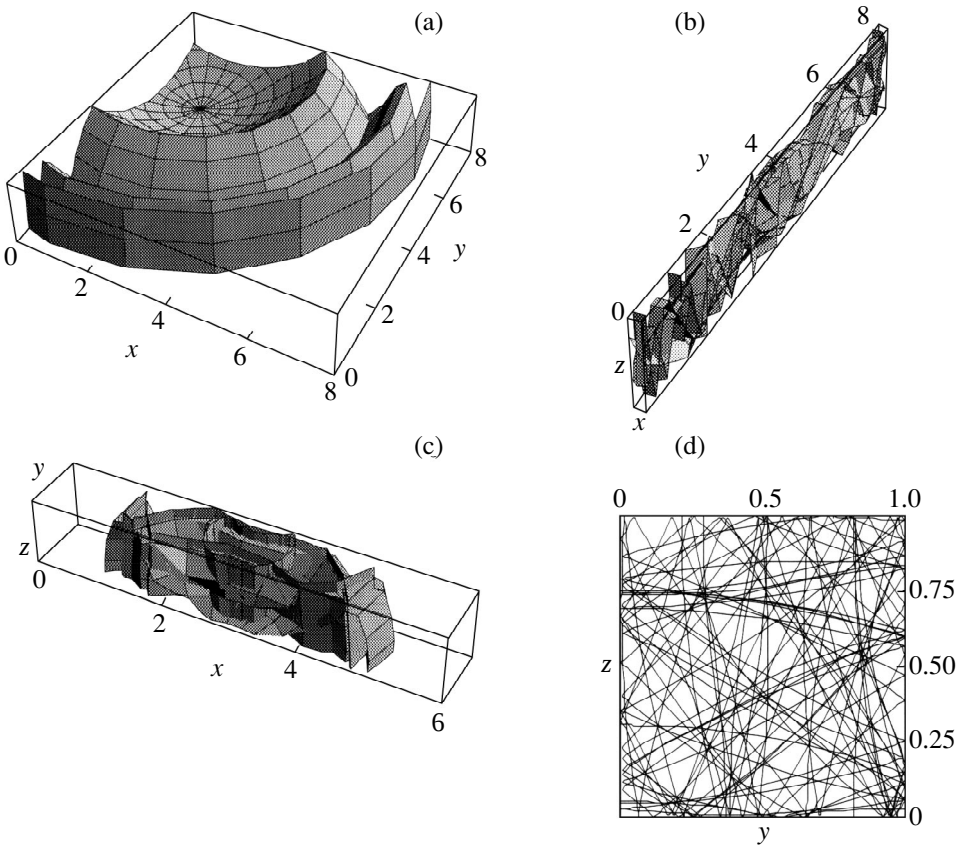


Fig. 1. The structure of disturbances in a bounded space: (a) configuration of a wave in a flat layer, (b) configuration of eight waves in a layer, (c) configuration of a wave in a tube, (d) configuration of eight waves in a tube.

When a wave propagates from the point $p_{\mathbf{r}}$ in the half-space that is not bounded by the upper wall, $z > 0$, the position of the wave front in the coordinate system related to the point $p_{\mathbf{r}}$ will be defined by the radius vector $\mathbf{r}(\xi, \eta, \zeta)$ and $|\mathbf{r}| = \tau$. Every event of reflection of the wave front will deform the latter. As a result, the front will acquire a more complex configuration. For $\tau + p_\tau > 0$, its coordinates x, y, and z will be defined by the relations

$$x = (\tau + p_\tau)\sin\theta\cos\varphi + p_x,$$

$$y = (\tau + p_\tau)\sin\theta\sin\varphi + p_y, \qquad (7a)$$

$$z = \left((\tau + p_\tau)\cos\theta - 2\mathrm{IP}\left[\frac{n+1}{2}\right]\right)(-1)^{n+p_z} + p_z.$$

For $\tau + p_\tau \le 0$, we have

$$x = p_x, \quad y = p_y, \quad z = p_z. \qquad (7b)$$

Here, IP (Integer Part) denotes the operation of separating out the integer part of the number in square brackets, and n denotes the number of wave reflections from the wall (positive integer),

$$n = \mathrm{IP}[(\tau + p_\tau)\cos\theta]. \qquad (8)$$

We will now find the surface of the wave front—a snapshot of sorts of the configuration of the front assumed by the front in time τ after initiation. Note that the procedure of solving the problem in the Mathematica-4 language simply reproduces the mathematical solution of the problem both in substance and in regards to the notation. First, the input data are described (p_τ, p_x, p_y, p_z), and the moment of time τ is preassigned, for which the shape of the wave front and the dimensions of the region being investigated (dx, dy, dz) are calculated. Further, formulas (7) and (8) are used to find three coordinates of the front at the preassigned moment of time τ, and then the conditions of graphic construction of the front surface are determined. (Usually indicated here are the ranges of variation of the variables, the angle of observation, the shadowing of surface areas, the labeling of coordinate axes, and other parameters.)

For definiteness, we will assume that the initial point is on the lower wall, and its coordinates and the investigation range have the following values:

$$\tau = 6, \quad p_\tau = 0, \quad p_x = 1.2, \quad p_y = 6.3,$$
$$p_z = 0, \quad dx = 8, \quad dy = 8, \quad dz = 1.$$

The surface is calculated with respect to the angle φ in the range from 0 to 2π and with respect to θ in the range

from 0 to $\pi/2$. The limits for variables x, y, and z correspond to the selected volume; however, in constructing a graph, the scale of z is enlarged by a factor of one-and-a-half for convenience of visual perception.

The configuration of this wave front is given in Fig. 1a. It has the form of a corrugated surface with a variable distance between the lines of bend and with a varying orientation of the normal to the surface of folds. Therefore, the propagation of only one wave develops a structure in the channel, the simple regularity of whose origin is hard to recognize immediately. Figure 1b shows what the space of a flat layer looks like under conditions of simultaneous propagation of eight waves whose initial parameters are varied in the following ranges:

$$p_\tau \in [0, 3.1], \quad p_x \in [0.2, 15.9],$$
$$p_y \in [0.1, 9.9], \quad p_z = \{0, 1\}.$$

For the inner details of the structure to be observed, it is not the entire channel region being investigated that is shown in the graph but rather a strip cut out from this region (rectangular parallelepiped) and having a width $x \in (4.8, 5.0)$ with the above-identified values of the limits of the remaining parameters ($y = 8$ and $z = 1$). The combination of waves gives a clear idea of the mechanism of formation of the structure of disturbances as a result of propagation of sound waves in a bounded space. Even the pattern of their interaction enables one to judge the behavior of the structure of disturbances.

A layer defined by two parallel walls is most convenient for applying this model. Indeed, in a channel like this, the disturbance of parameters at the observation point may be represented as a successive passage of spherical waves through this point and its virtual images which are its mirror reflections from the walls. Moreover, the position of an imaginary point is defined by only one coordinate which, on every reflection, changes to the double height of the layer $2d$ (in dimensionless coordinates, it changes to 2).

This simple useful rule may be readily extended to the case of a channel whose boundaries have the form of a tube with flat faces. For such a channel, the calculation scheme remains the same; however, because of the wave reflection from several walls, the number of imaginary points to be included increases. For example, in a square tube, the number of such points triples in every cycle of reflections, with the position of the imaginary point being defined by two, instead of one, coordinates which vary with the period $2d$ (here, d is the side of a square). We will treat this mechanism in more detail and find the wave configuration at the moment of time τ.

The wave configuration is not related to the place of its initiation in the direction of the x-axis of the tube and depends only on the coordinates along the y and z axes. On the faces parallel to the y-axis (horizontal walls, $z = 0$ or 1), we will denote these coordinates as p_{yy} and p_{yz}, and on the faces parallel to the z-axis (vertical walls, $y = 0$ or 1), as p_{zz} and p_{zy}. The values of p_{yy} and p_{zz} vary in the range from 0 to 1, while the values of p_{yz} and p_{zy} may only be discrete, 0 or 1.

We will now determine the number of reflections of a wave moving from one of the faces. We will distinguish between the cases of wave propagation in the direction from horizontal ($z = 0$ or $z = 1$) and from vertical ($y = 0$ or $y = 1$) faces. Each direction presumes reflections from the other three faces. In the case of propagation from the vertical face $z = 0$ (or $z = 1$), we will denote the number of reflections from the walls $y = 0$, $y = 1$, $z = 1$ (or $z = 0$) as n_{y0}, n_{y1}, and n_{yz}, respectively. For the direction of the wave motion from vertical faces, these designations may be written as n_{z0}, n_{z1}, and n_{zy}. By analogy with formula (8) for a flat layer, the number of wave reflections from different faces of the tube may be given by relations (9) in which, as previously, τ is the running time (in dimensionless units); the parameters $p_{\tau y}$ and $p_{\tau z}$ allow for the delay or, the other way around, for the advance of the wave initiation relative to the moment of time $\tau = 0$ for horizontal and vertical faces, respectively; and θ is the polar angle, $\theta \in [0, \pi]$.

For a horizontal face, $z = 0$ or $z = 1$,

$$\begin{aligned} n_{y0} &= \mathrm{IP}[(\tau + p_{\tau y})\sin(\theta - \pi/2) + p_{yy} - 1], \\ n_{y1} &= \mathrm{IP}[(\tau + p_{\tau y})\sin(\theta - \pi/2) + p_{yy}], \\ n_{yz} &= \mathrm{IP}[(\tau + p_{\tau y})\cos(\theta - \pi/2) + p_{yz}]; \end{aligned} \tag{9a}$$

for a vertical face, $y = 0$ or $y = 1$,

$$\begin{aligned} n_{z0} &= \mathrm{IP}[(\tau + p_{\tau z})\cos\theta + p_{zz} - 1], \\ n_{z1} &= \mathrm{IP}[(\tau + p_{\tau z})\cos\theta + p_{zz}], \\ n_{zy} &= \mathrm{IP}[(\tau + p_{\tau z})\sin\theta + p_{zy}]. \end{aligned} \tag{9b}$$

As a result, three coordinates corresponding to three waves reflected from three different faces of the tube space will emerge instead of each one of the coordinates y and z used in formulas (7) to describe the position of a wave reflected from the opposite end of a flat channel. In accordance with the notation adapted for reflected waves, we will introduce coordinates $y0$, $y1$, and zy instead of y and coordinates $z0$, $z1$, and yz instead of z. The relations defining their values are in many ways similar to formulas (7); therefore, in order to reduce the computations, we will treat waves in a single cross section of the tube, assuming that $\varphi = \pi/2$,

$$\begin{aligned} y0 &= ((\tau + p_{\tau y})\sin(\theta - \pi/2) + p_{yy} \\ &\quad - 2\mathrm{IP}[n_{y0}/2])(-1)^{n_{y0}}, \\ y1 &= ((\tau + p_{\tau y})\sin(\theta - \pi/2) + p_{yy} \\ &\quad - 2\mathrm{IP}[(n_{y1} + 1)/2])(-1)^{n_{y1}}, \end{aligned} \tag{10}$$

$$
\begin{aligned}
zy &= ((\tau + p_{\tau z})\sin\theta + p_{zy} \\
&\quad - 2\mathrm{IP}[(n_{zy} + 1)/2])(-1)^{n_{zy}}, \\
z0 &= ((\tau + p_{\tau z})\cos\theta + p_{zz} \\
&\quad - 2\mathrm{IP}[n_{z0}/2])(-1)^{n_{z0}}, \\
z1 &= ((\tau + p_{\tau z})\cos\theta + p_{zz} \\
&\quad - 2\mathrm{IP}[(n_{z1} + 1)/2])(-1)^{n_{z1}}, \\
yz &= ((\tau + p_{\tau y})\cos(\theta - \pi/2) + p_{yz} \\
&\quad - 2\mathrm{IP}[(n_{yz} + 1)/2])(-1)^{n_{yz}}.
\end{aligned}
\tag{11}
$$

In constructing the configuration of a wave, the position of a point on its front is defined by the coordinates $y0 - yz$, $zy - z1$ at $\theta < \pi/2$ and by the coordinates $y1 - yz$, $zy - z0$ at $\theta \geq \pi/2$.

Figure 1c shows the configuration of a wave during its propagation in a 1×1 square tube. The wave position is registered at the moment of time $\tau = 2$ after its initiation on the lower wall at the point $\{p_x, p_y, p_z\} = \{2, 0.6, 0\}$. One cannot but note that the structure of disturbances became irregular as a result of only two incomplete cycles of reflections of only one wave. It is obvious that the development of turbulence in a tube is markedly more intensive, as is seen in Fig. 1d, where this process is demonstrated using the example of propagation of eight waves (as for a flat layer in Fig. 1b). Because, in this case, it appears impossible to distinguish the details of configuration by its image in three-dimensional space, Fig. 1d shows its trace in the tube cross section ($x = 4$) at the moment of time $\tau = 3$. The distribution of initial disturbances on the wall surface and in time is preassigned by the law of chance.

Both in a flat channel and in a tube, the state of the structure of disturbances at late stages appears fully chaotic at first glance, and the structure cells that are formed are characterized by a complex distribution with respect to size and shape, so that the entire pattern makes an impression of turbulence in the channel. Proceeding from these examples and in view of the fluctuations of parameters in every single wave layer, one can understand how the pulsations are formed and what their structure is. The effect of intermittency, which is usually explained by the instability of the boundary layer, also finds a natural interpretation.

As was already noted, it is more convenient to analyze the processes for a flat layer, because this enables one to demonstrate all of the important singularities of the phenomenon and possibilities offered by the concept without resorting to cumbersome mathematical expressions in describing the geometry of a wave reflected from a complex surface, which would cause the paper to expand considerably. However, in this case as well, before turning to the analysis of pulsations, one must treat the problem of the site and sequence of initiation of a series of k disturbances. In fact, we have already come up against this problem. At $k = 1$, the initial parameters of disturbances were selected simply from considerations of obviousness, while at $k = 8$ they were preassigned by the law of random numbers.

The use of this law in the problem at hand involves certain special features. The sampling of random values of a parameter presumes a uniform distribution of probability in the preassigned range. A property that is common to any random number generator is the variation of the internal state on every reference to it. Therefore, repeated references will cause the generator to produce different numbers. The eventual outcome is clear from expressions (7) and (8). Assume that, as a result of the first reference, the generator gave a random value of some parameter, say, p_τ. If this value varies on subsequent requests, the wave coordinates will turn out to be random, and it will be impossible to describe any successive propagation of the wave. However, for other waves, one must find other random values which must also be reproduced during their lifetime. These requirements hold for the random values of the parameters p_x, p_y, and p_z as well.

We will treat this problem using the example of the Mathematica-4 system. Let the parameter p_τ vary in the range of $d\tau = \tau_2 - \tau_1$. We will set the generator to the SeedRandom [s] state, where s is an integer equal to the running time of the day (in fractions of a second). Then, the operation

$$p_\tau = \mathrm{Table}[(\tau_2 - \tau_1)\mathrm{Random}[], \{k\}]$$

will develop in the τ_2–τ_1 range a sequence p_{τ_k} of k constant numbers selected at random, which will be reproduced on every request of the value of p_{τ_k}. The random values of the initial coordinates of all k waves arising in the ranges $dx = x_2 - x_1$, $dy = y_2 - y_1$, and $dz = \{0, 1\}$ will be defined in the same manner.

The randomness of place and time of initiation of disturbances is but one of the options of distributions of p_τ and $p_{\mathbf{r}}$. Disturbances may arise at some set of fixed points associated, for example, with microprojections on the wall surface. Given this nature of disturbances, both a periodic (to be more precise, quasi-periodic) and a random behavior of the process in time is permissible. Finally, note that the preassigned channel size and wave velocity set up conditions under which the reflection of waves in a certain spectral range becomes preferable. However, as was demonstrated above, the reflections bring into action the mechanism of formation of disturbances whose position is independent of the singularities of the wall. In this case, disturbances will arise almost periodically. Different scenarios are apparently realized in experiments. And each one of those scenarios must affect, in its own manner, the structure of pulsation of parameters and, first of all, their spectra. At the same time, any regularity of the emergence of disturbances may be represented by appropriate expres-

sions for the parameters p_τ and $p_{\mathbf{r}}$, as in the case of random sequences. Then, the results of comparison of experimentally obtained and model spectra will help find this sequence and, along with it, the mechanism of the process effective in a concrete experiment. We will return to this problem in what follows; at this stage, we will investigate pulsations of velocity upon propagation of a single wave in the channel, and then the structure of pulsations for a series of waves.

4. PULSATIONS OF VELOCITY DURING PROPAGATION OF DISTURBANCE

We will determine the time dependence of velocity at an arbitrary point $\{x, y, z\}$ of a channel. It will be recalled that it is other than zero only inside a layer of thickness $2a$; therefore, pulsations at the point $\{x, y, z\}$ will arise only at the moment of arrival to this point of the leading wave front and will cease when the trailing wave front leaves this point. However, both the forward wave and all reflected waves pass through the above-identified point in the channel. Therefore, the velocity at that point will pulsate during the entire time of observation, and all waves must be included in calculating this velocity. We will judge the properties of these pulsations by the spectra which may be constructed using the Fourier transform of the time dependence of velocity.

Let the radius vector $\mathbf{r}_n(\xi, \eta, \zeta)$ of the wave front in free space, $|z| \geq 0$, connect the wave center p_x, p_y, p_z to the imaginary point corresponding to n reflections. At this point, pulsations of velocity would be observed on condition that $|r_n - \tau| < a$, where r_n is the distance to this point and τ is the actual radius of spherical wave in the free half-space. The velocity profile in the layer, $u(r_n - \tau)$, is formed by the distribution $P(r)$ of density in the initial disturbance of radius a, with its shape maintained during propagation, while the amplitude decreases both with time and on every reflection of the wave from the wall.

We will represent the distribution $P(r)$ in the form of the product of the constant P by some function, for example, cosine, with the argument $(r_n - \tau)/a$. For the velocity u at the point of space $\mathbf{r}_n$, we have

$$u = \begin{cases} 0, & |r_n - \tau| > a \\ B(1-\alpha)^{n(\tau)} \cos\left(\dfrac{\pi}{2}\dfrac{r_n - \tau}{a}\right)\dfrac{r_n - \tau}{\tau}, & |r_n - \tau| < a. \end{cases} \tag{12}$$

Without loss of generality, the coefficient $B = cP/2\rho$ may be taken to be equal to unity. We will express the components ξ, η, and ζ of the radius vector $\mathbf{r}_n$ in terms of the observation point coordinates x, y, and z in the related frame of reference. We will take into account the fact that two of these components, ξ and η, are not affected by the number of reflection, and they remain constant for any values of n, while the third component, ζ, depends both on n and on the wall on which the wave arises. In determining this correlation, we will use the fact that conjugate or imaginary points correspond to the observation point at the moments of passage of reflected waves through this point. We use the above-identified regularity of their position to find the correlation between the coordinates in two frames of reference. The observation point $\{x, y, z\}$ is known. Because this is a real point, one can immediately write the equalities $\xi = x$, $\eta = y$, and $\zeta = z$. The first imaginary point is a mirror reflection of the observation site in the wall plane and, for this point, $\xi = x$, $\eta = y$, and $\zeta = 2 - z$. Each subsequent imaginary point shifts to the distance $\delta\zeta = 2$ (dimensionless units). Then, Eqs. (7) yield

$$\xi = x - p_x, \quad \eta = y - p_y,$$
$$\xi_n = 2\mathrm{IP}\left[\frac{n+1}{2}\right](-1)^{p_z} + (z - p_z)(-1)^n, \tag{13}$$
$$r_n = \sqrt{\xi^2 + \eta^2 + \zeta_n^2},$$

where n is given by expression (8).

Formulas (12) and (13) enable one to simulate the distribution and properties of disturbances in a flat channel. We will treat this simulation in more detail, including the sequence of operations for determining pulsations of the velocity u and the spectrum of these pulsations during propagation of a single wave in the channel. In accordance with the adapted procedure, one must first indicate the values of parameters that preassign the process conditions, i.e., the input data. For example, for the size of initial disturbance we use the value $a = 0.1$ (a fraction of the separation between the walls), estimate the reflection losses as $\alpha = 0.1$, and for the observation site we select the coordinates $\{x, y, z\} = \{0.3, 0.5, 0.7\}$.

Because the scatter of the beginning and site of initiation for a single wave is of no importance, the distributions of the initial parameters p_τ and $p_{\mathbf{r}}$ and the ranges of their variation $d\tau = \tau_2 - \tau_1$ and $d\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$ may be eliminated from treatment. Instead, we will place the initial disturbance on the upper wall at the point with zero values of the remaining parameters. In describing pulsation, it is further important to preassign the maximum number $n_{\max}$ of wave reflections from the walls and the time T during which the process is analyzed. We will include 10 reflections. Accordingly, the duration of observation of the process will be restricted to $T = 10$. Then, the initial conditions will have the form

$$a = 0.1, \quad \alpha = 0.1, \quad \{x, y, z\} = \{0.3, 0.5, 0.7\},$$
$$\{p_\tau, p_x, p_y, p_z\} = \{0, 0, 0, 1\}, \quad n_{\max} = 10, \quad T = 10.$$

First of all, formulas (13) are used to find the coordinates $\{\xi_n, \eta_n, \zeta_n\}$ of imaginary points in the half-space, as well as the distances $\mathbf{r}_n$ to these points from the selected observation site $\{x, y, z\}$ and the numbers n of wave reflection.

Then, in accordance with formulas (12), velocity pulsations $u(\tau, n)$ are calculated at each passage of the wave

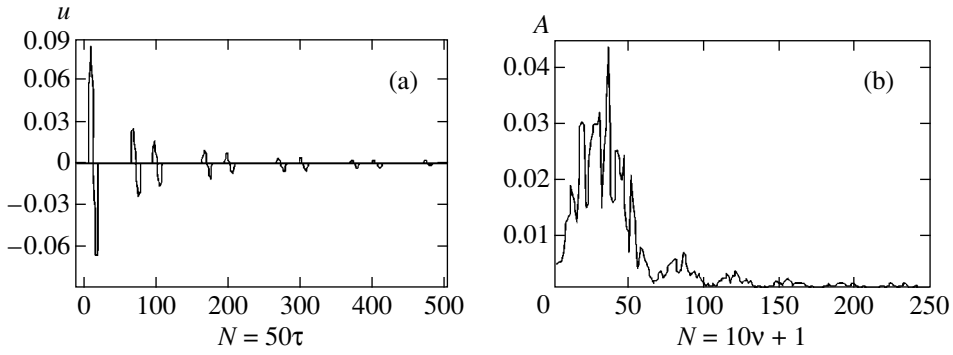


Fig. 2. (a) Velocity pulsations and (b) their spectra in a layer during propagation of a single wave: $a = 0.1$, $\alpha = 0.1$, $n = 10$, $\{x, y, z\} = \{0.3, 0.5, 0.7\}$, $\{p_\tau, p_x, p_y, p_z\} = \{0, 0, 0, 1\}$.

layer past the observation point $\{x, y, z\}$ located inside the channel. At the same time, these pulsations are added up to form a unified time dependence of velocity at the observation point in the form of the piecewise smooth function $u(\tau)$.

In conclusion, technical (but important) procedures are performed. Here, the function $u(\tau)$ is represented in the form of a table and is then used to construct graphs for velocity pulsations and for the spectrum of these pulsations. The spectrum (Fourier transform) calculation procedure involves breaking the observation time T into segments (resolvable elements) whose number N defines the spectral resolution and the range of frequencies ν of the spectrum. It is in the table that the values of T and N are found. In the case of single wave propagation being treated, at $T = 10$ it is quite sufficient to assume that $N = 500$. Then, the number of intervals N, the time τ (in dimensionless units), and the dimensionless frequency ν (in pulsations per unit time) will be related as $N = 50\tau$ and $N = 10\nu + 1$. The dependences of $u(\tau)$ (to be more precise, of $u(N)$) and of the amplitudes of the pulsation spectrum $A(\nu)$ (or $A(N)$) are given in Fig. 2. We use the example of a single disturbance to find out which parameters and how affect the spectrum of pulsations in the channel. Obviously, as regards the initial parameters $\{p_\tau, p_\mathbf{r}\}$, the wave start delay, p_τ, has no effect on the form of spectrum, while the removal of the source away from the observation site, $p_\mathbf{r}$, must deform the velocity pulsations and their spectrum, which is associated with the properties of the function r^{-1} and with the variation of the angle of reflection. The velocity u in the wave layer is proportional to τ^{-1}; therefore, the observation time T (or the number of reflections $n_{\max}$) governs the minimum amplitude of pulsations and, consequently, the high-frequency part of the spectrum that requires a respective number of resolvable elements N. Generally speaking, the wider the spectral band, the more accurately the model describes the real process; however, at some level, the accuracy becomes redundant. In the example being treated, at $a = \alpha = 0.1$ and $T = 10$, the amplitude of pulsations decreases by a factor of almost 300 during the time of observation. For our purposes, pulsations of a lower amplitude may be ignored.

Less obvious is the importance of the size a of initial disturbance, of the distribution P of parameters in the source, and of the loss factor α during wave reflection from the wall. Some preliminary information about the effect of these factors may be provided by the representation of function (12) in the integral form of the Fourier transform. Without going into details of the procedure of including function (12) into known expressions, note that, upon integration, two cofactors of this function would produce a relation in the form of sine integral si. In the low-frequency spectral range, it is equal to $\pi/2$; however, as the frequency increases, si is approximated by an expression of the type of $\cos\psi/\psi$, and the frequency ν is replaced in our case by the product $a\nu$. That is, the spectrum in some or other degree will be modulated, and the frequency band must increase with decreasing a. According to formulas (12), corrections into the shape of spectrum are introduced by the form of the function of velocity distribution P in the wave layer, as well as by the loss factor α which governs the jump of velocity pulsations upon wave reflection from the wall. In order to estimate this effect, we will use formulas (12) and (13) and the foregoing list of initial parameters and will successively, one by one, vary the values of the parameters a, P, and α, with the other parameters remaining as given in the list. The calculation results are given in Fig. 3, where the variations of the parameter $a \longrightarrow \{0.2, 0.1\}$ are represented by spectra *1* and *2*, respectively. The distribution P (see Eq. (12)) was given by the functions $P_3 = 1 - |r_n - \tau|/a$ and $P_4 = 1 - (|r_n - \tau|/a)^4$, as represented by spectra *3* and *4*. Spectra *5* and *6* correspond to the loss factors $\alpha \longrightarrow \{0.1; 0.5\}$. The vertical lines at $10\nu + 1 = 200$ indicate the sine curves added to the velocity u, with the amplitude g and frequency ν of each one of those sine curves giving an idea of the characteristics of pulsations in absolute magnitude.

It follows from graphs *1* and *2* in Fig. 3 that the frequency of the maximum of the first harmonic is defined by the width of spherical wave layer and equal to $1/2a$.

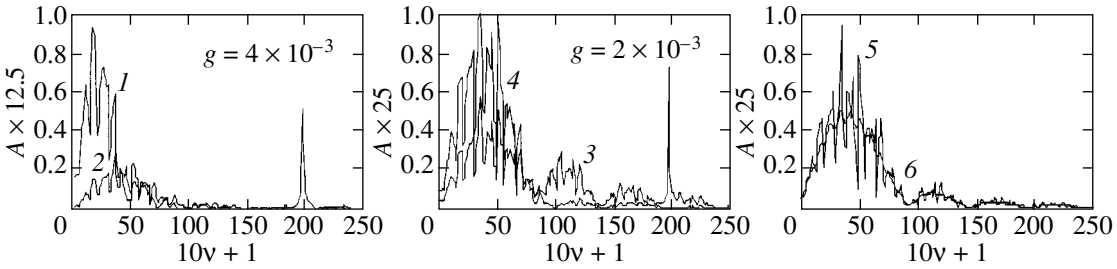


Fig. 3. Spectra of pulsations at the point $\{x, y, z\} = \{0.3, 0.5, 0.7\}$ with varying disturbance parameters: (*1*, *2*) $a = \{0.2, 0.1\}$; (*3*, *4*) $P = \{1 - |r_n - \tau|/a,\ 1 - (|r_n - \tau|/a)^4\}$; (*5*, *6*) $\alpha = \{0.1, 0.5\}$.

Note the uniformity of modulation of all spectra. The estimates reveal that, starting with the second harmonic, the modulation frequency remains a constant quantity for the given value of a; with a varying from 0.01 to 0.2, its ratio to the basic frequency and the product $a\nu_m$, where ν_m is the frequency of harmonic m, are virtually invariable. Hence it follows that the entire spectral band is inversely proportional to the size of primary disturbance, as is in fact demonstrated by curves *1* and *2* in Fig. 3. In so doing, the spectral density decreases, and a comparison of the amplitudes reveals the extent of this decrease.

The function P likewise affects the spectrum of pulsations. It follows from a comparison of curves *2*, *3*, and *4* (Fig. 3), where, respectively,

$$P \propto \cos\frac{\pi}{2}\frac{|r_n - \tau|}{a}, \quad P \propto 1 - \frac{|r_n - \tau|}{a},$$

and

$$P \propto 1 - \left(\frac{|r_n - \tau|}{a}\right)^4,$$

that this effect is analogous to the spectrum of the function P proper and, the closer the distribution P to rectangular pulse, the stronger the development of higher-order harmonics. Therefore, in estimating the effect of the form of function P on the pulsation spectrum, one can be guided simply by the properties of the Fourier transform of this function.

The importance of the loss factor is clear from a comparison of curves *5* and *6* (Fig. 3), where the values of α are 0.1 and 0.5. For a relatively low level of losses, a rapid oscillations are observed in spectrum *5* along the entire spectral curve. As the losses increase, this oscillation levels off, and spectrum *6* becomes smoother and follows the averaged amplitudes of spectrum *5*.

We treated the effect of various factors on the spectral properties of velocity pulsations, when a single disturbance in the form of a spherical wave layer propagated in a channel formed by two parallel walls. The width of the wave layer was equal to the diameter of initial disturbance. The amplitude-frequency characteristic, or pulsation spectrum, depended on the position of the wave initiation point relative to the observation point rather than on the moment of its initiation.

Proceeding from these results, one can expect that a sequence of waves initiated at one and the same site will produce the same spectrum irrespective of the behavior of this sequence. The frequency band of the spectrum will depend on the number of reflections taken into account; however, if the number of reflections is preassigned, the frequency band will be defined by the wave layer thickness. The spectrum must contain harmonics whose half-space and period of arrangement will be defined by the same wave layer thickness, and the damping decrement and the number of harmonics will be defined by the form of the distribution of parameters in the initial disturbance. An increase in the loss factor upon wave reflection from the wall must cause a decrease in the amplitude of high-frequency modulation of each harmonic of the spectrum and smooth out the spectral characteristic.

5. STRUCTURE OF PULSATIONS FOR A SEQUENCE OF DISTURBANCES

Given the properties of pulsations during propagation of a single wave (or waves issuing from a single point) in a channel, one can proceed to simulation of turbulence for an arbitrary sequence of disturbances; that is, in determining the behavior of velocity pulsations at the observation point, one must include the contribution made by disturbances originating from different sites (on the walls) and at different moments of time. Of real importance are the versions of initial conditions when a series of disturbances arises at random or fixed points p_r, and the moments of time p_τ may be both random and periodic, as well as quasi-periodic. Turbulence experiments usually involve recording velocity pulsations at a fixed point of the channel or taking a snapshot of distribution of disturbances in the channel cross section. Therefore, in the case of propagation of a series of disturbances in a channel, one needs, in addition to a spectrum at a point, which depends on the frequency of pulsations per unit time, ν, to simulate the structure of pulsations, i.e., their spectrum over the height z and along the channel, for example, on the x axis, with a spatial frequency ν expressed in pul-

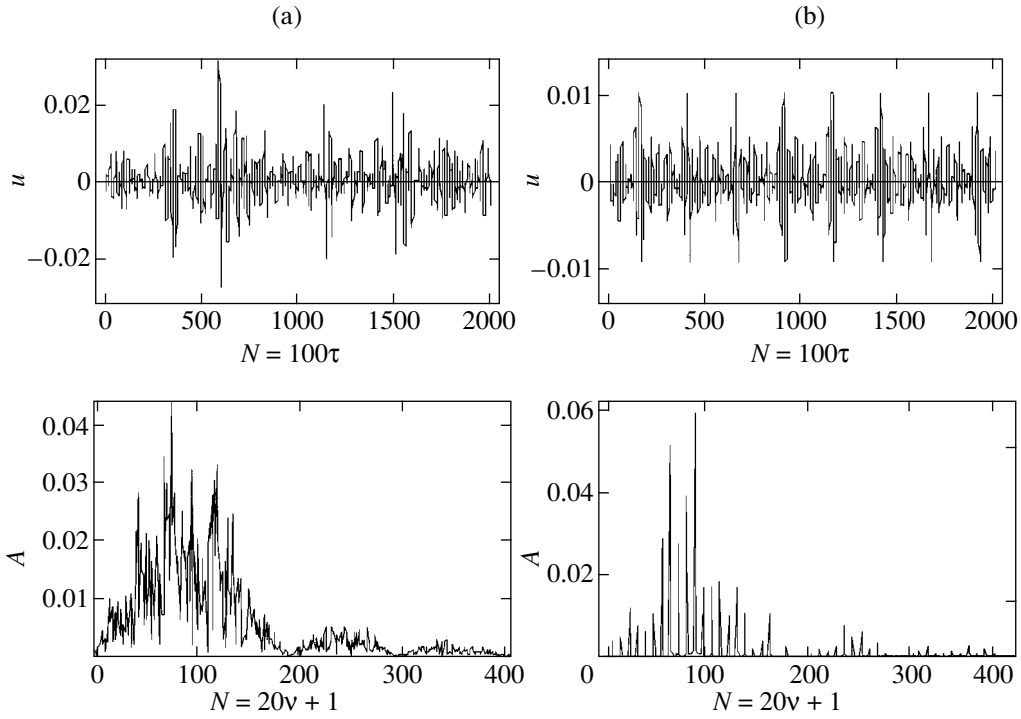


Fig. 4. Velocity pulsations and their spectra at the point $x = y = 10$, $z = 0.7$, in the interval of $\tau \in [30, 60]$: (a) $p_{\mathbf{r}}$ and p_τ, random functions; (b) $p_{\mathbf{r}}$—five fixed points, p_τ—function with the period of $T/20$.

sations per unit length. In so doing, the question arises as to the extent to which the process is steady-state.

We will first treat velocity pulsations in time and, for definiteness, select the minimum channel volume and time, which correspond to the dynamic equilibrium of the state of the structure at the observation point. Previously, it was assumed that the lifetime of disturbance $\tau = 10$. It is expedient that the region around the observation point, or the volume being investigated, should be limited to the size equivalent to this time interval. For $\tau = 10$, the size of this volume will be $\{x_1 - x_2, y_1 - y_2, z_1 - z_2\} = \{0–20, 0–20, 0–1\}$ with the center at the point $x = y = 10$. From the standpoint of the model being treated, the determination of the time t of relaxation of the structure to the steady state is analogous to the problem of finding the level of water in a vessel under conditions of the constant (on the average) delivery of water and the efflux velocity defined by the height of this level. Accordingly, the balance between the initiation and decay of waves will be defined by an expression containing $\exp(-t/\tau)$, and one can put $t = 3\tau$, for which the number of waves that are simultaneously present in the channel will differ from the equilibrium case by not more than 5%. Then, the observation of the process must not be started before the moment of time 3τ. The duration of observation $\tau_2 - \tau_1$ depends on the objective of the problem and technical capabilities. If k disturbances arise during the entire period T ($T > \tau$) of development of the process, $kn\tau/T$ reflected waves on the average will be constantly present in the channel. The reciprocal of this quantity points to the resolution β of the process in time or space.

The time dependence of velocity pulsations in the field of action of several waves is found in the same manner as in the case of several reflections of a single wave. The piecewise smooth function $u(x_1, x_2, y_1, y_2, z_1, z_2, \tau_1, \tau_2, p_x, p_y, p_z, p_\tau, x, y, z, n)$ is summed at the observation point $\{x, y, z\}$ over all k disturbances for the sequence of moments of time τ during the period of their emergence T, starting with the moment of time t. In the examples treated here, the value of $k = 100$ is taken for disturbances. The observation point is preassigned by the coordinates $\{x, y, z\} = \{10, 10, 0.7\}$, and the period being investigated is given by the interval $\{t, T\} = (30, 50)$; i.e., the spectrum is constructed only for the last 20 units of time.

The methods of preassigning the coordinates for a series of k disturbances were discussed in part above. In the Mathematica-4 system, their distribution by the law of chance is described by the function TRandom[], $\{k\}$ with the random number generator set to the s state. For an ordered arrangement of disturbances, use can be made of a function of the form of (T/k)(Range$[k] - 1/2$), where T/k is the period, and the operator Range$[k]$ generates a natural numerical series. An addition to this function of a random element varying within a single period gives a quasi-periodic function which is made up of two previous ones: (T/k)(Range$[k]$ – Random[], k). In the same manner, one can preassign any other law of distribution of coordinates and time of emergence of

disturbances. In determining velocity pulsations, the total time of observation of the process was broken into $N = 2000$ intervals, which corresponded to spectral resolution $\beta = 0.01$. Figure 4 gives the pulsations and spectra for random and fully ordered sequences of p_τ and p_r. The top graphs gave the dependences of velocity u on time τ, and the bottom graphs show the spectra A of velocity pulsations as a function of frequency ν expressed by the number N of intervals β. The graphs in Fig. 4a correspond to a random distribution of all initial parameters of disturbances. Figure 4b is an opposite version, in which the wave sequence is fully ordered, and disturbances at five uniformly arranged points arise periodically, with a period of $0.05T$. Note that the pulsations of velocity $u(\tau)$, as well as the spectrum $A(\nu)$, depend on the position $\{x, y, z\}$ of the observation point. In the first version, they are further affected by the realization of random distributions $\{p_\tau, p_x, p_y, p_z\}$. However, with a random distribution of disturbances, the general form and singularities of pulsations are preserved and resemble the case of single wave. This leads one to extend the inferences made in analyzing a single realization to the entire class of random distributions. The situation is different in the case of ordering, as is well seen in the top graphs of Fig. 4b, where the ordering of initiation of disturbances results in a clearly periodic behavior of pulsations. We will take a look at how this tendency shows up in spectra.

The spectra were investigated in the frequency band $\nu = 50$ (pulsations per unit time equal to the time during which the wave crosses the channel), which is governed by the selected resolution β in view of the rules of Fourier transformation. The spectrum is a set of harmonics whose amplitude decreases very rapidly; therefore, Fig. 4 gives only parts of spectra, defined by the frequency $\nu \approx 20$. The general form of the spectra in Fig. 4 reminds one of the spectral distribution of pulsations developed by a single wave, $k = 1$. As in the case of a single wave, a maximum of the envelope of first harmonic is located at the frequency $1/2a$, and the shape and decay of other harmonics are defined by the properties of initial disturbance, the distribution of parameters in which is given in this case by the harmonic function $P \propto \cos((r - \tau)/a)$. Proceeding from the similarity of spectra at $k = 1$ and $k \gg 1$, one could expect the number k to affect the function $u(\tau)$ and the spectral density $A(\nu)$ more than the form of spectral distribution of pulsations. However, the participation of a series of waves in the process leads to a redistribution of the spectral density. In particular, the half-width of the harmonic and its modulation vary. The only source of all of these changes may only be provided by the space-time characteristics of distribution of disturbances in the initial sequence, including the number of disturbances k.

We will estimate the main parameters of this sequence for the conditions corresponding to the above-identified parameters. On the average, two disturbances arise per unit time, with eight units of surface area of both walls per each one of those disturbances; i.e., the average frequency of initiation of disturbances is two disturbances per unit time, and the spatial frequency is 0.125 disturbance per unit area, or, in terms of the length r (the distance to the observation point), 3.5 disturbances per unit length. The fundamental harmonic has a half-width of $\delta\nu = 5.5$. With a random distribution of disturbances in a sequence (Fig. 4a), several bands incorporating sharp maxima may be identified in the spectrum. The first such maximum is located at a frequency $\nu = 2$, which coincides with the average frequency of initiation of disturbances. However, the highest amplitude is exhibited by the maximum at a frequency $\nu = 3.5$, which coincides with the average spatial frequency of distribution of disturbances (3.5). Note that quantitative agreement between the time and spatial frequencies is possible by virtue of the scales adapted for the dimensionless time τ and length r.

When a partial (spatial) order is introduced into the sequence of wave initiation, the harmonic half-width narrows down and the contrast between individual maxima corresponding to different frequencies and their combinations increases appreciably. This effect increases in the case of fully ordered series of disturbances with respect to both space and time. The effect of periodicity, at which the process is characterized, in addition to frequencies in space and, on the average, with respect to time, is further characterized by the fixed frequency of emergence of a group of five disturbances ($\nu = 0.4$), is demonstrated by the spectrum in Fig. 4b. This spectrum consists of only isolated maxima. Altogether, the fundamental harmonic includes 23 maxima, and the average (over all maxima) frequency shift is 0.4, which is equal to the frequency of emergence of a group of disturbances. These results may be treated as a corroboration of the assumptions of the direct dependence of the behavior of time spectrum modulation on the regularities of the initiation of disturbances. They may prove useful from the standpoint of preliminary estimation of the expected results in the simulation of turbulence, as well as in approximate analysis of the process.

Spatial spectra of pulsations must correspond to similar regularities. In turning to these spectra, we will use the same conditions as those adapted for analyzing pulsations in time. The wave contribution to the structure of pulsations will be included during the assumed lifetime $\tau = 10$. During this time, the wave passes the distance $r = 10$ and is reflected from the walls no more than 10 times. In order to investigate the spatial structure in the steady-state mode and, at the same time, reduce the number of versions of the problem, we will set the dimensions of the working volume in accordance with the relations

$$\{x_1 - x_2, y_1 - y_2, z_1 - z_2\} = \{0\text{–}20, 0\text{–}20, 0\text{–}1\}.$$

We will take the number of disturbances k for this volume to be $k = 100$. For the above-identified conditions, the steady-state mode sets in starting with the moment of time $\tau = T = 50$. The dependence of the velocity pulsations on the coordinates in some direction will be

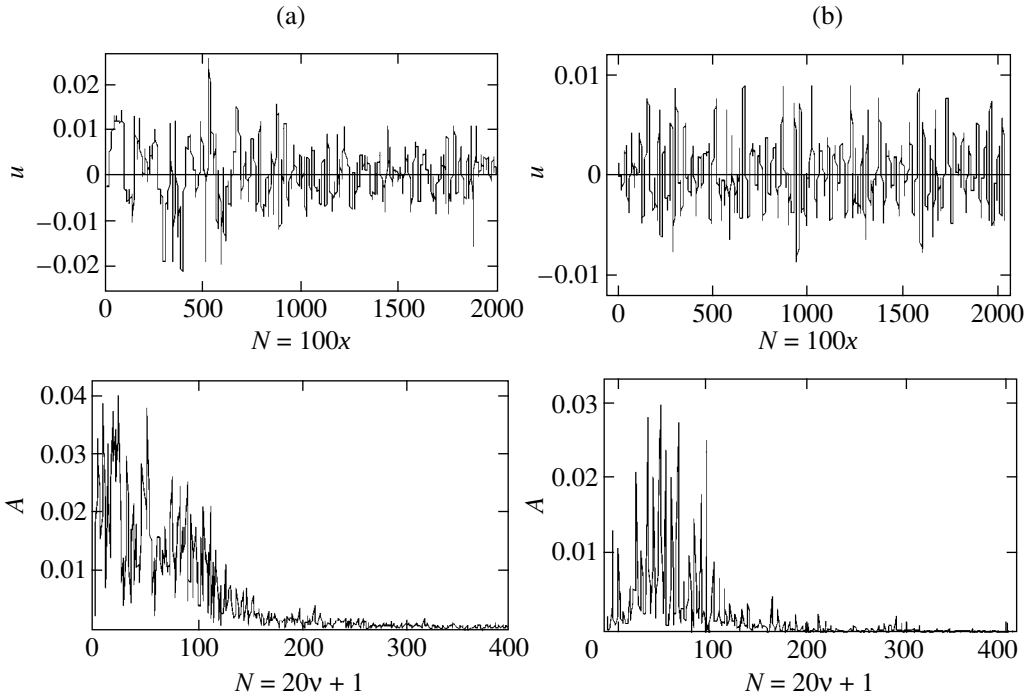


Fig. 5. The parameters of pulsations in the interval of $x \in [10, 30]$ ($y = 10$, $z = 0.7$) at the moment of time $T = 50$: (a) $p_{\mathbf{r}}$ and p_τ, random functions; (b) $p_{\mathbf{r}}$—five fixed points, p_τ—function with the period of $T/20$.

determined using the procedure of obtaining time spectra of pulsations at the observation point.

According to this procedure, in order to obtain the spectrum of pulsations in time, the observation point $\{x, y, z\}$ is fixed, and, at the moments of time $\tau \in [t, T]$, the piecewise smooth function

$$u(x_1, x_2, y_1, y_2, z_1, z_2, \tau_1, \tau_2, p_x, p_y, p_z, p_\tau, x, y, z, n)$$

is calculated and then summed for all k disturbances.

On the other hand, in order to obtain the spatial spectrum in a preassigned direction, the moment of time T is fixed, and the piecewise smooth function

$$u(x_1, x_2, y_1, y_2, z_1, z_2, T, p_x, p_y, p_z, p_\tau, r, n)$$

is calculated and then summed over all k waves at points $\mathbf{r} \in [\mathbf{r}_1, \mathbf{r}_2]$ of the segment being investigated in the preassigned direction. However, some difference exists between different directions. We will treat the procedure of obtaining the spectrum in the transverse and longitudinal directions of the selected channel volume parallel to the z axis and x axis. The first of these directions may be defined by the coordinates x and y, for example, $x = y = 10$, which corresponds to the volume center, with the value of the parameter z varied in the range $z \in [0, 1]$ with a step of 0.01. Similarly, a line parallel to the x axis may be preassigned by the coordinates y and z. However, unlike a line parallel to z whose variation limits are automatically preassigned by the channel walls, no such natural limitation exists for the x axis.

In this case, the choice of boundary conditions must provide, first, the formation of a steady-state structure and, second, the previous specific density of perturbations in the entire region being investigated. This requirement may be satisfied in a natural manner by increasing the extent of the region of initiation of disturbances on each side along x to a distance corresponding to the wave lifetime. While increasing the length of the volume being investigated, one must change the number of initial disturbances accordingly. In the case of $\tau = r = 10$, this matching produces $k = 200$ and $\{x_1 - x_2, y_1 - y_2, z_1 - z_2\} = \{0–40, 0–20, 0–1\}$. The structures of velocity pulsations in the direction of the x axis with $y = 10$ and $z = 0.7$ are given in Fig. 5 for the moment of time $T = 50$. Here, the variable parameter passes the interval $x \in [10, 30]$ with a step of 0.01. The top graphs of Fig. 5 represent the velocity pulsations for the above-identified (Fig. 4) two types of sequences of initial disturbances, and the bottom graphs give their spectra as functions of $N = 20\nu + 1$, where ν is the spatial frequency along x. The spectra are limited by the frequency $\nu \approx 20$.

As is demonstrated by the results of comparing Figs. 4 and 5, the spatial characteristics of pulsations in the direction of the x axis are largely similar to the properties of spectra in time. Comments on the spectra of pulsations in time (Fig. 4) are largely true of the spatial spectra in Fig. 5 as well. In particular, the ordering of the sequence of primary disturbances causes an increase in the contrast of spectral lines. The difference is observed only for the low-frequency (long-wave) region in which the

density of spatial spectra exceeds that of spectra in time. This difference is quite explicable.

Indeed, the characteristic time of variation of velocity at the observation point is associated, first of all, with the thickness $2a$ of the spherical wave layer, so that, at $a = 0.1$, the frequency $\nu = 5$. In this case, the spectral density in other frequency ranges depends on the length of coherence of a pulsation train formed as a result of the superposition of primary waves. Obviously, the possibility of the formation of a train whose length appreciably exceeds $2a$ is less, the greater this length. This is what is observed in Fig. 4, where the spectral density in the low-frequency region decreases almost to zero. On the other hand, when a wave crosses a segment in which a spatial spectrum is investigated, the length of coherence increases automatically with the radius of the spherical wave layer in the process of its propagation. In this case, the density of the low-frequency part of the spectrum depends on the number of waves which have the appropriate parameters and site of initiation. As is seen in Fig. 5, the spectrum in this region is by no means zero.

In this respect, the spatial spectrum in the direction of the z axis is representative. Similarly to the time spectrum, this spectrum lacks the low-frequency component; however, this is due to the fact that the length of the segment being investigated is limited. At the same time, in the region of other frequencies, the spectral distribution on the z axis almost coincides with the envelope of the spatial spectrum in the direction of the x axis. The calculation results demonstrate that the described properties of spectra show up for all forms of sequences of initial disturbances. Note further that the density of the low-frequency part of the spectrum depends also on the width $2a$ of the spherical wave layer. As the layer expands, the spectral density in the low-frequency region increases, because the length of coherence increases both in the layer proper and in the train of pulsations during the superposition of waves. In spatial spectra, this process accelerates further as a result of increase in the volume of coherence.

The foregoing results indicate that the concept of turbulence based on the superposition of acoustic waves enables one to calculate the time and space characteristics of the turbulent process if the boundary and initial conditions are preassigned. Because no additional restrictions are introduced into the suggested model, it may find fairly wide application. For example, it may be used to describe the structure of disturbances of low-temperature plasma, where acoustic (ion-sound) waves exist.

6. COMPARISON OF MODEL CALCULATION RESULTS WITH EXPERIMENTAL DATA

In order to compare the developed model of turbulence with experimental data and to provide an example of its application, we will turn to the results of studies of plasma channels developed by a Bessel beam of laser radiation. Because the experimental procedure and results belong to a new line of investigations, we will provide some explanations essential from the standpoint of understanding subsequent description.

Bessel beams [4] are formed by conical focusing lenses, i.e., axicons, which transform a plane wave front to a conical one with the angle γ at the cone base. Over the axicon focal length, $L \approx R/\gamma$ ($2R$ is the incident wave diameter), the wave beam is diffraction-compensated, and the transverse distribution of the field in the beam is constant and described by the zero-order Bessel function $J_0(kr \sin\gamma)$, where $k = 2\pi/\lambda$ and λ is the radiation wavelength. Zeros of the Bessel function divide the entire beam into cylindrical parts. Their radii are defined by the equalities $kr_i \sin\gamma = 2.40, 5.52, 8.65, 11.79$. With the beam intensity exceeding the breakdown threshold, a plasma channel is formed over the focal segment length. In the channel, the plasma temperature reaches 50 eV; the electron density, 5×10^{19} cm^{-3}; the ion sound velocity $(ZkT_e/m_i)^{1/2}$, 3.5×10^6 cm/s.

Two interference patterns of a channel are given in Fig. 6a. They were obtained as a result of joint research by the Institute of High Temperatures of the Russian Academy of Sciences (IVTAN) and University of Maryland, USA, in the facility described in [5] in the laboratory headed by H.M. Milchberg with the participation of L.Ya. Margolin and the present author. The experiments were made under the following conditions. An axicon with a base angle of 30° ($\gamma = 18°$) transformed heating radiation ($\lambda = 1.06$ μm, $E = 0.6$ J, $\tau = 100$ ps, $D = 1$ cm) to a Bessel beam with the diameter of the central part $2r_b \approx 2.6$ μm and approximately 1.5 cm long. The state of the plasma channel over a length of 512 μm was assessed by the interference patterns, whose enlarged image was recorded by a CCD camera in the light of probing laser radiation (0.53 μm, 70 ps). The numerals at the images of interference patterns indicate the numbers of pixels of the CCD camera, each sized 1.6 μm. The interference patterns in Fig. 6a were obtained for nitrous oxide at pressures of 200 torr (top frame) and 500 torr (bottom frame) 250 ps after the triggering of the heating pulse. This medium was selected in view of the low ionization potential and the low threshold of optical breakdown, owing to which a stable plasma channel could be formed. Against the background of equally inclined interference bands, one can see the contours of plasma channels with diameters of 57.3 and 41.0 μm, respectively.

The deviation of interference band from its initial direction, or the band shift, points to the difference of optical lengths during the beam propagation in undisturbed gas and in the plasma channel, which enables one to use the band shift to assess the structure of disturbances in the channel. As is demonstrated by the interference patterns, the general form of the band shift in the channel depends on pressure. At 200 torr, the interference bands assume the form of meniscus, with their sequence remain-

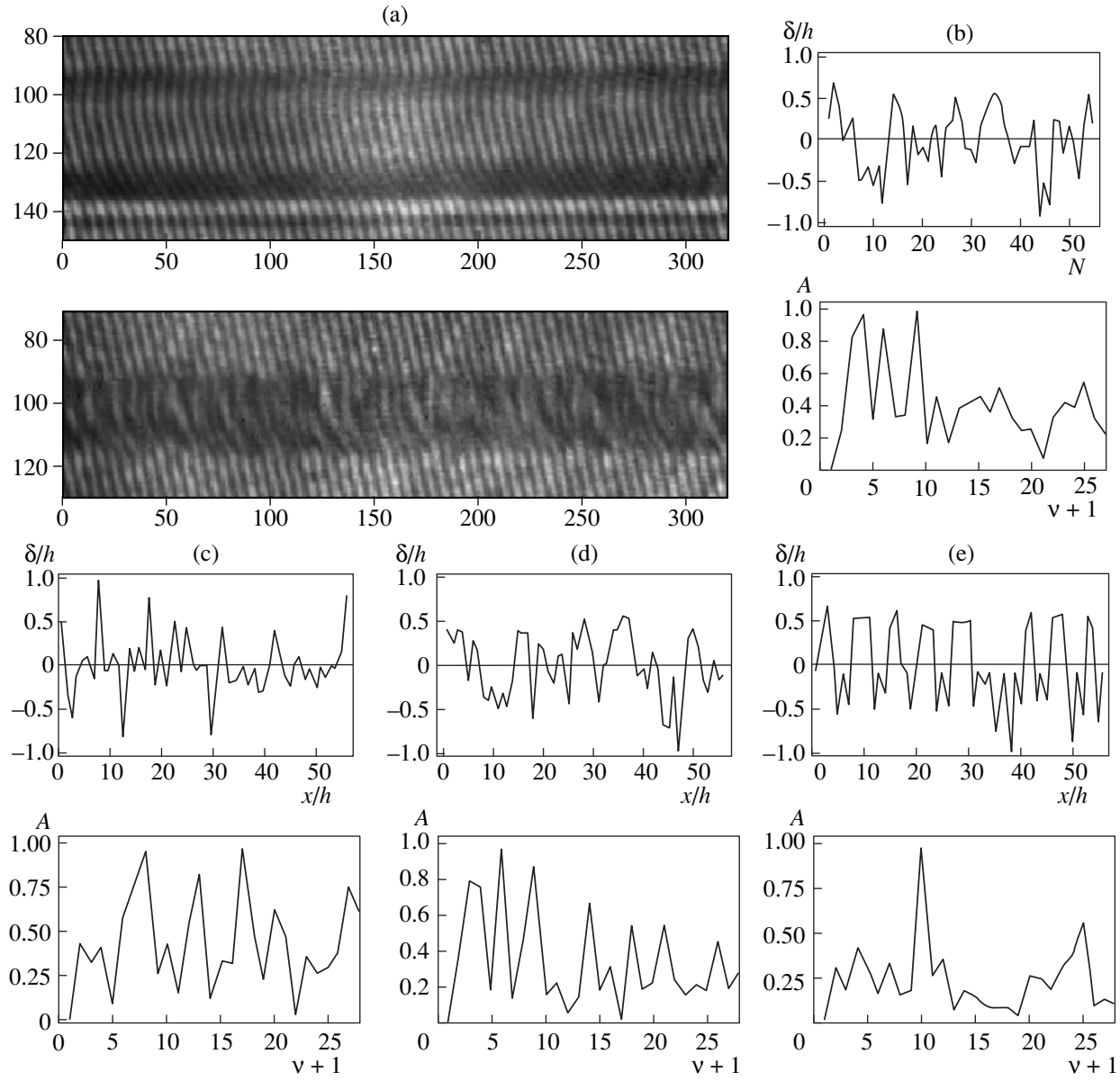


Fig. 6. Simulation of pulsations of parameters in a plasma channel along the x line, $r_0 = 0.5$: (a) interference patterns at pressures of 200 torr (top) and 500 torr (bottom); (b) experiment, band shift and spectrum; (c–e) calculation of pulsations at (c) $a = 0.063$, (d) 0.145, (e) 0.227.

ing regular. As the pressure increases, this regularity is gradually disturbed and, at 500 torr, it almost disappears. With a subsequent minor increase in pressure or delay, the band trace is lost, and it is impossible to measure its shift. Therefore, the interference pattern, recorded at 500 torr with a delay of 250 ps, was used to compare disturbances in the channel with the prediction data about pulsations.

The selected experimental procedure, defined by the high requirements of space-time resolution, makes it possible to determine the spatial characteristics of pulsations only in the axial direction, along the x axis. Indeed, the time characteristics would call for a large number of interference patterns, and a much higher density of interference bands is required for a radial distribution that is analogous to the function $u(z)$. At the same time, 56 bands come into the view of the interferometer in the longitudinal direction, along the x axis. All in all, the shift δ of interference bands does not exceed the order of interference h (distance between the bands). We assume them to be resolvable elements and obtain the above-mentioned possibility of constructing the curve of dependence of the band shift on longitudinal coordinate $\delta(x)$, analogous to the distribution of velocity pulsations $u(x)$. The interference pattern enables one to find the distribution of shifts $\delta(x)$ along lines parallel to the axis but arranged

at different distances from the latter (an analog of the channel height).

Given in Fig. 6b by way of example are the results of measurements of the band shift on the half-radius level, $r/R = 1/2$. The top graph represents the $\delta(N)$ correlation, where N is the interference band number equal to x/h, and the bottom graph represents the respective spatial spectrum. Here, the spectral density is expressed in relative units (as a ratio to its maximum), and the frequency ν is expressed in pulsations per single interference band, i.e., per distance h between poles. In comparing these results with the graphs in Fig. 5, note that the structure of disturbances in a plasma channel (with allowance for limited resolution) has features in common with structures of pulsations in a flat channel for ordered or partially ordered sequences of initial disturbances. However, a more detailed comparison calls for refinement of the model. Therefore, we will treat the process of channel formation in more detail.

The source of energy during channel formation is obviously provided by the field of Bessel beam. As is known [4], this field has a structure. Its properties in the radial direction are associated with the form of the Bessel function, and in the longitudinal direction it is shaped up under the effect of nonlinear processes and forms a chain of maxima divided by distance $l = 2\lambda/\sin^2\gamma$ [4] (in our case, $l \approx 21$ μm). We will assume for the time being that the breakdown of gas occurs in the volume of such a maximum, and that this breakdown may be treated as a microexplosion, with the channel formation being described proceeding from the problem of a series of point explosions. At the initial stage, the wave propagation may be regarded as self-similar. However, we are interested in late stages, at which the conditions of self-similarity are disturbed, and the wave approximates an acoustic wave. A direct calculation of wave propagation is associated with errors due to inaccurate data about the process of energy contribution during the breakdown of gas [6]. In order to verify this assumption, we will use the solution of the problem of point explosion with counter-pressure [7], in which characteristic scales of length and time appear. In particular, the length scale is usually provided by the sphere radius $R = (\varepsilon/p)^{1/3}$, where ε is the energy of microexplosion and p is the initial gas pressure. We compare two processes (Fig. 6a), which are registered under identical conditions at the same time (one extra parameter is eliminated) and differ by the initial pressure alone, to find the ratio of their diameters. This ratio is 1.40, which almost coincides with the value of 1.38 calculated for the length scale R. Therefore, this assumption is confirmed. However, the question remains as to the site and volume of the microexplosion. The problem is that it is almost impossible to obtain these data experimentally. It turns out, however, that the problem may be solved using the suggested model.

Let us formulate the problem. Microsites of breakdown serve the function of primary disturbances. In the radial direction, they may arise inside of the main part of the Bessel beam or within several of its rings, whose size on the scale of the channel radius (41.0 μm at 500 torr) is $a = 0.063$, 0.145, 0.227, and so on. In the direction along the channel, the number of the rings must not exceed the number of diffraction maxima, which is 24 or 25 over a length of 512 μm but may be less. The problem reduces to determining these parameters. In order to solve the problem, we will, first of all, set the parameters of the model in correspondence with the experimental conditions. Note that the parameters of the model, adapted in constructing the graphs in Fig. 5, corresponded to the steady-state mode of the process and described local pulsations of velocity in a flat channel with disturbances located on its walls. In the experiment, the process is characterized by a number of distinctions.

At the initial stage of the channel expansion, when the first disturbances only form the radial dimensions of the cylinder, reflected waves may be ignored, and the state of the structure may be characterized by the parameter $T = 1$. Then, by virtue of the symmetry of the propagation of pulsations, their geometry may be described by two coordinates, namely, longitudinal coordinate x and radial coordinate r. Further, the structure of pulsations in the model should be analyzed in the same segment as in the experiment: $x \in [0, 25R]$, where R is the channel radius. We express the coordinate x (for convenience of comparison with the measurement results) in terms of the distance h between interference bands to derive $\{x_1, x_2\} = \{0, 55\}$ for the extent of this segment. As is known, the shift of interference bands produces a pattern of distribution of disturbances, which is averaged along the chord of cylindrical channel, and the disturbances arise on its axis. Let the averaging take place on a distance equal to half the channel radius, i.e., impact parameter $r_0 = R/2$. Then, the path y along the half of the chord with the coordinates $x = (x_2 - x_1)/2$ and $z = R/2$ varies within zero to $r_0\tan\varphi$, where φ is the azimuth angle. The summation over the length y of the chord must include all disturbances that arise at the points p_x in the region $\{x_1, x_2\}$ at the moments p_τ of the time of the effect of a heating pulse of duration τ_0.

The treated properties of Bessel beam and plasma channel enable one to find the parameters of the model. A version of these parameters for describing pulsations in the plasma channel has the form

$$a = 0.063, \quad T = 1, \quad k = 25, \quad \tau_0 = 0.4,$$
$$x_1 = -1, \quad x_2 = 56, \quad y_1 = 0, \quad y_2 = r_0\tan(\pi/3),$$
$$r_0 = 0.5, \quad m = 11, \quad \beta = 1,$$
$$Q = 1, \quad q = 0, \quad \theta = 0.$$

In addition to the conditions specified previously, this list includes the length of the part of the channel being investigated $\{x_1, x_2\}$ and the number of disturbances k

with due regard for edge effects ($T = 1$), the resolvable element β, the form of function $P(r/a)$, the number m of elements of summation along the chord, the longitudinal distribution of disturbances p_x, the shift Q of the entire sequence of disturbances (within the band), the range q of random deviations of coordinates in the case of partial violation of periodicity of sequence p_x, the time distribution of disturbances p_τ, and a possible delay of breakdown θ. In the experiment, the pulsations expressed in terms of the order of interference h are constructed as functions of the number of the interference band (with resolution of $\beta = 1$, single band). Dimensionless quantities on the scale of the channel radius R are used in the calculations. The coefficient μ was introduced for comparison of the model and measured data, and all graphs are given as the ratio of quantities to their maximum values.

The results of applying the model to initial perturbations of sizes $a = 0.063$, 0.145, and 0.227 (corresponding to the central part of the Bessel function, to its first and second rings) are given in Figs. 6c–6e. The graphs in Fig. 6c represent the structure of pulsations for parameters corresponding to the foregoing version. Here, the size of disturbances is limited to the central part of the Bessel beam, $a = 0.063$, the periodic (along the x axis) arrangement of disturbances and the random law of their initiation within the duration of the heating pulse are assumed, $\tau_0 = 0.4$, and the function $P(r/a) = 1$ is selected.

The wave front of the Bessel beam travels through the channel region being investigated in a time of approximately $\tau = 0.01$ (less than 2 ps); therefore, the delay θ in this stage was not included. The model dependences in Fig. 6d are constructed for the case of $a = 0.145$, when the diameter of initial pulsations is limited by the second zero of the Bessel function. Here, it was still assumed that the disturbances are arranged periodically along the x axis and, in time, by the law of chance in accordance with the previously described procedure. For the subsequent ring of the Bessel function, where $a = 0.227$, the model graphs (with the same assumptions) are given in Fig. 6e.

The results of comparison of Figs. 6c and 6e (the central part, $a = 0.063$, and the second ring, $a = 0.227$, of the Bessel function) with the experimental data (Fig. 6b) point to an inconsistency between the selected parameters of the model and the conditions of plasma channel development. A much better agreement is exhibited by the graphs in Fig. 6d for the first ring, $a = 0.145$. The characteristics of the process, which provided for the best agreement between the predicted and experimentally obtained dependences, were determined by exhaustion of parameters of the model. The number of disturbances k, the distribution p_x (including the shift Q of the entire sequence and the range of scatter q), the duration τ_0 of the period of initiation, and the form of the function $P(r/a)$ were varied for $a = 0.145$.

The calculation results have demonstrated that the structure of disturbances in the channels is described most accurately for the following values of the parameters being varied: $k = 25$, $\tau_0 = 0.2$, $Q = 0.17$, $q = 0$, and $P(r/a) = 1$. According to these data, disturbances arise within the second ring of the Bessel function with a radius of 3 μm during a period of time limited by the range $\tau = 0.2$, or approximately 50 ps. The disturbances are arranged regularly with a period of 21 μm, and the scatter does not exceed several percent.

Therefore, the comparison of the results of model calculations with the measurement data reveals that the model correctly reflects the processes observed in the experiment and proves the validity of the model. At the same time, the use of the model made it possible to determine the singularities of formation of a plasma channel in the field of Bessel beam, which cannot be measured experimentally.

7. CONCLUSION

The developed concepts of the nature of turbulence made it possible to formulate the principles of simulating turbulent process, which are valid at any stage of development of this process and for different initial and boundary conditions. Models have been developed for a flat channel defined by two walls and for a square tube; these models describe the distribution of pulsations of hydrodynamic parameters, as well as space-time Fourier spectra of these pulsations.

A flat channel has simple boundary conditions. Indeed, in a channel like this, the disturbance of parameters at the observation point may be represented as a passage of spherical surface through this point and its mirror images in the channel walls. The shift of such an imaginary point is defined by only one of its coordinates; on every reflection, this coordinate varies by the double height of the channel, $2d$. This simple rule may be readily extended to the case of a tube with flat faces. In this case, however, the number of imaginary points to be included increases. For example, in a square tube, the number of such points triples in every cycle of reflections and the position of these points will be defined by two coordinates with the period $2d$ (here, d is a side of square) in each direction. The example of the steady-state mode of turbulence in a flat channel is used to demonstrate all important singularities of the phenomenon and, at the same time, the potentialities of the suggested concept. The parameters affecting the structure of pulsations have been determined, as well as the associated effects.

The results of model calculations are compared with the data of measurements of the structure of pulsations of parameters in a plasma channel formed by a Bessel beam of laser radiation. The results of comparison reveal that the model reflects correctly the processes observed in the experiment and prove the efficiency of the model. The model made it possible both to describe the mechanism of formation and the structure of disturbances in a channel and to make a number of inferences

about the singularities of interaction between a Bessel beam and matter, which cannot be measured experimentally. For example, the size of the sites of breakdown of gas in the field of Bessel beam and their arrangement along the channel axis were found, as well as a period of time of initiation that proved to be shorter than the heating pulse duration. Note that we are dealing with physical phenomena developing over micron distances and picosecond times, the measurements of which present certain difficulties.

The results obtained in this study indicate that the concept of turbulence based on the superposition of acoustic waves enables one to calculate the time and space characteristics of the turbulent process if the boundary and initial conditions are preassigned. Because no additional restrictions are introduced into the suggested model, it may find fairly wide application.

REFERENCES

1. H. A. Lorentz, Über die Entstehung turbulenter Flüssigkeitsbewegungen und über den Einfluß dieser Bewegungen bei der Strömung durch Röhren, in *Abhandlungen über theoretische Physik* (B. G. Teubner, Leipzig, 1907), p. 43.
2. L. N. Pyatnitsky, Zh. Éksp. Teor. Fiz. **113**, 191 (1998) [JETP **86**, 107 (1998)].
3. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 6: *Fluid Mechanics* (Nauka, Moscow, 1986; Pergamon, New York, 1987).
4. N. E. Andreev, Yu. A. Aristov, L. Ya. Polonskiĭ, and L. N. Pyatnitsky, Zh. Éksp. Teor. Fiz. **100**, 1756 (1991) [Sov. Phys. JETP **73**, 969 (1991)]; N. E. Andreev, L. Ya. Margolin, I. V. Pleshanov, and L. N. Pyatnitsky, Zh. Éksp. Teor. Fiz. **105**, 1232 (1994) [JETP **78**, 663 (1994)].
5. T. R. Clark and H. M. Milchberg, Phys. Rev. E **57**, 3417 (1998); J. Fan, T. R. Clark, and H. M. Milchberg, Appl. Phys. Lett. **73**, 3064 (1998).
6. H. Hora, *Physics of Laser-Driven Plasmas* (Wiley, New York, 1981; Énergoatomizdat, Moscow, 1986).
7. Ya. B. Zel'dovich and Yu. P. Raizer, *Physics of Shock Waves and High-Temperature Hydrodynamic Phenomena* (Fizmatgiz, Moscow, 1963; Academic, New York, 1966, 1967), Vols. 1, 2.

Translated by H. Bronstein

Journal of Experimental and Theoretical Physics, Vol. 92, No. 4, 2001, pp. 594–607.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 4, 2001, pp. 685–699.

PLASMA, GASES

Multipetal Vortex Structures in Two-Dimensional Models of Geophysical Fluid Dynamics and Plasma¶

V. P. Goncharov[a, *] and V. I. Pavlov[b, **]

[a]*Institute of Atmospheric Physics, Russian Academy of Sciences, Moscow, 109017 Russia*
**e-mail: vponom@atm.phys.msu.su*
[b]*U.F.R. des Mathématiques Pures et Appliquées, U.S.T.L., 59655 Villeneuve d'Ascq Cedex, France*
***e-mail: vipavlov@omega.univ_lille1.fr*

Received September 14, 2000

Abstract—A new class of strongly nonlinear steadily rotating vortices is found. The Hamiltonian contour dynamics is proposed as a new approach for their study in some models of geophysical fluid dynamics and plasma. Using the Euler description as a starting point, we present a systematic procedure to reduce the two-dimensional dynamics of constant-vorticity and constant-density patches to the Hamiltonian dynamics of their contours for various parametrizations of the contour. The special Dirac procedure is used to eliminate the constraints arising in the Hamiltonian formulations with the Lagrangian parametrization of the contour. Numerical estimations illustrating the physical significance of the results and the range of model parameters where these results can be applicable are presented. Possible generalizations of the approach based on the application of the Hamiltonian contour dynamics to nonplanar and 3D flows are discussed. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The purpose of this paper is the analytical and numerical study of a new class of strongly nonlinear steadily rotating vortices that can exist in two-dimensional flows with an internal scale similar to the Rossby deformation radius in quasigeostrophic models of geophysical fluid dynamics [1]. We show that these vortices can have a nontrivial multipetal structure and must rotate with comparatively small velocities under the assumption that their characteristic scales are sufficiently large compared to the internal one.

We also present a new approach based on a Hamiltonian version of the contour dynamics. The fact that equations of the contour dynamics are strongly nonlinear and genuinely nonlocal gave impetus to the progress and application mainly of numerical methods for their solution [2]. The analytical versions involving small parameters used for deriving and solving the approximate (local) equations of contour dynamics are only applicable in fluid models with an exterior characteristic scale (e.g., the depth of the unperturbed layer [3]) or with an internal one (e.g., the Rossby radius [4]). Because the solution of problems of this type essentially depends on choosing dynamic variables parametrizing the boundary, it is desirable to have a sufficiently flexible formulation of the equations of contour dynamics such that these equations could be easily reformulated from one phase space into another. In using approximate methods, it is important to keep in mind that all the information on the internal symmetry properties responsible for the dynamical individuality of the Hamiltonian system is contained in the Poisson brackets. Thus, in order to prevent the loss of internal symmetry properties of the system, we must use the approximations where one quantity—the Hamiltonian of the system—is subjected to these approximations but the original Poisson brackets remain intact. The need to use asymptotic methods is the principal reason for refusing traditional formulations, which are not only incompatible with these requirements but also not infrequently lead to cumbersome and recurrent calculations.

This paper is organized as follows. In Section 2, we construct local Poisson brackets for an incompressible nonuniform fluid. Relying heavily on this result as a fundamental principle, in Section 3 we derive a hierarchy of the reduced Poisson brackets specially adapted to the Hamiltonian description of models of the contour dynamics. The contour parametrization plays a decisive role. The occurrence of constraints is the indispensable feature of those Hamiltonian formulations that use the Lagrangian coordinates for this purpose. To eliminate the constraints, Dirac's procedure is used. In Section 4, we consider multipetal vortex structures in the Hasegawa–Mima model and the axial model of electronic fluid as examples of models admitting a direct application of the obtained results. We focus our attention on the study of steadily rotating multipetal vortex structures without contour self-intersections. Some numerical estimates and concluding remarks are presented in Section 5.

¶ This paper was submitted by the authors in English.

1063-7761/01/9204-0594$21.00

2. POISSON BRACKETS FOR AN INCOMPRESSIBLE NONUNIFORM EULERIAN FLUID

The equations of motion for a nonuniform incompressible fluid are formulated in terms of the Eulerian variables: the mass density ρ, the velocity $\mathbf{v}$, and the pressure p, as

$$\partial_t v_i + v_k \partial_k v_i = -\frac{1}{\rho}\partial_i p + \frac{1}{\rho} f_i, \tag{2.1}$$

$$\partial_t \rho + v_k \partial_k \rho = 0, \tag{2.2}$$

$$\partial_k v_k = 0, \tag{2.3}$$

where $\mathbf{f}$ is the result of exterior forces that do not violate the conservativeness of the fluid. This means that equations of motion (2.1)–(2.3) preserve the total energy H given by the sum of the kinetic energy T and the potential U of the fluid,

$$\begin{gathered} H = T + U, \\ T = \int \frac{\mathbf{v}^2}{2\rho} d\mathbf{x}, \quad U = U[\rho], \end{gathered} \tag{2.4}$$

where U is in general an arbitrary functional of the density ρ. For simplicity, we assume that the fluid is unbounded.

We now find the evolution equation for the momentum density $\boldsymbol{\pi} = \rho\mathbf{v}$. Equations (2.1) and (2.2) imply

$$\begin{aligned} &\partial_t \pi_i + v_k(\partial_k \pi_i - \partial_i \pi_k) \\ &= -\partial_i\left(p + \rho\frac{\mathbf{v}^2}{2}\right) + \frac{\mathbf{v}^2}{2}\partial_i \rho + f_i. \end{aligned} \tag{2.5}$$

Taking the curl of Eq. (2.5) and thereby eliminating the gradient term involving the pressure, we obtain the equation

$$\partial_t \gamma_i = e^{imn}\partial_m\left[e^{nkl} v_k \gamma_l - \frac{\mathbf{v}^2}{2}\partial_n \rho + f_n\right], \tag{2.6}$$

which describes the evolution law for the vorticity of the momentum density $\boldsymbol{\gamma} = \nabla \times \boldsymbol{\pi}$ under the action of exterior conservative forces.

We now show that the equations of motion for the incompressible inhomogeneous fluid reformulated in terms of the momentum density vorticity are Hamiltonian with the local Poisson brackets $\{\gamma_i, \gamma_k'\}$ and $\{\rho, \gamma_k'\}$. First, we compute the Poisson bracket $\{\rho, \gamma_k'\}$. Because the model is expected to be Hamiltonian, we have every reason to write

$$\partial_t \rho = \{\rho, H\} = \int\left[\{\rho, \gamma_k'\}\frac{\delta T}{\delta \gamma_k'} + \{\rho, \rho'\}\frac{\delta U}{\delta \rho'}\right] d\mathbf{x}'. \tag{2.7}$$

Comparing Eq. (2.7) with continuity condition (2.2) leads us to

$$\int\left[\{\rho, \gamma_k'\}\frac{\delta T}{\delta \gamma_k'} + \{\rho, \rho'\}\frac{\delta U}{\delta \rho'}\right] d\mathbf{x}' + v_k \partial_k \rho = 0. \tag{2.8}$$

We next introduce a local term in the integrand using the δ-function and express the velocity components v_l in terms of the functional derivatives $\delta T/\delta\gamma_k$ as

$$v_l = \frac{\delta T}{\delta \pi_l} = \int \frac{\delta T}{\delta \gamma_k'}\frac{\delta \gamma_k'}{\delta \pi_l} d\mathbf{x}' = e^{lki}\partial_k \frac{\delta T}{\delta \gamma_i}, \tag{2.9}$$

which can be directly obtained from Eq. (2.4). Upon integrating by parts and after some algebra in Eq. (2.8). We obtain

$$\begin{aligned} &\int \frac{\delta T}{\delta \gamma_k'}[\{\rho, \gamma_k'\} - e^{kml}\partial_l \rho \partial_m \delta(\mathbf{x} - \mathbf{x}')] d\mathbf{x} \\ &\qquad + \int \{\rho, \rho'\}\frac{\delta U}{\delta \rho'} d\mathbf{x}' = 0. \end{aligned}$$

This implies that

$$\{\rho, \gamma_k'\} = e^{kml}\partial_l \rho \partial_m \delta(\mathbf{x} - \mathbf{x}'), \quad \{\rho, \rho'\} = 0. \tag{2.10}$$

It remains to compute the Poisson bracket $\{\gamma_i, \gamma_k'\}$. Using the same reasoning as for the density, we can write the equation of motion for the vorticity of the momentum density $\boldsymbol{\gamma}$ as

$$\begin{aligned} \partial_t \gamma_i &= \{\gamma_i, H\} \\ &= \int\left[\{\gamma_i, \gamma_k'\}\frac{\delta T}{\delta \gamma_k'} + \{\gamma_i, \rho'\}\frac{\delta T}{\delta \rho'}\right] d\mathbf{x}' + \{\gamma_i, U\}. \end{aligned} \tag{2.11}$$

With the bracket $\{\rho, \gamma_k'\}$ already computed and

$$\frac{\delta T}{\delta \rho} = \frac{1}{2} v_k^2,$$

Eq. (2.11) can be rewritten as

$$\begin{aligned} \partial_t \gamma_i &= \int \{\gamma_i, \gamma_k'\}\frac{\delta T}{\delta \gamma_k'} d\mathbf{x}' \\ &\quad - e^{iml}\partial_m\left(\frac{1}{2} v_k^2 \partial_l \rho\right) + \{\gamma_i, U\}. \end{aligned} \tag{2.12}$$

Comparing Eqs. (2.12) and (2.6), we obtain

$$\begin{aligned} &\int \{\gamma_i, \gamma_k'\}\frac{\delta T}{\delta \gamma_k'} d\mathbf{x}' - e^{imn}\partial_m(e^{nkl} v_k \gamma_l) \\ &\qquad + \{\gamma_i, U\} - e^{imn}\partial_m f_n = 0. \end{aligned}$$

If we introduce the local term $e^{imn}\partial_m(e^{nkl} v_k \gamma_l)$ into the integral using the δ-function and replace the velocity

components v_l in accordance with Eq. (2.9), after the integration by parts we obtain

$$\int \frac{\delta T}{\delta \gamma_k'}[\{\gamma_i, \gamma_k'\} - e^{ipj}e^{jln}e^{knm}\partial_p\gamma_l\partial_m\delta(\mathbf{x}-\mathbf{x}')]d\mathbf{x}'$$
$$+ \{\gamma_i, U\} - e^{imn}\partial_m f_n = 0.$$

This immediately implies that the Poisson bracket for the vector field $\boldsymbol{\gamma}$ and the relation between the exterior force and the potential energy are given by

$$\{\gamma_i, \gamma_k'\} = e^{ipj}e^{jln}e^{knm}\partial_p\gamma_l\partial_m\delta, \tag{2.13}$$

$$\{\gamma_i, U\} = e^{imn}\partial_m f_n. \tag{2.14}$$

We note that the resulting force **f** can be found from Eq. (2.14) up to a gradient term. This fact is a consequence of the invariance of the equations of motion (2.1)–(2.3) under the transformation

$$p \longrightarrow p + \phi, \quad f_i \longrightarrow f_i - \partial_i\phi,$$

where ϕ is an arbitrary function whose choice has no influence on the physical implications of the theory. Thus, it follows from Eq. (2.14) that no structure other than

$$f_i = \frac{\partial\rho}{\partial x_i}\frac{\delta U}{\delta\rho}$$

is admissible for the external forces in the case where $U = U[\rho]$.

Collecting Eqs. (2.10) and (2.13), we find the complete system of Poisson brackets in the phase space ($\boldsymbol{\gamma}$, ρ),

$$\{\rho, \rho'\} = 0, \tag{2.15}$$

$$\{\rho, \gamma_k'\} = e^{kml}\partial_l\rho\partial_m\delta, \tag{2.16}$$

$$\{\gamma_i, \gamma_k'\} = e^{ipj}e^{jln}e^{knm}\partial_h\gamma_l\partial_m\delta. \tag{2.17}$$

Therefore, the equations of motion for an incompressible nonuniform fluid corresponding to these Poisson brackets take the form

$$\partial_t\boldsymbol{\gamma} = \{\boldsymbol{\gamma}, H\} = \nabla\times\left(\left[\boldsymbol{\gamma}, \nabla\times\frac{\delta H}{\delta\boldsymbol{\gamma}}\right] + \frac{\delta H}{\delta\rho}\nabla\rho\right), \tag{2.18}$$

$$\partial_t\rho = \{\rho, H\} = -\left(\nabla\times\frac{\delta H}{\delta\boldsymbol{\gamma}}\right)\cdot\nabla\rho. \tag{2.19}$$

The results obtained in Eqs. (2.15)–(2.19) can be considered as a generalization of the well-known Hamiltonian description of an incompressible homogeneous fluid (see, for example, [5–10]) and are used in what follows as a fundamental principle in constructing a hierarchy of reduced Poisson brackets for various models of contour dynamics.

3. HAMILTONIAN VERSION OF THE CONTOUR DYNAMICS

We begin with a two-dimensional plane flow where the curl of the momentum is normal to the flow plane and hence has only the component

$$\boldsymbol{\gamma} = \{0, 0, \gamma\}, \quad \gamma = \varepsilon^{ik}\partial_i\pi_k, \tag{3.1}$$

where ε^{ik} is the unit antisymmetric tensor (with $\varepsilon^{12} = 1$). In this case, Poisson brackets (2.15)–(2.17) for an incompressible inhomogeneous fluid can be reformulated for the dynamical variables γ and ρ as

$$\{\rho, \rho'\} = 0, \tag{3.2}$$

$$\{\rho, \gamma'\} = \varepsilon^{ki}\partial_i\rho\partial_k\delta(\mathbf{x}-\mathbf{x}'), \tag{3.3}$$

$$\{\gamma, \gamma'\} = \varepsilon^{ki}\partial_i\gamma\partial_k\delta(\mathbf{x}-\mathbf{x}'). \tag{3.4}$$

It is well known that two-dimensional dynamics of patches of a constant vorticity and density can be reduced to the dynamics of their contours, ignoring the description of the rest of the fluid. However, it is a nontrivial fact that the description of the contour evolution can take various forms depending on the variables used; this deserves attention from both practical and theoretical standpoints.

For simplicity, we consider a single domain G^+ bounded by a closed fluid contour that separates it from the rest of the fluid in an exterior region G^-. Denoting the vorticity and the density inside and outside accordingly as ω^+, ρ^+, and ω^-, ρ^-, we use the respective + and – superscripts for labeling variables in the internal domain G^+ and in the exterior region G^-. Using this notation, we can write the momentum and the mass density as

$$\begin{aligned}\boldsymbol{\pi} &= \rho^+\mathbf{v}^+\theta^+ + \rho^-\mathbf{v}^-\theta^-,\\ \rho &= \rho^+\theta^+ + \rho^-\theta^-,\end{aligned} \tag{3.5}$$

where θ^+ and θ^- are the mutually complementary substantive functions

$$\theta^+ = \begin{cases}1 & \text{if } \mathbf{x}\in G^+\\ 0 & \text{if } \mathbf{x}\in G^-,\end{cases} \qquad \theta^- = \begin{cases}1 & \text{if } \mathbf{x}\in G^-\\ 0 & \text{if } \mathbf{x}\in G^+,\end{cases}$$

such that

$$\theta^+ + \theta^- = 1, \quad \theta^+\theta^- = 0. \tag{3.6}$$

We note that, by definition, a substantive θ-function characterizing a fluid domain has the dynamical property

$$\partial_t\theta + v_k\partial_k\theta = 0,$$

implying that the corresponding domain moves together with the fluid.

Inserting $\boldsymbol{\pi}$-representation (3.5) in Eq. (3.1) yields

$$\gamma = \rho^+\omega^+\theta^+ + \rho^-\omega^-\theta^- + \beta, \tag{3.7}$$

where the variable β can be expressed as

$$\beta = (\rho^+ v_k^+ - \rho^- v_k^-)\varepsilon^{ik}\partial_i\theta^+. \quad (3.8)$$

It is easily seen that β has a δ-functional character and thus describes a vortex sheet whose density is specified by the jump of the tangential momentum across the contour.

As the first step, we transform Poisson brackets (3.2)–(3.4) from the phase space (γ, ρ) into the space of dynamical variables (β, θ^+). In accordance with Eqs. (3.5)–(3.7), we have

$$\rho = \rho^- + (\rho^+ - \rho^-)\theta^+, \quad (3.9)$$

$$\gamma = \rho^-\omega^- + (\rho^+\omega^+ - \rho^-\omega^-)\theta^+ + \beta. \quad (3.10)$$

Depending on the existence of a mass density jump across the contour, insertion of Eqs. (3.9) and (3.10) into Eqs. (3.2)–(3.4) leads to two types of Poisson brackets.

3.1. Piecewise-Constant Vortex Models without Mass Density Jumps

We first consider the degenerate case where the mass density jump is absent, and therefore $\rho^+ = \rho^- = \rho_0$. In this case, the vortex sheet density is a constant of motion and its presence modifies the Hamiltonian of the model but has no influence on the Poisson bracket $\{\theta^+, \theta^{+\prime}\}$ that completely determines the contour evolution. Taking this into account, we can set $\beta = 0$ for simplicity of computing. Inserting Eq. (3.10) in Eq. (3.4), we then obtain

$$\{\theta^+, \theta^{+\prime}\} = \nu^{-1}\varepsilon^{ik}\partial_k\theta^+\partial_i\delta(\mathbf{x} - \mathbf{x}'), \quad (3.11)$$

where $\nu = \rho_0(\omega^+ - \omega^-)$.

Which of the Hamiltonian versions of contour dynamics follows from Eq. (3.11) depends on how we parameterize the substantive θ^+-function. The simplest parameterization can be achieved with the Heaviside function

$$\theta^+(\eta - x_2) = \begin{cases} 1 & \text{if } \eta \geq x_2 \\ 0 & \text{if } \eta < x_2, \end{cases}$$

where the variable $\eta = \eta(x_1, t)$ specifies the contour shape. The corresponding version of the Hamiltonian description defined by the Poisson bracket $\{\eta, \eta'\}$ can be derived directly from Eq. (3.11) if we use the trivial relation

$$\eta = \int x_2 \frac{d}{dx_2}\theta^+(\eta - x_2)dx_2$$

that maps the dynamics in the phase space of $\boldsymbol{\gamma}$ into the phase space of η. After some algebra, we then find

$$\{\eta, \eta'\} = \int x_2 x_2' \frac{d^2}{dx_2 dx_2'}\{\theta^+, \theta^{+\prime}\}dx_2 dx_2'$$

$$= -\nu^{-1}\frac{\partial}{\partial x_1}\delta(x_1 - x_1').$$

It is noteworthy that the same Poisson bracket characterizes the KdV-type equations. Hamiltonian formulations based on this version of Poisson brackets are preferable for the study of multilayer models [3].

A more general parametrization can be realized when the contour C bounding the domain G^+ is given in the parametric form

$$\mathbf{x} = \hat{\mathbf{x}}(s, t),$$

where s is the contour arc length. The vector $\mathbf{t} = \partial\hat{\mathbf{x}}/\partial s$ tangential to the contour satisfies the normalization condition

$$\mathbf{t}^2 = 1. \quad (3.12)$$

We note that the θ^+-functions admit an analytical representation through the contour integral,

$$\theta^+ = \frac{i}{2\pi}\int_C \frac{\hat{z}_s ds}{z - \hat{z}}, \quad (3.13)$$

where $z = x_1 + ix_2$ and $\hat{z} = \hat{x}_1 + i\hat{x}_2$ are complex-valued notations for the vectors $\mathbf{x} = (x_1, x_2)$ and $\hat{\mathbf{x}} = (\hat{x}_1, \hat{x}_2)$, and i is the imaginary unit. Representation (3.13) can be obtained as a consequence of the Cauchy formula, which is well known in the theory of functions of a complex variable. Using another formula [11],

$$\frac{\partial}{\partial\bar{z}}\frac{1}{z} = \pi\delta(\mathbf{x}),$$

the z derivative of the θ^+-function can be easily calculated from Eq. (3.13) as

$$\frac{\partial\theta^+}{\partial\bar{z}} = \frac{i}{2}\int_C \hat{z}_s\delta(\mathbf{x} - \hat{\mathbf{x}})ds.$$

With this result, we can find the usual and variational derivatives of the θ^+-function,

$$\partial_i\theta^+ = \int_C n_i\delta(\mathbf{x} - \hat{\mathbf{x}})ds, \quad (3.14)$$

$$\frac{\delta\theta^+}{\delta\hat{x}_i} = -n_i\delta(\mathbf{x} - \hat{\mathbf{x}}), \quad (3.15)$$

where $\mathbf{n}$ is the unit normal vector related to the unit tangent vector $\mathbf{t}$ as $n_i = \varepsilon^{ki}t_k$.

We now find the expression for Poisson bracket (3.11) in the phase space of the dynamic variables $\hat{\mathbf{x}}(s, t)$. We

first express the left-hand side of Eq. (3.11) in terms of the bracket $\{\hat{x}_i, \hat{x}'_k\}$,

$$\{\theta^+, \theta^{+\prime}\} = \int\int_C \frac{\delta\theta^+(\mathbf{x})}{\delta\hat{x}_i(s)}\frac{\delta\theta^+(\mathbf{x}')}{\delta\hat{x}_k(s')}\{\hat{x}_i, x'_k\}dsds'.$$

Using Eq. (3.15), we obtain

$$\{\theta^+, \theta^{+\prime}\} = \int\int_C \delta(\mathbf{x}-\hat{\mathbf{x}})\delta(\mathbf{x}'-\hat{\mathbf{x}}')n_i n'_k\{\hat{x}_i, \hat{x}'_k\}dsds'. \quad (3.16)$$

On the other hand, using Eq. (3.14), the right-hand side of Eq. (3.11) can be presented as

$$\nu^{-1}\varepsilon^{ik}\partial_k\theta^+\partial_i\delta(\mathbf{x}-\mathbf{x}') = \nu^{-1}\int\int_C \delta(\mathbf{x}-\hat{\mathbf{x}})\delta(\mathbf{x}'-\hat{\mathbf{x}}')\frac{\partial\delta(s-s')}{\partial s}dsds'. \quad (3.17)$$

Comparing Eqs. (3.16) and (3.17) yields the integral equality

$$\int\int_C \delta(\mathbf{x}-\hat{\mathbf{x}})\delta(\mathbf{x}'-\hat{\mathbf{x}}') \times [\nu n_i n'_k\{\hat{x}_i, \hat{x}'_k\} - \partial_s\delta(s-s')]dsds' = 0,$$

whence it follows that

$$\nu n_i n'_k\{\hat{x}_i, \hat{x}'_k\} = \partial_s\delta(s-s'). \quad (3.18)$$

Because the bracket is skew-symmetric, the general solution of Eq. (3.18) for $\{\hat{x}_i, \hat{x}'_k\}$ can be written as

$$\nu\{\hat{x}_i, \hat{x}'_k\} = n_i n'_k\partial_s\delta(s-s') + t_i n'_k a(s, s') - t'_k n_i a(s', s) + t_i t'_k b(s, s'), \quad (3.19)$$

where $a(s', s)$ and $b(s, s')$ are some structure functions and, in addition, $b(s, s')$ must be antisymmetric,

$$b(s, s') = -b(s', s).$$

The choice of the structure functions $a(s', s)$ and $b(s, s')$ cannot be arbitrary but must be matched with constraint (3.12), which means that $\mathbf{t}^2$ is the integral of motion for contour dynamics models with any Hamiltonian. Geometrically, Eq. (3.12) specifies a surface in the phase space $\hat{\mathbf{x}}(s, t)$ such that all the trajectories of real motions lie on this surface. Similar integrals of motion are known as Casimir invariants, or annihilators, of Poisson brackets, i.e., $\{\mathbf{t}^2, \hat{x}'_k\} = 0$. This immediately implies

$$t_i\partial_s\{\hat{x}_i, \hat{x}'_k\} = 0. \quad (3.20)$$

Inserting Eq. of (3.19) into this condition, we obtain

$$\partial_s a(s, s') = -t_i\frac{\partial n_i}{\partial s}\partial_s\delta(s-s'), \quad (3.21)$$

$$\partial_s b(s, s') = t_i\frac{\partial n_i}{\partial s}a(s', s). \quad (3.22)$$

Solving Eqs. (3.21) and (3.22) for the structure functions $a(s', s)$ and $b(s, s')$, we find

$$a(s, s') = \frac{\partial}{\partial s'}[\kappa'\sigma(s'-s)],$$

$$b(s, s') = \frac{1}{2}(\kappa'^2 + \kappa^2)\sigma(s'-s),$$

where

$$\kappa = n_i\partial_s t_i = -t_i\partial_s n_i$$

is the contour curvature and

$$\sigma(s-s') = \frac{1}{2}\mathrm{sgn}(s-s').$$

Thus, the Poisson bracket in the phase space $\hat{\mathbf{x}}(s, t)$ is expressible as

$$\{\hat{x}_i, \hat{x}'_k\} = \nu^{-1}\Big[n_i n'_k\partial_s\delta(s-s') - t_i n'_k\frac{\partial}{\partial s'}[\kappa'\sigma(s-s')] + t'_k n_i\frac{\partial}{\partial s}[\kappa\sigma(s'-s)] + \frac{1}{2}t_i t'_k(\kappa'^2 + \kappa^2)\sigma(s-s')\Big]. \quad (3.23)$$

Now, the equations of contour dynamics can be written in the Hamiltonian form as

$$\partial_t\hat{x}_i = \{\hat{x}_i, H\} = \nu^{-1}\Big[n_i\frac{\partial}{\partial s}\Big(n_k\frac{\delta H}{\delta\hat{x}_k}\Big) + t_i\int_C\kappa'\sigma(s-s')\frac{\partial}{\partial s'}\Big(n'_k\frac{\delta H}{\delta\hat{x}'_k}\Big)ds' + n_i\frac{\partial}{\partial s}\kappa\int_C\sigma(s'-s)t'_k\frac{\delta H}{\delta\hat{x}'_k}ds' + \frac{1}{2}t_i\int_C(\kappa'^2 + \kappa^2)\sigma(s'-s)t'_k\frac{\delta H}{\delta\hat{x}'_k}ds'\Big]. \quad (3.24)$$

We emphasize that constraint (3.12) must be used only after all the variational derivatives are taken in Eqs. (3.24).

In most fluid dynamics models commonly arising in applications, the Hamiltonians are constructed such that

$$t_i\frac{\delta H}{\delta\hat{x}_k} = 0.$$

In this case, Eqs. (3.24) can be presented as

$$n_i\left(\partial_t\hat{x}_i - \nu^{-1}\frac{\partial}{\partial s}\frac{\delta H}{\delta\hat{x}_i}\right) = 0. \tag{3.25}$$

Recalling that in these models

$$\frac{\delta H}{\delta\hat{x}_k} = n_k\nu\hat{\psi},$$

where $\hat{\psi}$ is the streamfunction given on the contour, we obtain from Eq. (3.25) the equations of contour dynamics in the traditional form

$$n_i\partial_t\hat{x}_i = \frac{\partial\hat{\psi}}{\partial s}.$$

The equations of motion of this type were used in [4] to derive the equations of contour dynamics in the weak-curvature approximation for the Hasegawa–Mima model of plasma.

To eliminate the constraint from the Hamiltonian formulation of the contour dynamics, we introduce two new variables φ and ρ as

$$t_1 = \rho\cos\varphi, \quad t_2 = \rho\sin\varphi, \tag{3.26}$$

where $\varphi(t, s)$ is the inclination angle of the unit tangent vector **t** to the axis x_1. In terms of the new variables, constraint (3.12) becomes

$$\rho = 1.$$

Following [12], we define the total Hamiltonian as the superposition

$$H_D = H + \lambda_i I_i$$

involving the original Hamiltonian H and a linear combination of the constraints

$$I_i = \int_C t_i ds = 0,$$

with λ_i being some multipliers that must be determined. The constraints of this type are not a prerogative of closed contours for which the identities

$$\int_C t_i ds \equiv \int_C \frac{\partial\hat{x}_i}{\partial s} ds \equiv 0$$

are quite evident. The same constraints are also valid for open contours if we assume that the contours are closed at infinity. In what follows, for simplicity, we consider an open contour C running in the x_1-direction from $-\infty$ to $+\infty$. We note that in the weak-curvature approximation, the descriptions of models with closed and open contours are locally equivalent. In this situation, the results obtained for open contours can be extended to closed ones.

The multipliers λ_i can be determined from the requirement that the equation of motion for the variable φ on the surface of the constraint $\rho = 1$ must be defined by the Poisson bracket $\{\varphi, \varphi'\}$ as

$$\partial_t\varphi = \{\varphi, H_D\} = \int_{-\infty}^{\infty}\{\varphi, \varphi'\}\frac{\delta H_D}{\delta\varphi'}ds'. \tag{3.27}$$

Using the formulas for the variational derivatives

$$\begin{aligned}\frac{\delta\varphi}{\delta\hat{x}_i'} &= \frac{n_i}{\rho^2}\partial_s\delta(s-s'),\\ \frac{\delta\rho}{\delta\hat{x}_i'} &= \frac{t_i}{\rho}\partial_s\delta(s-s'),\end{aligned} \tag{3.28}$$

we find that

$$\begin{aligned}\partial_t\varphi = \{\varphi, H_D\} &= \int_{-\infty}^{\infty}\frac{\partial\varphi}{\partial\hat{x}_i''}\{\hat{x}_i'', \hat{x}_i'\}\frac{\delta H_D}{\delta\hat{x}_k'}ds''ds'\\ &= -\int_{-\infty}^{\infty} n_i\frac{\partial\{\hat{x}_i, \hat{x}_k'\}}{\partial s}\frac{\partial}{\partial s'}\left(n_k'\frac{\delta H_D}{\delta\varphi'} + t_k'\frac{\delta H_D}{\delta\rho'}\right)ds'.\end{aligned} \tag{3.29}$$

Integration by parts brings Eq. (3.29) to the form

$$\begin{aligned}\partial_t\varphi = &\int_{-\infty}^{\infty} n_i n_k'\frac{\partial^2\{\hat{x}_i, \hat{x}_k'\}}{\partial s\partial s'}\frac{\delta H_D}{\delta\varphi'}ds'\\ &- n_i\frac{\partial\{\hat{x}_i, \hat{x}_k'\}}{\partial s}\left(n_k'\frac{\delta H_D}{\delta\varphi'} + t_k'\frac{\delta H_D}{\delta\rho'}\right)\Bigg|_{-\infty}^{+\infty}.\end{aligned} \tag{3.30}$$

Under the assumption that the perturbation on the contour vanishes at infinity, and therefore, φ and its derivatives tend to zero as $s \longrightarrow \pm\infty$, the last term in Eq. (3.30) can be written as

$$\begin{aligned}&n_i\frac{\partial\{\hat{x}_i, \hat{x}_k'\}}{\partial s}\left(n_k'\frac{\delta H_D}{\delta\varphi'} + t_k'\frac{\delta H_D}{\delta\rho'}\right)\Bigg|_{\pm\infty}\\ &= \nu^{-1}\left(\varphi_{sss} + \frac{1}{2}\varphi_s^3\right)\left(\frac{\delta H}{\delta\rho}\Bigg|_{\pm\infty} + \lambda_1\right).\end{aligned}$$

In accordance with Eqs. (3.28), we have

$$\begin{aligned}\{\varphi, \varphi'\} &= \int_{-\infty}^{\infty}\frac{\delta\varphi}{\delta\hat{x}_i''}\frac{\delta\varphi'}{\delta\hat{x}_k'''}\{\hat{x}_i'', \hat{x}_k'''\}ds''ds'''\\ &= n_i n_k'\frac{\partial^2\{\hat{x}_i, \hat{x}_k'\}}{\partial s\partial s'},\end{aligned} \tag{3.31}$$

and it is therefore easy to conclude that Eq. (3.30) can be rewritten in form (3.27) only if the last term in Eq. (3.30) can be eliminated. There is no way of doing this except by setting

$$\lambda_1 = -\frac{\delta H}{\delta\rho}\Bigg|_{\pm\infty}.$$

Because the theory is independent of λ_2, this multiplier can be chosen arbitrarily without affecting the equation of motion. For simplicity, we put $\lambda_2 = 0$.

The explicit form of the Poisson bracket $\{\varphi, \varphi'\}$ can be found by inserting Poisson bracket (3.23) into Eq. (3.31) and by using the Frenet formulas

$$\partial_s t_i = \kappa n_i, \quad \partial_s n_i = -\kappa t_i, \quad \kappa = \varphi_s. \tag{3.32}$$

By a direct calculation, we obtain

$$\{\varphi, \varphi'\} = -\nu^{-1}\Big[\partial_s^3\delta(s-s') + 2\varphi_s\partial_s(\varphi_s\delta(s-s'))$$
$$+ \sigma(s-s')\Big(\varphi_s'\Big(\varphi_{sss} + \frac{1}{2}\varphi_s^3\Big) + \varphi_s\Big(\varphi_{sss}' + \frac{1}{2}\varphi_s'^3\Big)\Big)\Big].$$

Thus, we have obtained the Poisson bracket for one more Hamiltonian version of contour dynamics. The corresponding equation of motion (3.29) can now be written as

$$\partial_t\varphi = \{\varphi, H_D\} = -\nu^{-1}\Big[\partial_s^3\frac{\delta H_D}{\delta\varphi} + 2\varphi_s\partial_s\varphi_s\frac{\delta H_D}{\delta\varphi}$$
$$+ \Big(\varphi_{sss} + \frac{1}{2}\varphi_s^3\Big)\int_{-\infty}^{\infty}\sigma(s-s')\varphi_s'\frac{\delta H_D}{\delta\varphi'}ds' \tag{3.33}$$
$$+ \varphi_s\int_{-\infty}^{\infty}\sigma(s-s')\Big(\varphi_{sss}' + \frac{1}{2}\varphi_s'^3\Big)\frac{\delta H_D}{\delta\varphi'}ds'\Big].$$

Because the constraint $\rho = 1$ can now be imposed directly on the total Hamiltonian H_D before evaluating the Poisson bracket, Dirac's total Hamiltonian is given by

$$H_D = \left[H - \frac{\delta H}{\delta\rho}\Big|_{s=\infty}\int_{-\infty}^{\infty}\cos\varphi ds\right]_{\rho=1}. \tag{3.34}$$

3.2. Piecewise-Uniform Models with Vorticity and Density Jumps

When a piecewise-uniform model admits density jumps, i.e., $\rho^+ \neq \rho^-$, the vortex sheet density

$$\mu(s,t) = (\rho^-\hat{v}_i^- - \rho^+\hat{v}_i^+)t_i, \quad \hat{v}_i^\pm = v_i^\pm\big|_{\mathbf{x}=\hat{\mathbf{x}}}$$

is no longer a constant of motion. In this case, the evolution of the contour is therefore defined in the phase space of two variables θ^+ and β, where in accordance with Eqs. (3.8) and (3.14), β is related to μ as

$$\beta = (\rho^+v_k^+ - \rho^-v_k^-)\varepsilon^{ik}\partial_i\theta^+ = \int_C\mu(s,t)\delta(\mathbf{x}-\hat{\mathbf{x}})ds.$$

Inserting Eqs. (3.9) and (3.10) into Eqs. (3.2)–(3.4) gives the Poisson brackets

$$\{\theta^+, \theta^{+\prime}\} = 0, \tag{3.35}$$

$$\{\theta^+, \beta'\} = \varepsilon^{ik}\partial_k\theta^+\partial_i\delta(\mathbf{x}-\mathbf{x}'), \tag{3.36}$$

$$\{\beta, \beta'\} = \nu\varepsilon^{ik}\partial_k\theta^+\partial_i\delta(\mathbf{x}-\mathbf{x}') + \varepsilon^{ik}\partial_k\beta\partial_i\delta(\mathbf{x}-\mathbf{x}'), \tag{3.37}$$

where $\nu = \rho^+\omega^+ - \rho^-\omega^-$.

The reformulation of contour dynamics from the (θ^+, β) phase space into the $(\hat{\mathbf{x}}, \mu)$ phase space is carried out in much the same way as in the previous subsection. Following this procedure, we obtain from Eqs. (3.35)–(3.37) that the Poisson brackets $\{\hat{x}_i, \hat{x}_k'\}$ and $\{\hat{x}_i, \mu'\}$ satisfy

$$n_in_k'\{\hat{x}_i, \hat{x}_k'\} = 0, \tag{3.38}$$

$$n_i\partial_s[\mu't_k'\{\hat{x}_i, \hat{x}_k'\}] - n_i\{\hat{x}_i, \mu'\} = \partial_s\delta(s-s'), \tag{3.39}$$

$$\partial_s\partial_s'[\mu'\mu t_it_k'\{\hat{x}_i, \hat{x}_k'\}] - \partial_s[\mu t_i\{\hat{x}_i, \mu'\}] - \partial_s'[\mu't_i'\{\mu, \hat{x}_i'\}] + \{\mu, \mu'\} = \nu\partial_s\delta(s-s'). \tag{3.40}$$

Finding the Poisson brackets must be matched with constraint (3.12). As noted above, this constraint means that the quantity $\mathbf{t}^2$ is a Casimir invariant and hence commutes with the variables making up the basis of the phase space. Therefore, condition (3.20) must be complemented by one more condition

$$t_i\partial_s\{\hat{x}_i, \mu'\} = 0. \tag{3.41}$$

Solving (3.38)–(3.40) with conditions (3.20) and (3.41), we obtain

$$\{\hat{x}_i, \hat{x}_k'\} = 0, \tag{3.42}$$

$$\{\hat{x}_i, \mu'\} = -n_i\partial_s\delta(s-s') + t_i\partial_s'[\kappa'\sigma(s-s')], \tag{3.43}$$

$$\{\mu, \mu'\} = \nu\partial_s\delta(s-s') + \partial_s\partial_s'[(\kappa'\mu + \kappa\mu')\sigma(s-s')]. \tag{3.44}$$

To eliminate the constraints, by analogy with the previous subsection, we introduce two new variables φ and ρ in accordance with Eq. (3.26) under the constraint $\rho = 1$. The Poisson brackets on the $(\hat{x}_i, \mu)$ phase space can be easily transformed into the (φ, μ) space. In fact, only the first two brackets (3.42) and (3.43), where the dynamical variables $\hat{x}_i$ appear, must be reformulated. The required formulas can be obtained using Eq. (3.28) and take the form

$$\{\varphi, \varphi'\} = n_in_k'\frac{\partial}{\partial s}\frac{\partial}{\partial s'}\{\hat{x}_i, \hat{x}_k'\}, \tag{3.45}$$

$$\{\varphi, \mu'\} = n_i\frac{\partial}{\partial s}\{\hat{x}_i, \mu'\}. \tag{3.46}$$

Inserting the Poisson bracket (3.42) and (3.43) in Eqs. (3.45) and (3.46) and using Frenet formulas (3.32), we obtain

$$\{\varphi, \varphi'\} = 0,$$

$$\{\varphi, \mu'\} = -\partial_s^2\delta(s-s') + \varphi_s\partial_s'[\varphi_s'\sigma(s-s')],$$

$$\{\mu, \mu'\} = \nu\partial_s\delta(s-s') + \partial_s\partial_s'[(\varphi_s'\mu + \varphi_s\mu')\sigma(s-s')].$$

If we restrict our consideration to open contours running from $-\infty$ to $+\infty$ in the x_1-direction, the corresponding Dirac's total Hamiltonian H_D can be determined in the same way as in the previous subsection, with the same results as in Eq. (3.34). Thus, contour dynamics corresponding to a given system of the Poisson brackets is described by the equations

$$\partial_t\varphi = \{\varphi, H_D\} = -\frac{\partial^2}{\partial s^2}\frac{\delta H_D}{\delta\mu} - \varphi_s\int_{-\infty}^{\infty}\varphi_s'\sigma(s-s')\frac{\partial}{\partial s'}\frac{\delta H_D}{\delta\mu'}ds',$$

$$\partial_t\mu = \{\mu, H_D\} = \frac{\partial^2}{\partial s^2}\frac{\delta H_D}{\delta\varphi} + \frac{\partial}{\partial s}\left[\varphi_s\int_{-\infty}^{\infty}\varphi_s'\sigma(s-s')\frac{\delta H_D}{\delta\mu'}ds'\right] + \nu\frac{\partial}{\partial s}\frac{\delta H_D}{\delta\mu} - \frac{\partial}{\partial s}\left[\int_{-\infty}^{\infty}(\varphi_s'\mu + \varphi_s\mu')\sigma(s-s')\frac{\partial}{\partial s'}\frac{\delta H_D}{\delta\mu'}ds'\right].$$

4. *N*-PETAL STRUCTURES IN TWO-DIMENSIONAL FLUID MODELS

4.1. Hamiltonian Formulation of the Problem

The simplest models that admit a direct application of the obtained results are a quasigeostrophic barotropic model, a model of plasma based on the Hasegawa–Mima equation, and an axial model of electronic vortices. These models are known [1, 13] to belong to vorticity-like systems governed by the equation

$$\partial_t\omega + (\partial_1\psi)\partial_2\omega - (\partial_2\psi)\partial_1\omega = 0, \quad (4.1)$$

where the potential vorticity ω and the streamfunction ψ are functions of the x_1 and x_2 coordinates in the horizontal plane and are related by

$$\omega = \left(\Delta - \frac{1}{r^2}\right)\psi,$$

where r is an internal scale treated as the Rossby deformation radius and $\Delta = \partial_1^2 + \partial_2^2$ is the two-dimensional Laplace operator. For the Hasegawa–Mima model, the parameter r is treated as the Larmor ion radius r_L given by

$$r_L = \left(\frac{m_iT_ec^2}{B_0^2e^2}\right)^{1/2}, \quad (4.2)$$

where m_i is the ion mass, T_e is the electron temperature, e is the electron charge, c is the velocity of light, and B_0 is the induction of an ambient uniform magnetic field. The electric potential Φ and the electron number density n_e can be expressed in terms of the streamfunction as

$$\Phi = \frac{B_0}{c}\psi, \quad n_e = n_0\exp\left(\frac{B_0e}{T_ec}\psi\right), \quad (4.3)$$

where n_0 is the unperturbed plasma density.

In the axial model of electronic fluid with constant density, the parameter r must be chosen as the skin layer width r_S given by

$$r_S = c\left(\frac{m_e}{4\pi ne^2}\right)^{1/2},$$

where m_e is the electron mass and n is the constant plasma density. In this model, the magnetic field B is related to the streamfunction ψ by

$$B = -\frac{4\pi ne}{c}\psi.$$

It is easy to verify that the vorticity-like models governed by equation of motion (4.1) are Hamiltonian, namely, characterized by the Poisson bracket of the same type as Eq. (3.4),

$$\{\omega, \omega'\} = \varepsilon^{ki}\partial_i\omega\partial_k\delta(\mathbf{x}-\mathbf{x}'),$$

and have the Hamiltonian

$$H = -\frac{1}{2}\int\psi\omega d\mathbf{x},$$

which can be rewritten solely in terms of the potential vorticity as

$$H = -\frac{1}{2}\int\omega\omega' G(\mathbf{x}, \mathbf{x}')d\mathbf{x}d\mathbf{x}'.$$

Green's function G is found as the solution of the problem

$$\left(\Delta - \frac{1}{r^2}\right)G = \delta(\mathbf{x}-\mathbf{x}')$$

and has the explicit form

$$G(\mathbf{x}, \mathbf{x}') = -\frac{1}{2\pi}K_0\left(\frac{|\mathbf{x}-\mathbf{x}'|}{r}\right),$$

where K_0 denotes the modified Bessel function of zero-order.

As already proved, the reduction of the description of vorticity-like systems in Eq. (4.1) to contour dynamics becomes possible if the entire fluid can be decomposed into domains such that each of them moves with the fluid and has a constant potential vorticity. For the unbounded fluid with a single vortex patch is embedded in a background shear flow, the distribution of the potential vorticity ω can be presented as

$$\omega = \omega^+\theta^+ + \omega^-\theta^-, \quad \theta^+ + \theta^- = 1,$$

where ω^+, θ^+ and ω^-, θ^- have the same meaning as before. The corresponding Hamiltonian is then given by

$$H = -\frac{\nu^2}{2}\int \theta^+\theta^{+\prime} G(\mathbf{x}, \mathbf{x}')d\mathbf{x}d\mathbf{x}', \quad \nu = \omega^+ - \omega^-.$$

After some manipulations, this can be expressed in terms of contour-dynamical variables as

$$H = \frac{(r\nu)^2}{2}\int \theta^+\theta^{+\prime}[\delta(\mathbf{x}-\mathbf{x}') - \Delta G]d\mathbf{x}d\mathbf{x}'$$

$$= \frac{(r\nu)^2}{2}\left(\int \theta^+ d\mathbf{x} + \int G\frac{\partial\theta^+}{\partial x_i}\frac{\partial\theta^{+\prime}}{\partial x_i'}d\mathbf{x}d\mathbf{x}'\right) \qquad (4.4)$$

$$= \frac{(r\nu)^2}{2}\left(\int \theta^+ d\mathbf{x} + \int_C G(\hat{\mathbf{x}}, \hat{\mathbf{x}}')t_i t_i' ds ds'\right).$$

We note that the first integral

$$I = \int \theta^+ d\mathbf{x} = -\frac{1}{2}\int_C \hat{x}_i n_i ds$$

has a simple geometric meaning of the vortex patch area and is a Casimir invariant (belongs to the annihilator of Poisson bracket (3.23)). Therefore, it does not affect the equation of motion and can be omitted in defining the Hamiltonian. Thus, we obtain from Eq. (4.4)

$$H = -\frac{(r\nu)^2}{4\pi}\int\int_C K_0\left(\frac{|\mathbf{x}-\mathbf{x}'|}{r}\right)t_i t_i' ds ds'. \qquad (4.5)$$

The following analysis is carried out in the weak curvature approximation where the characteristic curvature radius R of the contour is much larger than the internal scale (deformation radius) r, which allows the introduction of a small parameter $\varepsilon = r/R$. In this case, it is possible to develop the local representation for the Hamiltonian in Eq. (4.5),

$$H = \int_C h[s; \rho, \varphi]ds, \qquad (4.6)$$

where the Hamiltonian density h is expressible as a power series in the small parameter ε,

$$h = \frac{(r\nu)^2}{4\pi}$$

$$\times\left(-\pi r\rho + \frac{3}{8}\pi\frac{r^3}{\rho}\varphi_s^2 - \frac{8}{3}\frac{r^4}{\rho^3}\varphi_s(\rho_s\varphi_s - \rho\varphi_{ss}) + O(\varepsilon^4)\right).$$

Inserting Eq. (4.6) in Eq. (3.34) and neglecting the fourth-order terms in ε, we find Dirac's Hamiltonian for contour dynamics in vorticity-like systems under consideration,

$$H_D = \frac{r^3\nu^2}{4}\int_C\left(\cos\varphi + \frac{3}{8}r^2\varphi_s^2\right)ds. \qquad (4.7)$$

It is interesting to note that because $H \sim O(\varepsilon^2)$, the main contribution to Dirac's Hamiltonian is given solely by the constraint functional. In the leading-order approximation, therefore, Eq. (4.7) becomes

$$H_D = \frac{r^3\nu^2}{4}\int_C \cos\varphi ds. \qquad (4.8)$$

In accordance with Eq. (3.33), we now obtain the contour dynamics equation

$$\partial_t\varphi = \{\varphi, H_D\} = -\frac{r^3\nu}{4}\left(\varphi_{sss} + \frac{1}{2}\varphi_s^3\right). \qquad (4.9)$$

4.2. Steadily Rotating Localized Vortex Structures

We consider solutions of Eq. (4.9) that manifest themselves as stationary vortex structures rotating with a constant angular velocity ω_0. These solutions have the form

$$\varphi(t, s) = \tilde{\varphi}(s - ct) - \omega_0 t, \qquad (4.10)$$

where $\omega_0 > 0$ for the clockwise rotation and $\omega_0 < 0$ for the counterclockwise rotation. Inserting Eq. (4.10) in Eq. (4.9) and choosing the spatial scale R as

$$R = \frac{r}{2}\left(\frac{\nu}{\omega_0}\right)^{1/3}, \qquad (4.11)$$

we introduce the dimensionless variables

$$\tilde{s} = \frac{s - ct}{R}, \quad \tilde{\kappa} = \frac{\partial\tilde{\varphi}}{\partial\tilde{s}}$$

and obtain the equation

$$\left(\frac{\partial\tilde{\kappa}}{\partial\tilde{s}}\right)^2 = -\frac{1}{4}\tilde{\kappa}^4 + c_1\tilde{\kappa}^2 + \tilde{\kappa} + c_2, \qquad (4.12)$$

where c_2 is an integration constant and $c_1 = c(2\omega_0 R)^{-1}$.

According to the theory of elliptic functions [14], Eq. (4.12) has two sets of periodic solutions expressed in terms of elliptic functions,

$$\tilde{\kappa} = b + \frac{a-b}{1-\alpha \mathrm{F}(\lambda\tilde{s}|m)}, \tag{4.13}$$

where F is one of the Jacobi elliptic functions (either sn or dn) and m is the parameter of these functions, with the vertical line symbolizing the m-dependence. We note that depending on the type of the Jacobi elliptic functions, the independent basic parameters α and m parametrize all the others parameters a, b, λ, and consequently, c_1 and c_2.

To derive the equations describing the boundary shape of vortex structures rotating in the horizontal z plane, we must integrate the equation

$$\frac{\partial \hat{z}}{\partial \tilde{s}} = \exp(i\tilde{\varphi}), \tag{4.14}$$

where $\hat{z} = (\hat{x}_1 + i\hat{x}_2)/R$ is the dimensionless complex coordinate of the contour. It can be directly verified that if $\tilde{\kappa}$ satisfies Eq. (4.12), the solution of Eq. (4.14) is given by

$$\hat{z}(\tilde{s}) = 2\left[\frac{\partial \tilde{\kappa}}{\partial \tilde{s}} + i\left(c_1 - \frac{\tilde{\kappa}^2}{2}\right)\right]\exp(i\tilde{\varphi}). \tag{4.15}$$

4.3. Classification of Solutions

In this subsection, we focus our attention on the classification of those solutions of Eq. (4.12) that correspond to multipetal vortex structures without self-intersection of the contour. For this purpose, we perform both analytical and numerical investigation of the problem in Eqs. (4.13) and (4.15), restricting our study to the case where F = sn. As becomes apparent after a close examination, the solutions of the second type with F = dn do not contain vortices without contour self-intersections.

With F = sn, the periodic solution for the contour curvature (4.13) takes the form

$$\tilde{k} = b + \frac{a-b}{1-\alpha \mathrm{sn}(\lambda\tilde{s}|m)}. \tag{4.16}$$

If the independent parameters α and m are considered as basic, all the other parameters a, b, and λ can be expressed as

$$a = -2^{-1/3}\frac{\alpha(1+m-2\alpha^2)}{[(1-m)^2\alpha(m-\alpha^4)]^{1/3}},$$

$$b = 2^{-1/3}\frac{\alpha^2 + m(\alpha^2-2)}{\alpha[(1-m)^2\alpha(m-\alpha^4)]^{1/3}},$$

$$\lambda = 2^{-1/3}\frac{\sqrt{(\alpha^2-m)(1-\alpha^2)}}{[(1-m)^2\alpha(m-\alpha^4)]^{1/3}}.$$

The parameters c_1 and c_2 are expressed in terms of a and b as

$$c_1 = \frac{ba}{2} - \frac{1}{a+b}, \quad c_2 = -\frac{1}{4}(b + a + b^2a^2).$$

We emphasize that the conditions of the contour continuity (smoothness) and reality of solutions to Eq. (4.12) impose the following restrictions on the parameters α and m:

$$0 \le \alpha \le 1, \quad m < \alpha^2. \tag{4.17}$$

It follows from Eq. (4.15) that in order to find the boundary shape we must know the slope angle $\tilde{\varphi}(\tilde{s})$ in addition to the variable $\tilde{\kappa}$. This can be computed by integrating (4.16) along the contour line,

$$\tilde{\varphi}(\tilde{s}_0) = \int_0^{\tilde{s}} \tilde{\kappa}(s)ds = b\tilde{s} + \frac{a-b}{\lambda}\Pi(\alpha^2; \mathrm{am}(\lambda\tilde{s}|m)\big|m) \tag{4.18}$$
$$- 2\mathrm{Im}\{\ln[\mathrm{cn}(\lambda\tilde{s}|m)\sqrt{\alpha^2-m} + i\,\mathrm{dn}(\lambda\tilde{s}|m)\sqrt{1-\alpha^2}]\},$$

where $\Pi(u; \vartheta|m)$ is the incomplete elliptic integral of the third kind and the Jacobi amplitude $\mathrm{am}(u|m)$ is defined by

$$\mathrm{am}(u|m) = \arcsin(\mathrm{sn}(u|m)).$$

As mentioned above, our study is restricted to vortex structures with a finite area bounded by a closed contour without self-intersections. It is worth noting that the elimination of self-intersecting contours corresponding to rather exotic vortex formations from the consideration is motivated by the weak-curvature approximation used in deriving Eq. (4.9), but is not at all dictated by intrinsic reasons of fluid dynamics. In other words, the exact equations of motion for the two-dimensional ideal fluid admit the existence of solutions with such a contour topology.

Obviously, considering such contours requires a generalization of model assumptions in the initial statement of the problem. Because the vortex region becomes multiply connected when the contour admits self-intersections, the corresponding piecewise-constant vorticity distribution can be rather specific. If the topology of the contour self-intersection is known, the vorticity distribution can be easily reproduced because the vorticity jump must remain invariant when going around the contour in one of the directions (see Fig. 1). In essence, the question of whether to include solutions of this type into the framework of our scheme is the question of whether a global behavior of solutions is sensitive to a local violation of the weak-curvature approximation. The answer can be found by comparing numerical and analytical solutions. If these solutions are insensitive, they have every ground for being included and can be improved using various numerical procedures similar to the "contour surgery" proposed in [2].

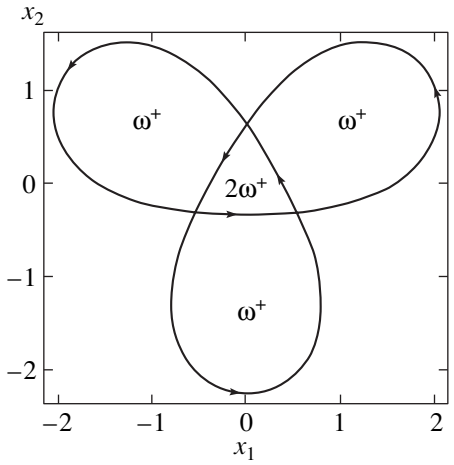


Fig. 1. The geometry of a three-petal vortex region of piecewise-constant vorticity with a self-intersecting contour. The vorticity distribution is ω^+ in petals and $2\omega^+$ in the core, so the jump in vorticity is the invariant ω^+ in tracing the contour.

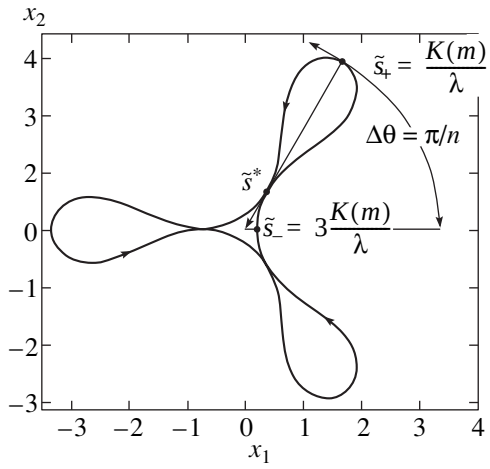


Fig. 2. Three-petal vortex structure. The point $\tilde{s}_+ = K(m)/\lambda$ lies in the petal tip and $\tilde{s}_- = 3K(m)/\lambda$ lies between the petals. $\tilde{s}^*$ is the self-contacting point of the contour.

Because the contour is closed and its curvature is a periodic function of $\tilde{s}$, the boundary shape of the vortices must have an n-petal structure. An example of this structure is given in Fig. 2. From this figure and the analysis of Eq. (4.16), it is clear that the contour curvature of the n-petal vortex structure, being an oscillatory function with the period $4K(m)/\lambda$, has extrema at the points

$$\tilde{s}_- = (4j-1)\frac{K(m)}{\lambda}, \quad \tilde{s}_+ = (4j-3)\frac{K(m)}{\lambda},$$
$$j = 1, 2, \ldots, n,$$

where $K(m)$ is the complete elliptic integral of the first kind. At these points, the contour curvature takes the extreme values

$$\tilde{\kappa}_+ = b + \frac{a-b}{1-\alpha} = 2^{-1/3}\frac{m(\alpha+2)-\alpha(1+2\alpha)}{[(1-m)^2\alpha(m-\alpha^4)]^{1/3}},$$
$$\tilde{\kappa}_- = b + \frac{a-b}{1+\alpha} = 2^{-1/3}\frac{m(\alpha-2)-\alpha(1-2\alpha)}{[(1-m)^2\alpha(m-\alpha^4)]^{1/3}}. \quad (4.19)$$

The subscript notation $\mp$ means that $f_\mp = f(\tilde{s}_\mp)$. The relative position of the turning points $\tilde{s}_-$ and $\tilde{s}_+$ depends on the parameters α and m. To establish which of them is at the tip of the petal and which is in the trough between the petals, it is necessary to compute the distances between these points and the symmetry center (the coordinate origin). For this purpose, we introduce ρ and θ as the polar coordinates,

$$\hat{z}(\tilde{s}) = \rho e^{i\theta}.$$

In accordance with Eqs. (4.15) and (4.12), the variables ρ and θ are then given by

$$\rho^2 = 4(c_1^2 + c_2 + \tilde{\kappa}), \quad (4.20)$$

$$\theta = \arctan\left(\frac{\partial\tilde{\kappa}/\partial\tilde{s}}{c_1 - \tilde{\kappa}^2/2}\right) + \tilde{\varphi}. \quad (4.21)$$

Expressing c_1 and c_2 in terms of α and m and using Eq. (4.19), we find from Eq. (4.20) that

$$\rho_+^2 = 2^{2/3}\frac{[m(1+2\alpha)-\alpha^3(\alpha+2)]^2}{\alpha(m-\alpha^4)[(1-m)^2\alpha(m-\alpha^4)]^{1/3}},$$

$$\rho_-^2 = 2^{-1/3}\frac{[m(1-2\alpha)-\alpha^3(\alpha-2)]^2}{\alpha(m-\alpha^4)[(1-m)^2\alpha(m-\alpha^4)]^{1/3}}.$$

The relative position of the turning points depends on whether 1 is greater or less than the ratio

$$\left(\frac{\rho_-}{\rho_+}\right)^2 = 1 + \frac{8\alpha(\alpha^2-m)(m-\alpha^4)}{[m(1+2\alpha)-\alpha^3(\alpha+2)]^2}.$$

It is easy to see that the inequality $\alpha^4 \le m \le \alpha^2$ entails the inequality $\rho_- \ge \rho_+$; therefore, in this interval of the parameters, the tops of the petals lie at the points $\tilde{s}_-$. In the event that $m \le \alpha^4$ (and consequently, the reverse inequality $\rho_- \le \rho_+$ holds), the tips of the petals lie at the points $\tilde{s}_+$.

It is amply clear that in the region of the permissible parameters (4.17), not all solutions (4.16) correspond to vortex structures with closed contours. The condition under which periodic solution (4.16) corresponds to a closed contour can be formulated as

$$\Delta\theta = \theta_- - \theta_+ = \frac{\pi}{n}. \quad (4.22)$$

This condition has a simple geometrical interpretation shown in Fig. 2. From this figure, it is easy to see that

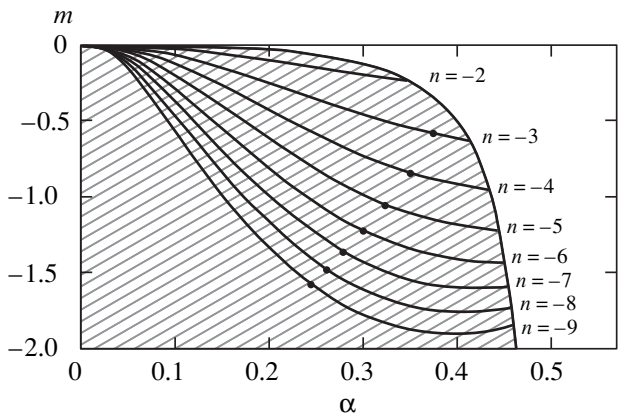


Fig. 3. The family of n-petal vortex regimes in the plane αm. The characteristic curves assign the dependence $m_n(\alpha)$ for $n = -1, -2, \ldots, -9$. The limit points where the corresponding vortex structure has the contour with a self-contact are marked as •.

$2\Delta\theta$ is merely the angular distance between neighboring petals. To evaluate its value, it suffices to note that the position vector and the tangent one are mutually orthogonal at the turning points. It thus follows from Eq. (4.21) that

$$\theta_{\pm} = \tilde{\varphi}_{\pm} + \frac{\pi}{2}\Delta_{\pm},$$

where the sign function $\Delta_{\pm}$ is defined as

$$\Delta_{\pm} = \operatorname{sgn}[m(1 \pm 2\alpha) - \alpha^3(\alpha \pm 2)]. \qquad (4.23)$$

The expression for $\tilde{\varphi}_{\pm}$ can be easily found from Eq. (4.18) as

$$\tilde{\varphi}_{\pm} = \frac{4j - 2 \mp 1}{\lambda} \times [bK(m) + (a-b)\Pi(\alpha^2|m)] - \pi, \qquad (4.24)$$

where

$$\Pi(u|m) = \Pi\left(u; \frac{\pi}{2}\middle|m\right)$$

is the complete elliptic integral of the third kind.

Equations (4.23) and (4.24) allow us to rewrite Eq. (4.22) as

$$bK(m) + (a-b)\Pi(\alpha^2|m) = \frac{\pi}{2}\lambda\left(\frac{1}{n} - \Delta\right), \qquad (4.25)$$

where

$$\Delta = \frac{1}{2}(\Delta_- - \Delta_+) = \frac{1}{2}\{\operatorname{sgn}[m(1-2\alpha) - \alpha^3(\alpha-2)] - \operatorname{sgn}[m(1+2\alpha) - \alpha^3(\alpha+2)]\}.$$

The analysis shows that Eq. (4.25) has solutions in the form of n-petal structures in the region

$$m \le \alpha^3 \frac{\alpha - 2}{1 - 2\alpha},$$

where $\Delta = 0$ for $n \le -2$. In Fig. 3, this region is marked by a shaded background. The solutions are presented by the characteristic curves that determine the dependence $m_n(\alpha)$ for every n. For a fixed n, the multipetal structure can therefore be described by a single parameter α. The vortex shapes for $n = 2, 3$ depending on α are shown in Figs. 4 and 5. For every n-petal regime, the characteristic curve has a limit point where the corresponding vortex structure has a self-contacting contour. Solutions without intersections of contours are on the left of the point and those with self-intersections are on the right.

A prerequisite to the formation of a self-contact in a contour can be formulated on the basis of geometrical

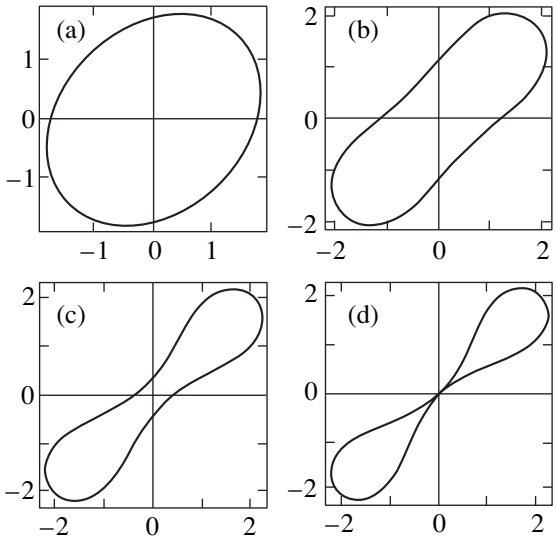


Fig. 4. Shapes of boundaries for double-petal vortex structures: $\alpha = 0.050$ (a), 0.200 (b), 0.300 (c), 0.353 (d).

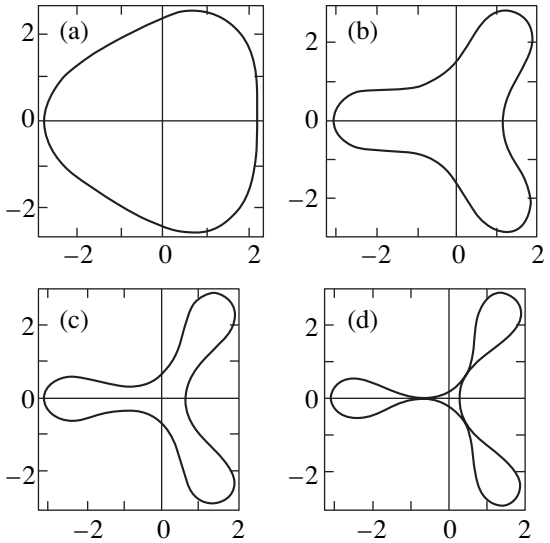


Fig. 5. Shapes of boundaries for three-petal vortex structures: $\alpha = 0.050$ (a), 0.200 (b), 0.300 (c), 0.371 (d).

Values of parameters characterizing the limiting regimes

n	α	m	$\tilde{\kappa}_-$	$\tilde{\kappa}_+$	ρ_-	ρ_+
–2	0.352823	–0.245778	–0.456761	1.79081	0	2.12018
–3	0.371469	–0.580662	–0.820287	1.95339	0.193635	3.3365
–4	0.348897	–0.844407	–1.01623	2.03108	0.42446	3.51701
–5	0.323504	–1.0545	–1.15832	2.08942	0.635998	3.65998
–6	0.300157	–1.22456	–1.27263	2.13903	0.83048	3.78634
–10	0.231285	–1.66566	–1.60011	2.29932	1.49709	4.22362

considerations following from Fig. 2. At the tangency point $\tilde{s}^*$ the angles θ and $\tilde{\varphi}$ are related by

$$\theta(\tilde{s}^*) = \tilde{\varphi}(\tilde{s}^*).$$

Equation (4.21) now implies the condition

$$\tilde{\kappa}^2(\tilde{s}^*) = 2c_1.$$

One more condition is obtained by taking into account that in tracing the contour from the point $\tilde{s}_+$ to the tangency point $\tilde{s}^*$, the tangent vector is rotated through $\pi/2$, and therefore

$$\tilde{\varphi}(\tilde{s}^*) - \tilde{\varphi}_+ = \pi/2.$$

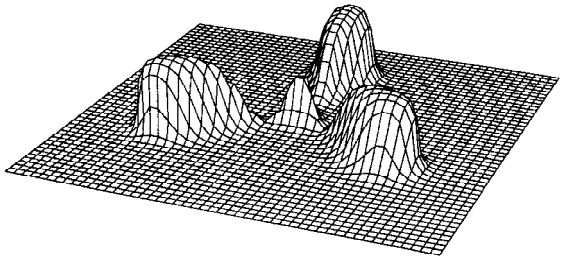


Fig. 6. Surface plot of the streamfunction field for the limiting three-petal vortex structure.

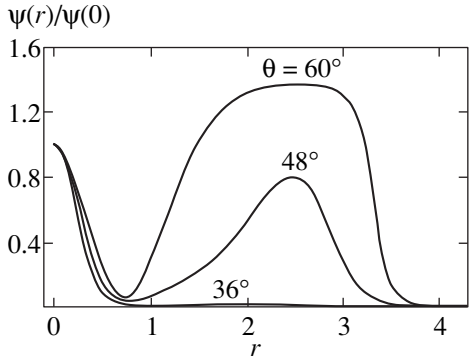


Fig. 7. The radial profile of the streamfunction for the limiting three-petal vortex structure given in Fig. 6. The profiles correspond to the directions $\theta = 60°$, 48°, 36°.

Using the relation

$$\tilde{\varphi}_+ = \frac{\pi}{2}\left(\frac{1}{n} - \Delta\right) - \pi,$$

which follows from Eqs. (4.24) and (4.25) with $j = 1$, we obtain the conditions

$$\tilde{\varphi}(s^*) = \frac{\pi}{2}\left(\frac{1}{n} - \Delta - 1\right),$$

$$\tilde{\kappa}^2(s^*) = 2^{-2/3} \times \frac{2\alpha^4(1+m) + \alpha^2(1+m(m-10)) + 2m(m+1)}{[(1-m)^2\alpha(m-\alpha^4)]^{2/3}}.$$

Together with Eq. (4.25), these conditions fix all the parameters of the limiting regimes presented in the table. In the quasigeostrophic barotropic model, the physical interpretation of ψ is the pressure deviation, and in the plasma model based on the Hasegawa–Mima equation, this quantity characterizes the electric potential. To illustrate the spatial and temporal character of distributions of ψ, we assume for simplicity that the background vorticity is absent, i.e., $\omega^- = 0$. Using the results obtained in Section 4.1, we can then establish the formula

$$\psi(\mathbf{x}) = \omega^+ \int \theta^+{}' G(\mathbf{x}, \mathbf{x}') d\mathbf{x}' = \omega^+ R^2 \times \mathrm{Im} \int_C \left[\frac{1}{|\hat{z} - z|} - \varepsilon^{-1} K_1\left(\frac{|\hat{z} - z|}{\varepsilon}\right)\right] \frac{\hat{z}_s(\hat{\bar{z}} - \bar{z})}{|\hat{z} - z|} ds, \quad (4.26)$$

where

$$z = (x_1 + ix_2)/R \text{ and } \varepsilon = |r/R|.$$

The distribution $\psi(\mathbf{x})/\psi(0)$ associated with the presence of the three-petal vortex of limiting type is calculated in accordance with Eq. (4.26) and is shown in Fig. 6. The radial profiles corresponding to this vortex are presented in Fig. 7.

5. CONCLUDING REMARKS

To gain greater insight into the physical significance of the results and decide in which range of parameters these results can be applicable, we make some estimates for the Hasegawa–Mima model and for the axial model of electronic vortices, in parallel. We note that for these models, the values of the r_L and r_S parameters cover a broad range. According to factual evidence [15], the Larmor ion radius r_L measures 10^3 cm for the interplanetary gas and 10^{-2} cm for the solar corona. Depending on the type of plasma, the skin layer width r_S varies between 5×10^5 and 5×10^{-3} cm.

To illustrate the results obtained in more detail, we consider the Hasegawa–Mima model of plasma with the parameters $T_e = 10^4$ K, $n_0 = 10^{14}$ cm^{-3}, $B_0 = 10^4$ G, and $m_i = 1.67 \times 10^{-24}$ g, which are typical for a low-pressure gas discharge. In accordance with Eq. (4.2), we find $r_L \approx 10^{-2}$ cm. Because the theory of limiting vortex structures has only two control parameters (the angular rotation velocity ω_0 and the vorticity jump $\nu = \omega^+ - \omega^-$), we put $\omega_0 = 10$ s^{-1}, $\omega^- = 0$, and $\omega^+ = 10^6$ s^{-1} in order to calculate some characteristics of a three-petal drift vortex. In this case, Eq. (4.11) gives $R \approx 10 r_L = 10^{-1}$ cm, and therefore, each petal of the vortex structure has the radial length $\rho_+ R \approx 3.3 \times 10^{-1}$ cm. Next, upon numerical integration with $\varepsilon = r_L/R \approx 10^{-1}$, we obtain from Eq. (4.26) that $\psi(0) \approx 5.07 R^2 \omega^+$. Thus, we can estimate the magnitudes of the electric potential Φ and the electron number density n_e at the center of the three-petal drift vortex. It follows from Eq. (4.3) that $\Phi(0) \approx 4.4 \times 10^2$ V and $n_e(0) \approx 1.5 \times 10^{16}$ cm^{-3}.

We note, in closing, some possible generalizations of the Hamiltonian versions of 2D contour dynamics. The technique that we have described can also be used for 3D vortex objects, for example, in quasigeostrophic baroclinic models of geophysical fluid dynamics. The Hamiltonian versions of 2D contour dynamics can be successfully applied to the study of nonplanar models in all the cases where the velocity field is invariant along the vorticity field direction. Typical examples are flows on the sphere and also flows with the rotational and helical spatial symmetry of the vortex field.

This work was partly supported by the Russian Foundation for Basic Research (grant no. 00-05-64019-a).

REFERENCES

1. J. Pedlosky, *Geophysical Fluid Dynamics, 2nd ed.* (Springer-Verlag, New York, 1986).
2. D. G. Dritschel, J. Comput. Phys. **77**, 240 (1988).
3. V. P. Goncharov and V. I. Pavlov, Eur. J. Mech. B/Fluids **19**, 831 (2000).
4. A. V. Gruzinov, Pis'ma Zh. Éksp. Teor. Fiz. **55**, 75 (1992) [JETP Lett. **55**, 76 (1992)].
5. V. I. Arnol'd, Ann. Inst. Fourier (Grenoble) **16**, 319 (1966).
6. P. J. Olver, J. Math. Anal. Appl. **89**, 233 (1982).
7. H. D. I. Abarbanel, D. D. Holm, J. E. Marsden, and T. S. Ratiu, Philos. Trans. R. Soc. London, Ser. A **318**, 349 (1986).
8. V. P. Goncharov and V. I. Pavlov, *Problems of Hydrodynamics in Hamiltonian Description* (Mosk. Gos. Univ., Moscow, 1993).
9. V. P. Goncharov and V. I. Pavlov, Eur. J. Mech. B/Fluids **16**, 509 (1997).
10. V. P. Goncharov and V. I. Pavlov, Nonlinear Processes Geophys. **1**, 219 (1998).
11. E. Madelung, *Die Mathematischen Hilfsmittel des Physikers* (Springer-Verlag, Berlin, 1957; Nauka, Moscow, 1968).
12. P. A. M. Dirac, Proc. R. Soc. London, Ser. A **246**, 326 (1958).
13. V. I. Petviashvili and O. A. Pohotelov, *Solitary Waves in Plasma and Atmosphere* (Énergoatomizdat, Moscow, 1989).
14. *Heigher Transcendental Functions (Bateman Manuscript Project)*, Ed. by A. Erdelyi (McGraw-Hill, New York, 1953, 1953, 1955; Nauka, Moscow, 1965, 1966, 1967), Vols. 1–3.
15. N. A. Krall and A. W. Trivelpiece, *Principles of Plasma Physics* (McGraw-Hill, New York, 1973; Mir, Moscow, 1975).

Journal of Experimental and Theoretical Physics, Vol. 92, No. 4, 2001, pp. 608–616.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 4, 2001, pp. 700–709.

PLASMA, GASES

Distribution Function of the Number of Particles in the Condensate of an Ideal Bose Gas Confined by a Trap

V. A. Alekseev

Lebedev Physical Institute, Russian Academy of Sciences, Leninskiĭ pr. 53, Moscow, 117924 Russia
e-mail: valeks@sci.lebedev.ru

Received September 15, 2000

Abstract—The distribution function $W_0(n_0)$ of the number n_0 of particles in the condensate of an ideal Bose gas confined by a trap is found. It is shown that at the temperature below the critical temperature T_c this function has a Gaussian shape and depends on the trap potential via two parameters only. The center of this function shifts to larger values of n_0 with decreasing temperature and its width tends to zero, which corresponds to the suppression of fluctuations. In the narrow vicinity of the critical temperature $|T - T_c| \le T_c/\sqrt{N}$, where N is the number of particles in the trap, the distribution function changes and at the temperature above the critical one it takes the usual form $W_0(n_0) = [1 - \exp(\mu)]\exp(\mu n_0)$, where μ is the chemical potential in temperature units. In the limit $N \longrightarrow \infty$, this change occurs at a jump. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

After the first experimental observations of Bose condensation of atoms in a magnetic parabolic trap [1], many theoretical papers devoted to the study of this new state of matter have appeared. However, the fundamental question of the distribution function $W_0(n_0)$ of the number of particles in the condensate, i.e., the number n_0 of particles in the ground state with the energy E_0, has not been discussed in fact so far.

In this paper, the distribution function $W_0(n_0)$ of the number of particles in the ground state of an ideal Bose gas confined by a trap is found and its temperature dependence is investigated. It is shown that below the critical temperature, the distribution $W_0(n_0)$ exhibits qualitative changes, which are accompanied by the suppression of fluctuations and the change in the gas heat capacity. When the number N of particles confined by the trap is large (formally, at $N \longrightarrow \infty$), these changes occur at a jump, i.e., acquire the character predicted by Landau for phase transitions in crystals [2].

The distribution $W_0(n_0)$ is obtained by summing

$$W_0(n_0) = \sum_{n_1 + n_2 + \ldots = N - n_0} W(n_0, n_1, \ldots) \qquad (1)$$

and the Gibbs distribution

$$W(n_0, n_1, \ldots) = \frac{1}{S}\exp(-\varepsilon_0 n_0 - \varepsilon_1 n_1 - \ldots), \qquad (2)$$

where $\varepsilon_k = E_k/T$, E_k is the energy of the kth state of a particle in the trap, T is the temperature in the energy units, and the total energy of the gas $E = E_0 n_0 + E_1 n_1 + \ldots$ is specified by the number n_k of particles in the states with the energy E_k. The summation in Eq. (1) is performed over all positive values of n_1, n_2, … (except n_0), which satisfy the condition

$$n_1 + n_2 + \ldots = N - n_0. \qquad (1a)$$

The statistical sum S (normalization) can be found by summation of

$$S = \sum_{n_0 + n_1 + \ldots = N} W(n_0, n_1, \ldots) \qquad (3)$$

over all positive values of n_k (including n_0), which satisfy the condition

$$n_0 + n_1 + \ldots = N. \qquad (4)$$

Usually, instead of fulfillment the exact condition (4), an additional parameter μ (the chemical potential in the temperature units) is introduced into the Gibbs distribution,

$$W(n_0, n_1, \ldots) = S^{-1} \times \exp(\mu(n_0 + n_1 + \ldots) - \varepsilon_0 n_0 - \varepsilon_1 n_1 - \ldots), \qquad (5)$$

and the statistical sum is found by summation over all $n_k \ge 0$. As a result, we obtain

$$S = \prod_k (1 - \exp(\mu - \varepsilon_k))^{-1},$$
$$W(n_0, n_1, \ldots) = \prod_k W_k(n_k), \qquad (6a)$$
$$W_k(n_k) = (1 - \exp(\mu - \varepsilon_k))\exp((\mu - \varepsilon_k)n_k).$$

In particular, when the energy is measured relative to the ground-state energy, the distribution of the number of particles in the ground state has the form

$$W_0(n_0) = (1 - e^{\mu})e^{\mu n_0}. \quad (6b)$$

Expressions (6a) and (6b) directly yield the average value of the number $\tilde{n}_k$ of particles in the state with the energy E_k (hereafter denoted by a tilde)

$$\tilde{n}_k = (e^{\varepsilon_k - \mu} - 1)^{-1}. \quad (7)$$

After that, the chemical potential μ is defined by the condition

$$\sum_k \tilde{n}_k = N. \quad (8)$$

Therefore, the distribution (6a) satisfies not the exact condition (4) but condition (8), which is fulfilled only for average values.

Meanwhile, the distribution (6a) leads to a contradiction. Namely, (6a) yields the well-known relations

$$\langle n_k^2 \rangle = 2\tilde{n}_k^2 + \tilde{n}_k, \quad \langle \Delta n_k^2 \rangle = \tilde{n}_k^2 + \tilde{n}_k. \quad (9)$$

In the limit $T \longrightarrow 0$, all the particles should be in the ground state, so that $\tilde{n}_0 = N$, and we have from (9)

$$\langle \Delta n_0^2 \rangle = N^2 + N,$$

which is meaningless. The authors of paper [3] (see also references therein) have suggested solving this problem still using distribution (6) but calculating the root-mean-square fluctuation of the total number of excited particles

$$N_* = \sum_{k \neq 0} n_k.$$

Then, the exact condition (4) yields

$$\langle \Delta n_0^2 \rangle = \langle \Delta N_*^2 \rangle = \langle N_*^2 \rangle - \langle N_* \rangle^2. \quad (10)$$

Because for $T \longrightarrow 0$, we certainly have $\langle N_*^2 \rangle \longrightarrow 0$ and $\langle N_* \rangle^2 \longrightarrow 0$, and the contradiction related to (9) is removed. However, the result (10) does not correspond to the value of $\langle \Delta n_0^2 \rangle$ obtained directly from (6). This again emphasizes the fact that the correct distribution differs from the distribution (6a), which does not satisfy condition (4).

In this paper, it is shown that the distribution (6a) is valid at temperatures above the condensation temperature (critical temperature). The correct treatment of condition (4) results in the fact that below the condensation temperature ($T < T_c$) the distribution function of excited particles approximately retains the form (6a), whereas the distribution function (6b) of the number of particles in the condensate drastically changes and takes a Gaussian shape. In this case, in particular, the contradictions related to (9) are removed. The statistical dependence of quantities n_k determined by the condition (4) becomes essential at temperatures below the critical temperature, resulting in a change in the function $W_0(n_0)$ and suppression of fluctuations.

Note that, as it follows from the general considerations and is demonstrated in section 4 using the exact solution for a gas trapped in a parabolic potential, at a large number of particles N the distribution function $W_0(n_0)$ changes and the mean value $\tilde{n}_0$ vanishes very rapidly (i.e., in a rather narrow region near certain temperature T_c) but continuously (in the case of $\tilde{n}_0$, without a jump in the derivative as well). Only in the limiting case of $N \longrightarrow \infty$, this change at $T = T_c$ has the character of a jump and $\tilde{n}_0$ vanishes at this point with a nonzero derivative. Only in this case, the value T_c has an exact sense and can be defined as the temperature above which the population of the ground state is equal to zero. For finite N values, there is no possibility of such definition and T_c is a characteristic temperature in a narrow vicinity of which $\tilde{n}_0$ is changing from very large to very small values.

In this paper, the Bose gas confined by a trap, i.e., a system with a discrete spectrum, was considered. The condensation of such a gas is of the most interest from the practical point of view because this phenomenon has been experimentally observed. Generally speaking, the case of a free Bose gas is also very interesting. For some reasons, which will be discussed below, this case requires a special approach.

2. DEGENERATE BOSE GAS

Condition (1a) can be satisfied automatically by representing (1) in the form[1]

$$W_0(n_0) = \frac{1}{S}\exp(-\varepsilon_0 n_0) \sum_{n_1, n_2, \ldots}^{(n_0)} \exp(-\varepsilon_1 n_1 - \varepsilon_2 n_2 - \ldots) \times \frac{1}{2\pi i}\oint z^{(-N + n_0 - 1) + n_1 + n_2 + \ldots} dz, \quad (11)$$

where the sign (n_0) at the sum means that the summation over n_0 is excluded. The integration contour in (11) is a circle with a center at the point $z = 0$. Only when condition (1a) is satisfied, the integrand in (11) has a simple pole and the integral is equal to $2\pi i$. In other cases, the integral is zero, which allows one to perform summation independently (before integrating) over all values $0 \le n_k \le \infty$ in the right-hand side of (11), excluding, of course, the summation over n_0. However, the convergence of all the sums appearing, including the statistical sum S, should be provided. One can easily

[1] A similar method of the sum representation has been used in [4, 5].

see that this will be satisfied if the radius of the circle $|z|$, which can be conveniently written in the form

$$|z| = e^{\mu},$$

will be limited by the condition

$$e^{\mu - \varepsilon_0} < 1.$$

It is convenient to arrange the values of ε_k in the order $\varepsilon_0 < \varepsilon_1 \leq \varepsilon_2 \ldots$ (the ground state is always nondegenerate). Then, we can set $\varepsilon_0 = 0$ (that means $\varepsilon_{k\neq 0} > 0$) and require the fulfillment of the condition $\mu < 0$.

The summation in (11) gives

$$W_0(n_0) = \frac{1}{S}\frac{1}{2\pi i}\oint z^{-N+n_0-1}e^{G(z)}dz,$$

$$e^{G(z)} = \prod_{k\neq 0}(1 - ze^{-\varepsilon_k})^{-1}, \qquad (12)$$

$$G(z) = -\sum_{k\neq 0}\ln(1 - ze^{-\varepsilon_k}) \equiv \sum_{k\neq 0}\sum_{p=1}^{\infty}\frac{1}{p}z^p e^{-p\varepsilon_k}.$$

The function $G(z)$ has no singularities inside the circle with the radius $|z| = e^{\mu} < 1$ ($\mu < 0$), so that the equality $W_0(n_0) = 0$ is automatically satisfied for $n_0 > N$. For $n_0 = N$, we obtain from Eq. (12)

$$W_0(n_0 = N) = \frac{1}{S}e^{G(0)} = \frac{1}{S};$$

and for $n_0 = N - 1$, we have

$$W_0(n_0 = N-1) = \frac{1}{S}\left[\frac{d}{dz}e^{G(z)}\right]_{z=0} = \frac{1}{S}\sum_{k\neq 0}e^{-\varepsilon_k}.$$

At a very low temperature (in fact, it is necessary that the value of $e^{-\varepsilon_1}$ should be exponentially small, i.e., $\varepsilon_1 \gg 1$, which is certainly satisfied for $T \longrightarrow 0$ for a system with the discrete spectrum), the probability $W_0(n_0 = N-1)$ is small compared to the probability $W_0(n_0 = N)$ and rapidly decreases with further decreasing n_0, as one can easily see from Eq. (12). Therefore, we may restrict ourselves to two quantities

$$W_0(n_0 = N) = 1 - \sum_{k\neq 0}e^{-\varepsilon_k},$$

$$W_0(n_0 = N-1) = \sum_{k\neq 0}e^{-\varepsilon_k}, \qquad (13)$$

which yield

$$\langle n_0\rangle = N - \sum_{k\neq 0}e^{-\varepsilon_k}, \quad \langle \Delta n_0^2\rangle = \sum_{k\neq 0}e^{-\varepsilon_k},$$

so that $\langle \Delta n_0^2\rangle \longrightarrow 0$ at $T \longrightarrow 0$.

It is interesting to compare the distribution (13), which takes the form

$$W_0(n_0) = \delta_{n_0, N}$$

at $T \longrightarrow 0$, with the distribution (6b). For $T \longrightarrow 0$, it follows from (8) that

$$e^{-\mu} - 1 = 1/N$$

or

$$\mu \approx -1/N,$$

and the distribution (6b) takes the form

$$W_0(n_0) = \frac{1}{N}e^{-n_0/N};$$

i.e., it has a maximum at $n_0 = 0$, monotonically decreases with increasing n_0, and is nonzero in a broad range of values of n_0, decreasing only by a factor of e at the limiting value $n_0 = N$.

Similarly to (12), the joint distribution can be easily found:

$$W_{0,\, i\neq 0}(n_0, n_i) = \frac{1}{S}e^{-\varepsilon_i n_i}\frac{1}{2\pi i} \times \oint z^{-N-1+n_0+n_i}e^{G(z)}(1 - ze^{-\varepsilon_i})dz. \qquad (14)$$

At $T \longrightarrow 0$, only three quantities are substantial:

$$W_{0,\, i\neq 0}(n_0 = N, n_i = 0) = S^{-1},$$

$$W_{0,\, i\neq 0}(n_0 = N-1, n_i = 1) = S^{-1}e^{-\varepsilon_i},$$

$$W_{0,\, i\neq 0}(n_0 = N-1, n_i = 0) = S^{-1}\sum_{k\neq 0,\, i}e^{-\varepsilon_i}.$$

Finally, we obtain that the distribution

$$W_{i\neq 0}(n_i) = \sum_{n_0}W_{0,\, i\neq 0}(n_0, n_i)$$

of the number of particles in excited states described by the expressions

$$W_{i\neq 0}(n_i = 0) = 1 - e^{-\varepsilon_i},$$

$$W_{i\neq 0}(n_i = 1) = e^{-\varepsilon_i}. \qquad (15)$$

This result virtually coincides with Eq. (6a).

At higher temperatures, when $\varepsilon_1 \leq 1$ and the total number of particles in excited states is comparable with the number of particles in the ground state, it is impossible to obtain the result by such a simple method. In this case, however, one can use the circumstance that when the number of particles confined by a trap is large (this number in real experiments is between 10^3 and 10^8), the condition $\varepsilon_1 \ll 1$ is satisfied at temperatures that are much lower than the condensation temperature,

which allows one to study virtually completely the integrals in expressions (12) and (14).

3. CONDENSATION OF BOSE GAS IN A TRAP

Having made in Eq. (12) the change of variables

$$z = e^{\mu + ix}$$

we obtain

$$W_0(n_0) = \frac{1}{S} e^{\mu n_0} \int_{-\pi}^{\pi} \exp[-i(N-n_0)x + F(x)]dx,$$
$$F(x) = -\sum_{k\neq 0} \ln(1 - \exp[\mu + ix - \varepsilon_k]). \quad (16)$$

We omitted in Eq. (16) the factor

$$(2\pi)^{-1} e^{-\mu N}$$

that is independent of n_0, which affects only the normalization defined by the relation (16) itself. Below, we will follow this rule.

Then, by expanding the function $F(x)$ in a series at the point $x = 0$ retaining the three first terms, we find

$$F(x) = F(0) + ixA - x^2 D,$$
$$F(0) = -\sum_{k\neq 0} \ln(1 - e^{\mu - \varepsilon_k}), \quad A = \sum_{k\neq 0} \tilde{n}_k, \quad (17)$$
$$D = \frac{1}{2}\sum_{k\neq 0} (\tilde{n}_k + \tilde{n}_k^2),$$

where the quantities $\tilde{n}_k$ are defined by expression (7) (recall that $\varepsilon_0 = 0$). The first term in the expansion enters into the normalization after the substitution into Eq. (16) and can be omitted. The values of A and D depend on the parameter μ, which satisfies the condition $\mu < 0$, being arbitrary in other respects. Let us choose this parameter by requiring the fulfillment of the condition

$$N = \tilde{n}_0 + A = \tilde{n}_0 + \sum_{k\neq 0} \tilde{n}_k. \quad (18)$$

Condition (18) coincides with (8), and the parameter μ acquires the meaning of the chemical potential, the quantities $\tilde{n}_k$ not being now, however, the average numbers of particles. Note that μ and $\tilde{n}_0$ are related by the expression

$$\mu = -\ln(1 + 1/\tilde{n}_0),$$

which allows us to treat Eq. (18) as the equation for μ or $\tilde{n}_0$, depending on our choice.

Consider the temperature dependences of quantities A and D. At $T \longrightarrow 0$, we obtain $\varepsilon_{k\neq 0} \longrightarrow \infty$, so that $A = 0$, $D = 0$, $\tilde{n}_0 = N$, and $\mu = -1/N$ with a high accuracy. As temperature increases, the values of $\varepsilon_{k\neq 0}$ decrease, while the values of A and D increase, and when $T > T_*$, where T_* is a characteristic temperature that depends on the type of the trap potential and the number N of confined particles, the situation certainly occurs when A becomes of the order of N and the value of $\tilde{n}_0 = N - A$ still remains large (which means that the temperature T_* is much lower than the critical temperature, $T_* \ll T_c$), while the chemical potential equals $\mu = -1/\tilde{n}_0$; i.e., it is very small. It is important that in this case the value of D also becomes of the order of N, i.e., very large. As the temperature further increases, the values of $\varepsilon_{k\neq 0}$ decrease and condition (18) can be satisfied only at large values of $|\mu|$; i.e., the value of $\tilde{n}_0$ becomes small (the condensate fraction is absent). In this case, $A = N$ with a high accuracy and $D \geq N/2$.

This means that, beginning from temperatures $T > T_*$, which are much lower than the condensation temperature, the real part of the function $F(x)$ has a maximum at $x = 0$ and rapidly decreases in the vicinity of this point. Therefore, we may substitute the expansion (17) into Eq. (16) and perform integration using infinite limits. As a result, we find that the distribution of particles in the condensate is described by the expression

$$W_0(n_0) = S^{-1} \exp\left[\mu n_0 - \frac{(n_0 - \tilde{n}_0)^2}{4D}\right], \quad (19)$$
$$\mu = -\ln(1 + 1/\tilde{n}_0),$$

which has a universal shape because it depends on the trap potential only via parameters $\tilde{n}_0$ and D. All the average values are obtained from this distribution by differentiating the statistical sum

$$S = \sum_{n_0 = 0}^{N} \exp\left[\mu n_0 - \frac{(n_0 - \tilde{n}_0)^2}{4D}\right]$$

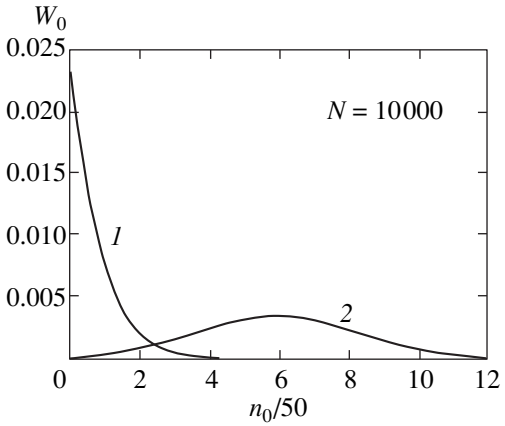
with respect to the explicit parameter μ.

The distribution (19) has qualitatively different forms for large and small values of $\tilde{n}_0$. In a broad temperature range below the condensation temperature, when the conditions

$$\tilde{n}_0^2 \gg D, \quad N - \tilde{n}_0 \gg 1 \quad (20)$$

are satisfied, the distribution function (19) is exponentially small at two its ends at $n_0 = 0$ and $n_0 = N$ [the second inequality is equivalent to the condition $D \gg 1$ and simultaneously provides the validity of Eq. (19)]. Therefore, we may calculate the statistical sum by passing from summation to integration with infinite limits, which gives

$$S = 2\sqrt{\pi D} \exp(\mu\tilde{n}_0 + \mu^2 D), \quad \mu = -1/\tilde{n}_0. \quad (21)$$



Distribution function (19) of the number of particles in the condensate of a Bose gas confined by a trap for different temperatures $T/T_c = 1.01$ (*1*) and 0.99 (*2*). The values of $\tilde{n}_0$ are calculated from expression (29), $D = (1/2)\gamma N(T/T_c)^3 \approx 6850(T/T_c)^3$. The number of particles in the trap $N = 10000$.

In this case, the distribution is described by a Gaussian with the width $\Delta n_0 = \sqrt{4D}$ slightly shifted by $-D/2\tilde{n}_0$ with respect to $\tilde{n}_0$ due to the factor

$$e^{\mu n_0} = e^{-n_0/\tilde{n}_0}.$$

The value of $\tilde{n}_0$ increases with decreasing temperature, while D decreases, and the distribution (19) narrows down,

$$\Delta n_0 = \sqrt{4D} \approx \sqrt{2(N - \tilde{n}_0)},$$

and its center shifts to larger values of n_0. All the average values can be easily calculated:

$$\langle n_0^k \rangle = \frac{1}{S}\frac{\partial^k S}{\partial \mu^k}.$$

The first and second moments are close to $\tilde{n}_0$ and $\tilde{n}_0^2$, respectively:

$$\begin{aligned} \langle n_0 \rangle &= \tilde{n}_0\left(1 - \frac{2D}{\tilde{n}_0^2}\right), \\ \langle n_0^2 \rangle &= \tilde{n}_0^2\left(1 - \frac{2D}{\tilde{n}_0^2} + \frac{4D^2}{\tilde{n}_0^4}\right). \end{aligned} \tag{22}$$

The relative root-mean-square fluctuation

$$\sqrt{\langle \Delta n_0^2 \rangle}/\tilde{n}_0 \approx \frac{\sqrt{2D}}{\tilde{n}_0}$$

is small in accordance with Eq. (20) and decreases with decreasing temperature, i.e., with the increasing number of particles in the condensate. Eventually, at very large temperatures, the condition of validity of distribution (19) (the second condition in Eq. (20)) is no longer fulfilled and the distribution takes the form of Eq. (13). As the temperature increases, the $\tilde{n}_0$ value is decreasing, the first condition in (20) fails to be valid, and the lower boundary $n_0 = 0$ of the distribution becomes more significant. Finally, at $\tilde{n}_0 \ll \sqrt{D}$ (although the condition $\tilde{n}_0 \gg 1$ can still be fulfilled, however, one can assume that the condensate fraction is already absent), the factor $e^{\mu n_0}$ becomes dominant in expression (19). By performing summation, we obtain Eq. (6b). The passage from the Gaussian shape of the distribution (19) to the shape (6b) occurs at the values

$$1 \ll \tilde{n}_0 \leq \sqrt{D} \approx \sqrt{N/2},$$

which, for large N, corresponds to a narrow vicinity of the critical point. The qualitative change in the shape of distribution function (19) in the vicinity of the critical point is shown in the figure.

The joint distribution $W_{0,\, i\neq 0}(n_0, n_i)$ can be obtained similarly. By using Eq. (14), making again the change of variables

$$z = e^{\mu + ix}$$

and performing integration, we find

$$\begin{aligned} W_{0,\, i\neq 0}(n_0, n_i) &= \frac{1}{S}\exp[\mu(n_0 + n_i) - \varepsilon_i n_i] \\ &\times\left\{\exp\left[-\frac{(n_i + n_0 - \tilde{n}_0)^2}{4D}\right]\right. \\ &\left. - \exp(\mu - \varepsilon_i)\exp\left[-\frac{(n_i + n_0 + 1 - \tilde{n}_0)^2}{4D}\right]\right\}. \end{aligned} \tag{23}$$

The distribution of the number of particles in excited states can be readily obtained from (23):

$$W_{i\neq 0}(n_i) = \sum_{n_0 = 0}^{N} W_{0,\, i\neq 0}(n_0, n_i).$$

When the number of particles in the condensate is large and the condition (20) is satisfied, one may replace summation by integration with infinite limits to obtain the distribution

$$W_{i\neq 0}(n_i) = (1 - e^{-\varepsilon_i})e^{-\varepsilon_i n_i}, \tag{24}$$

which coincides with (6a) for $\mu = 0$. Note that, with accuracy to the first-order terms in $\mu = -1/\tilde{n}_0$, the sum of averages corresponding to (24) is

$$\sum_{i\neq 0}\langle n_i \rangle = \sum_{i\neq 0}\tilde{n}_i - 2D/\tilde{n}_0.$$

Finally, we obtain from (22)

$$\sum_i \langle n_i \rangle = N.$$

When $\tilde{n}_0 \ll \sqrt{D}$ (however, the condition $\tilde{n}_0 \gg 1$ can be still valid), the factor $e^{\mu n_0}$ plays the dominant role in summation over $\tilde{n}_0$ in (23), as in (19), and we obtain the distribution (6a).

The average of products can be also obtained from the distribution (23). In particular, when the conditions (20) are satisfied, we obtain

$$\langle n_0 n_{i\neq 0} \rangle = \langle n_0 \rangle \langle n_i \rangle - \langle n_i \rangle^2 - \langle n_i \rangle^2.$$

In the conclusion of this section, we emphasize the following important circumstance. Upon condensation of the gas, the number $\tilde{n}_0$ of particles becomes of the order of N and the chemical potential of the order of $1/N$, i.e., very small. For this reason, it is important that the value of D would remain finite for $\mu \longrightarrow 0$. In the case of a system with the discrete spectrum, this requirement is certainly fulfilled because the largest of the terms $\tilde{n}_{k=1}^2$ in D remains finite for $\mu \longrightarrow 0$ (quantities $\tilde{n}_k$ cannot produce a singularity because they are related by expression (18)]. In the case of a free gas (a system with a continuous spectrum), the situation is different. In this case, $D \longrightarrow \infty$ at $\mu \longrightarrow 0$; i.e., the second derivative of the function $F(x)$ has a singularity at the point $x = 0$, and one can show that the term next to a linear one in the expansion of $F(x)$ is proportional to $|x|^{3/2}$. This, of course, drastically changes the situation and requires a special treatment.

4. PARABOLIC POTENTIAL

In this case, which is of most practical importance, we can calculate exactly the quantities $\tilde{n}_0$ and D, which determine the distribution (19). We will denote the energy levels of particles by introducing the vector index $\mathbf{k} = (k_x, k_y, k_z)$, $k_i = 0, 1, \ldots$ and represent the energy of a three-dimensional oscillator (in temperature units) in the form

$$\varepsilon_{\mathbf{k}} = \frac{\hbar}{T}(\mathbf{k} \cdot \boldsymbol{\omega}),$$

where $\boldsymbol{\omega} = (\omega_x, \omega_y, \omega_z)$ and $\omega_{x,y,z}$ are the frequencies of a parabolic trap. Then, the sum over all $\mathbf{k} \neq 0$ in Eq. (12) can be calculated exactly and we have

$$G(z) = \sum_{p=1}^{\infty} p^{-1} z^p g(\boldsymbol{\beta}, p),$$
$$g(\boldsymbol{\beta}, p) = \left[\prod_{i=x,y,z} (1 - e^{-\beta_i p})^{-1} \right] - 1, \quad \beta_i = \frac{\hbar \omega_i}{T}. \tag{25}$$

By making the change of variables

$$z = e^{\mu + ix}$$

in Eq. (12), we obtain again Eq. (16). The function $F(x)$ now has the form

$$F(x) = \sum_{p=1}^{\infty} p^{-1} e^{(\mu + ix)p} g(\boldsymbol{\beta}, p),$$

and the coefficients A and D in Eq. (17) take the form

$$A = \sum_{p=1}^{\infty} e^{\mu p} g(\boldsymbol{\beta}, p), \quad D = \frac{1}{2} \sum_{p=1}^{\infty} p e^{\mu p} g(\boldsymbol{\beta}, p). \tag{26}$$

The relation (18) also retains its meaning because, as one can see from Eqs. (25) and (26), the relation

$$A = \left(z \frac{dG}{dz} \right)_{z = e^{\mu}} = \sum_{\mathbf{k} \neq 0} (e^{\varepsilon_{\mathbf{k}} - \mu} - 1)^{-1} = \sum_{\mathbf{k} \neq 0} n_{\mathbf{k}}$$

is naturally fulfilled.

For small β (we will see below that this condition becomes valid at temperatures that are much lower than the critical temperature), the main contribution to sums A and D comes from small values of p, and we can write[2]

$$A = \frac{1}{\beta^3} \sum_{p=1}^{\infty} p^{-3} e^{\mu p}, \quad D = \frac{1}{2\beta^3} \sum_{p=1}^{\infty} p^{-2} e^{\mu p},$$
$$\beta = \frac{\hbar \omega}{T}, \quad \omega = (\omega_x \omega_y \omega_z)^{1/3}.$$

It is convenient to rewrite these relations by introducing the critical temperature T_c:

$$A = N t^3 \zeta^{-1}(3) \sum_{p=1}^{\infty} p^{-3} e^{\mu p},$$
$$D = \frac{1}{2} N t^2 \zeta^{-1}(3) \sum_{p=1}^{\infty} p^{-2} e^{\mu p}, \tag{27}$$
$$t = T/T_c, \quad T_c = \zeta^{-1/3}(3) \hbar \omega N^{1/3},$$

where $\zeta(s)$ is the Riemann zeta function. It is convenient to represent the relation (18), which determines μ or $\tilde{n}_0$, in the form

$$\tilde{n}_0/N + t^3 f(\mu) = 1, \quad f(\mu) = \zeta^{-1}(3) \sum_{p=1}^{\infty} p^{-3} e^{\mu p}. \tag{28}$$

[2] To simplify the representation, we neglect here the next-order term β^{-2}, which gives a correction to the critical temperature related to a finite number of particles in the trap and can be easily taken into account if necessary [6, 7].

For small μ, we can use in equation (28) the expansion of the function $f(\mu)$ with accuracy to the first-order terms in μ:

$$f(\mu) = 1 + \gamma\mu, \quad \gamma = \zeta(2)/\zeta(3) \approx 1.37$$

and the value $\tilde{n}_0 = -1/\mu$. As a result, the quadratic equation

$$-\frac{1}{N\mu} + \gamma t^3 \mu = 1 - t^3$$

is obtained [7], which yields the temperature dependences of the chemical potential and $\tilde{n}_0$:

$$\mu = \frac{1}{2\gamma t^3}[1 - t^3 - \sqrt{(1-t^3)^2 + 4\gamma t^3/N}],$$
$$\tilde{n}_0 = \frac{N}{2}[1 - t^3 + \sqrt{(1-t^3)^2 + 4\gamma t^3/N}]. \quad (29)$$

For $N \longrightarrow \infty$, it follows from these relations that at temperatures below the critical temperature ($T < T_c$), when $t < 1$, we have

$$\mu = 0, \quad \tilde{n}_0/N = 1 - t^3, \quad T < T_c, \quad (30)$$

while at temperatures above the critical temperature, but with the condition $t - 1 \ll 1$ still being valid, we find from (29)

$$\mu = -\frac{t^3 - 1}{\gamma t^3}, \quad \tilde{n}_0 = \frac{\gamma t^3}{t^3 - 1}, \quad (31)$$
$$T > T_c, \quad t - 1 \ll 1.$$

One can see from (29) that for finite values of N, the dependence (30) is realized provided two equivalent conditions

$$\tilde{n}_0 \gg \sqrt{N}, \quad 1 - t^3 \gg 1/\sqrt{N} \quad (32)$$

are fulfilled, which coincide in fact with the first of the conditions (20). Condition (32) becomes invalid in a narrow vicinity of the critical temperature, and one should use a more exact relation (29), which allows one to describe correctly a smooth decrease in $\tilde{n}_0$ and an increase in μ with increasing temperature up to the values $T \le T_c + \Delta T$ (where $\Delta T \ll T_c$), when $\tilde{n}_0$ is still large and μ is small [7]. The values of $\tilde{n}_0$ and μ obtained in this way should be used in the distribution (19).

Note now that the parameter $\beta = \hbar\omega/T$ can be rewritten by expressing T in terms of the critical temperature

$$\beta = \zeta^{1/3}(3) N^{-1/3} t^{-1},$$

which shows that the condition $\beta \ll 1$ proves to be fulfilled at large N beginning from temperatures that are much lower than the critical temperature but satisfy, however, the condition $T \gg T_*$, where $T_* = T_c N^{-1/3} \ll T_c$. Therefore, expansion (27) can be used only for rather high temperatures $T \gg T_*$ (it should be recalled that this expansion is valid for $\beta \ll 1$). Because, however, $T_* \ll T_c$, this expansion is valid beginning from temperatures that are much lower than the critical temperature, when $\tilde{n}_0$, as one can see from (30), is already very close to N; i.e., almost all the gas is already in the condensate. In this case, we can calculate $\tilde{n}_0$ and μ in the temperature region

$$T_* \ll T \le T_c + \Delta T, \quad T_* = T_c N^{-1/3} \ll T_c \quad (33)$$

using (29). In the region (33), the condition $\mu \ll 1$ is fulfilled, so that we can set $\mu = 0$ in the calculation of D. As a result, we find

$$D = \frac{1}{2}\gamma t^3 N,$$

and, as one can see from Eq. (33), the condition $D \gg 1$ is satisfied. As temperature further increases, D increases and approaches its limiting value $D = N/2$ at high temperatures. Because of this, beginning from temperatures that satisfy the condition $T \gg T_*$ but are still much lower than the critical temperature, we can use the asymptotic estimate of the integral (16), which leads to the distribution (19).

It should be emphasized that the region (33) where quantities (29) can be calculated contains not only temperatures below the critical temperature but also the vicinity of the critical temperature where the distribution (19) changes from Gaussian to the form (6b). As temperature approaches the critical temperature, the conditions (32) and, correspondingly, (20) are no longer fulfilled in the region,

$$|1 - t| \le 1/\sqrt{N},$$

and distribution (19) is not already Gaussian but still differs from (6b). However, in the region

$$1/\sqrt{N} \ll t - 1 \ll 1,$$

when expressions (29) are still valid, the condition $\tilde{n}_0 \ll \sqrt{D}$ is already fulfilled, and the distribution (19) takes the form (6b). Thus, as temperature increases in the narrow vicinity of the critical point

$$|1 - t| \le 1/\sqrt{N},$$

the distribution (19) qualitatively changes from Gaussian to the form (6b) (see figure). For large values of N, this change in the shape of the distribution function of the number of particles in the condensate occurs in a very narrow vicinity of the critical temperature, i.e., almost at a jump.

5. JUMP OF THE HEAT CAPACITY

The change in distribution (19) from a Gaussian shape, which it has in the region (20), (32), where the pronounced condensate fraction exists, to the form (6b), when the condensate fraction is virtually absent, is accompanied by an abrupt change in the heat capacity, which occurs at a jump at $N \longrightarrow \infty$. In the case of the

parabolic potential, the jump magnitude can be calculated exactly.

When $N \longrightarrow \infty$, the average values $\langle n_{i\neq 0}\rangle$ coincide with $\tilde{n}_{i\neq 0}$. Let us represent $\varepsilon_{\mathbf{k}}$ in the form

$$\varepsilon_{\mathbf{k}} = \beta(\boldsymbol{\Omega}\cdot\mathbf{k}), \quad \Omega_i = \omega_i/\omega.$$

Then, we obtain for the energy E of the gas

$$E = T\sum_{k\neq 0}\varepsilon_k\tilde{n}_k = T\sum_{k\neq 0}\frac{\beta(\boldsymbol{\omega}\cdot\mathbf{k})}{\exp(\beta(\boldsymbol{\omega}\cdot\mathbf{k})-\mu)-1}$$

$$= T\sum_{k\neq 0}\sum_{p=1}^{\infty}\beta(\boldsymbol{\omega}\cdot\mathbf{k})\exp(-(\beta(\boldsymbol{\omega}\cdot\mathbf{k})-\mu)p)$$

$$= T\beta\left(-\frac{\partial}{\partial\beta}\right)\sum_{p=1}^{\infty}\frac{1}{p}e^{\mu p}\left\{\left[\prod_{i=x,y,z}(1-\exp-\beta\Omega_i p)^{-1}\right]-1\right\}.$$

At a very low temperature $T \ll \hbar\omega_i$, this gives

$$E = \hbar\sum_i \omega_i\exp(-\hbar\omega_i/T);$$

i.e., the gas energy is exponentially small. At higher temperatures $T \gg T_*$, using the smallness of the parameter β, we find

$$E = 3N\frac{T^4}{T_c^3}\zeta^{-1}(3)\sum_{p=1}^{\infty}p^{-4}e^{\mu p}.$$

As a result, at $T \gg T_*$, the heat capacity has the form

$$C = \frac{dE}{dT} = 12Nt^3\zeta^{-1}(3)\sum_{p=1}^{\infty}p^{-4}e^{\mu p} + 3Nt^3\left(t\frac{\partial\mu}{\partial t}\right)\zeta^{-1}(3)\sum_{p=1}^{\infty}p^{-3}e^{\mu p}. \quad (34)$$

When $N \longrightarrow \infty$, the chemical potential μ in this expression is zero at $T \le T_c$. For $T \ge T_c$, it is determined by the equality

$$t^3 f(\mu) = t^3\zeta^{-1}(3)\sum_{p=1}^{\infty}p^{-3}e^{\mu p} = 1, \quad (35)$$

which follows from Eq. (28). Therefore, for $T \le T_c$, we obtain for the heat capacity the expression

$$C = 12\zeta^{-1}(3)\zeta(4)Nt^3 \approx 10.8Nt^3.$$

For $T = T_c$, we find from Eq. (35)

$$\frac{\partial\mu}{\partial t} = -3/\gamma \approx -2.19.$$

Finally, the jump in the heat capacity is

$$\Delta C = -9[\zeta(3)/\zeta(2)]N \approx -6.57N.$$

At large negative values of μ (large t), the first term dominates in sum (35), and we obtain

$$e^{\mu} = \zeta(3)t^{-3}, \quad t(\partial\mu/\partial t) = -3.$$

As a result, we obtain at high temperature from Eq. (34) the value of heat capacity $C = 3N$, which coincides with the heat capacity of an ideal gas confined by a trap, which is described by the classical distribution function

$$W(\mathbf{r},\mathbf{v},t) = \left(\frac{m\omega}{2\pi T}\right)^3 N\exp\left(-\frac{m\omega^2 r^2 + mv^2}{2T}\right),$$

where m, $\mathbf{r}$, and $\mathbf{v}$ are the particle mass, coordinate, and velocity, respectively. One can see from the exact expansion of the function $f(\mu)$ [see formula (A.1)] that the function $(\partial\mu/\partial t)$ has an infinite negative slope at the point $t = 1$. Therefore, the heat capacity has the same slope after the jump and, as a result, it rapidly acquires the asymptotic value $C = 3N$.

6. CONCLUSIONS

(1) The general integral representation (12) of the distribution function $W_0(n_0)$ of the number of particles in the condensate of an ideal Bose gas confined by a trap (a system with the discrete spectrum) allows one to study the variation of this function over the entire temperature range. At a very low temperature ($E_1/T \ll 1$), the distribution function has the form (13); i.e., it is close to unity at $n_0 = N$ and rapidly (exponentially) decreases at $n_0 < N$, taking the form

$$W_0(n_0) = \delta_{n_0,N}$$

at $T = 0$. Beginning from sufficiently high temperatures $T > T_*$ [the temperature T_* depends on the form of the trap potential and the number N of particles, however, for large N it is certainly much lower than the critical temperature ($T_* \ll T_c$)], the distribution function $W_0(n_0)$ takes the universal form (19): only the parameter n_0 and D appearing in this function depend on the trap potential. In a broad temperature range below the critical temperature, $T > T_*$, $T_c - T \gg T_c/\sqrt{N}$, distribution (19) is described by a Gaussian whose width and, hence, root-mean-square fluctuations tend to zero with decreasing temperature. As the temperature approaches the critical temperature, the width of distribution (19) increases; i.e., fluctuations are enhanced. In a narrow vicinity $|T - T_c| \le T_c/\sqrt{N}$ of the critical temperature, the distribution function (19) changes completely and takes the form (6b) at $T - T_c \gg T_c/\sqrt{N}$. In the limit $N \longrightarrow \infty$, this change occurs at a jump.

(2) In the case of a trap with a parabolic potential, the parameters $\tilde{n}_0$ and D, which determine the univer-

sal distribution (19), can be calculated exactly. In the entire temperature region of the condensate existence,

$$D = \frac{\zeta(2)}{2\zeta(3)} t^3 N,$$

and $\tilde{n}_0$ is calculated from expression (29).

(3) As follows from calculations [7], for typical experimental parameters of the trap, gas can be treated as ideal up to the values $N \leq 10^4$. This value of N is sufficiently high, so that the features of the condensation process considered above can be manifested. However, unlike the case of a free gas (confined by the vessel walls only), the gas confined by the trap is certainly no longer ideal when the number of particles is sufficiently large. The study of the influence of the interaction between gas particles on the features of the change in the distribution function $W_0(n_0)$ during gas condensation is one of the most important problems of the theory.

APPENDIX

Consider the equality

$$\int_1^\infty (e^{xz} - 1)^{-1} dz = \sum_{p=1}^\infty \int_1^\infty e^{-pxz} dz = \frac{1}{x} \sum_{p=1}^\infty p^{-1} e^{-px},$$

from which it follows that

$$\sum_{p=1}^\infty p^{-1} e^{-xp} = x \int_1^\infty (e^{xz} - 1)^{-1} dz = \int_x^\infty (e^z - 1)^{-1} dz.$$

The last integral can be rewritten in the form

$$\int_x^\infty (e^z - 1)^{-1} dz = \int_x^1 [(e^z - 1)^{-1} - z^{-1} + z^{-1}] dz$$

$$+ \int_1^\infty (e^z - 1)^{-1} dz = -\ln x - \int_0^x [(e^z - 1)^{-1} - z^{-1}] dz$$

$$+ \int_0^1 [(e^z - 1)^{-1} - z^{-1}] dz + \int_1^\infty (e^z - 1)^{-1} dz.$$

A sum of two last integrals is zero, and we find

$$f(\mu) = \zeta^{-1}(3) \sum_{p=1}^\infty p^{-3} e^{p\mu}$$

$$= 1 + \gamma\mu - \frac{\mu^2}{2\zeta(3)} [\ln(-\mu) - 3/2] \qquad \text{(A.1)}$$

$$- \zeta^{-1}(3) \int_0^{-\mu} dx_1 \int_0^{x_1} dx_2 \int_0^{x_2} dx_3 \left[\frac{1}{e^{x_3} - 1} - \frac{1}{x_3} \right].$$

ACKNOWLEDGMENTS

The author thanks D.D. Krylova for useful discussions. This work was partially supported by the State Science and Technology Program "Metrology."

REFERENCES

1. M. H. Anderson, J. R. Ensher, M. R. Matthews, *et al.*, Science **269**, 198 (1995); C. C. Bradley, C. A. Sackett, J. J. Tolett, *et al.*, Phys. Rev. Lett. **75**, 1687 (1995); K. B. Davis, M. O. Mewes, M. R. Andrews, *et al.*, Phys. Rev. Lett. **75**, 3969 (1995).
2. L. D. Landau, Zh. Éksp. Teor. Fiz. **7**, 19 (1937); Phys. Z. Sowjetunion **11**, 26 (1937).
3. M. Holthaus, E. Kalinowski, and K. Kirsten, cond-mat/9804171.
4. M. A. Leontovich, *An Introduction to the Thermodynamics. Statistical Physics* (Nauka, Moscow, 1983).
5. I. Fujiwara, D. ter Haar, and N. J. Wergeland, J. Stat. Phys. **2**, 329 (1970).
6. W. Ketterle and N. J. van Druten, Phys. Rev. A **54**, 656 (1996).
7. V. A. Alekseev and D. D. Krylova, Kvantovaya Élektron. (Moscow) **30**, 441 (2000).

Translated by M. Sapozhnikov

Journal of Experimental and Theoretical Physics, Vol. 92, No. 4, 2001, pp. 617–625.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 4, 2001, pp. 710–718.

PLASMA, GASES

On an Anomaly of the Thermal Filamentation Mechanism in a High-Power Electromagnetic Field of a Strongly Collisional Plasma

V. P. Silin

Lebedev Physical Institute, Russian Academy of Sciences, Leninskiĭ pr. 53, Moscow, 117924 Russia
e-mail: silin@sci.lebedev.ru

Received October 10, 2000

Abstract—It is shown that, under the conditions when the mean free path of electrons in a fully ionized plasma is small compared with the London plasma length, the thermal mechanism of inverse bremsstrahlung absorption and electronic conductivity has a stabilizing action on filamentation instability of high-power electromagnetic radiation, in contrast to the known case of comparatively low-intensity radiation. This new nonlinear property of a plasma is shown to be determined by a decrease in the effectiveness of plasma heating in a heating pumping field with increasing the intensity of radiation if electron oscillations have above-thermal rates.

1. The filamentation of electromagnetic radiation [1, 2] in a fully ionized plasma corresponds to the model of a strongly collisional plasma when the characteristic filament size is much larger than the mean free path of electrons. The physical nature of filamentation instability under these conditions corresponds to the thermal mechanism and is caused by two competing thermal processes. One of these is heating of plasma electrons as a result of inverse bremsstrahlung absorption. The second mechanism is related to the removal of heat, which was acquired by the plasma through absorption of radiation, from a filament. In view of this, temperature perturbation in a filament is directly proportional to the increase in heat caused by filamentation disturbance of the electromagnetic field and inversely proportional to the square of the wave vector of the filament and electronic heat conductivity characterizing heat transfer from the filament. Under stationary filamentation conditions, when the pressure of a plasma is almost constant, temperature increase caused by an increase of the electromagnetic field of filamentation perturbation decreases the electron density, which is similar to the ponderomotive force influence. This force virtually fully determines the development of short-wave perturbations characteristic of collisionless plasma conditions. Note that both ponderomotive and thermal mechanisms usually favor the appearance of filamentation. The thermal filamentation mechanism plays the determining role in long-wave perturbations when the size of a filament is larger than the mean free path of electrons. In contrast, the ponderomotive mechanism prevails in short-wave filamentation perturbations.

We show in this communication that such a generally accepted picture of radiation filamentation in a fully ionized plasma is not observed in a comparatively strong pumping field. Under strongly collisional conditions, the thermal mechanism based on inverse bremsstrahlung absorption and electronic heat conductivity not only does not favor radiation filamentation but plays the role of the stabilizing factor, the possibility of overcoming which under the action of ponderomotive radiation determines the filamentation threshold for a strongly collisional plasma in a high-power electromagnetic radiation field. This is a new nonlinear effect corresponding to the conditions of the action of high-power radiation on a plasma [3].

2. Consider a plasma in a strong high-frequency field

$$\mathscr{E} = \frac{1}{2}[\mathbf{E}(\mathbf{r}, t)\exp\{-i(\omega_0 t - \mathbf{k}_0\mathbf{r})\} + \mathbf{E}^*(\mathbf{r}, t)\exp\{i(\omega_0 t - \mathbf{k}_0\mathbf{r})\}]. \quad (2.1)$$

The transfer in such a plasma will be described by the equations for momenta from [4]. As the $\mathbf{E}(\mathbf{r}, t)$ amplitude in Eq. (2.1) slowly varies during high-frequency pumping field period $2\pi/\omega_0$, we can, according to [4], write for the electronic component

$$\frac{\partial n_e}{\partial t} + \mathrm{div}(n_e\mathbf{u}^e) = 0, \quad (2.2)$$

$$m_e n_e\left(\frac{\partial \mathbf{u}^e}{\partial t} + \left(\mathbf{u}^e\frac{\partial}{\partial \mathbf{r}}\right)\mathbf{u}^e\right) + \frac{\partial(n_e k_B T_e)}{\partial \mathbf{r}} + n_e\frac{\partial}{\partial \mathbf{r}}\frac{e^2|\mathbf{E}|^2}{4m_e\omega_0^2} = \mathbf{R}^{ei}, \quad (2.3)$$

$$\frac{\partial T_e}{\partial t} + \mathbf{u}^e\frac{\partial T_e}{\partial \mathbf{r}} + \frac{2}{3}T_e\mathrm{div}\mathbf{u}^e + \frac{2}{3n_ek_B}\mathrm{div}\mathbf{q}_e = \frac{2}{3n_ek_B}Q_{ei}. \quad (2.4)$$

Here, e is the charge of the electron, m_e is the mass of the electron, and k_B is the Boltzmann constant.

Note that the last term in the left-hand side of Eq. (2.3), which is the Miller force now often called the ponderomotive force, is to a substantial extent determined by the magnetic part of the Lorentz force. Consider the action of a strong pumping field on the plasma when the rate of nonrelativistic electron oscillations in the electric field,

$$\mathbf{u}_E(t) = \frac{ei}{2m_e\omega_0}\{\mathbf{E}\exp\{-i(\omega_0 t - \mathbf{k}_0\mathbf{r})\} - \mathbf{E}^*\exp\{i(\omega_0 t - \mathbf{k}_0\mathbf{r})\}\}, \quad (2.5)$$

far exceeds the speed of electron thermal motions. According to [4], the electron-ion friction force in this limit is given by

$$(\mathbf{R}^{ei})_k = m_e n_e \nu_{kj}(u_j^e - u_j^i), \quad (2.6)$$

$$\nu_{kj} = \frac{4\pi e^2 e_i^2 n_i \Lambda}{m_e^2} \times\left(\left\langle\frac{1}{u_E^3(t)}\right\rangle\delta_{kj} - 3\left\langle\frac{u_{Ek}(t)u_{Ej}(t)}{u_E^5(t)}\right\rangle\right). \quad (2.7)$$

Accordingly, the heat released in collisions between electrons and ions is

$$Q^{ei} = \langle Q_{IB}\rangle + \frac{1}{2}\mathbf{R}^{ei}(\mathbf{u}^e - \mathbf{u}^i). \quad (2.8)$$

Here, the heat absorbed by electrons as a result of the inverse bremsstrahlung effect is given by

$$\langle Q_{IB}\rangle = \frac{4\pi e^2 e_i^2 n_i n_e \Lambda}{m_e}\left\langle\frac{1}{u_E(t)}\right\rangle. \quad (2.9)$$

Angle brackets denote averaging over the period of the high-frequency pumping field.

Lastly, the density of the electron heat flow is given by the equation [4]

$$\mathbf{q}_e = \chi\left\{\frac{\partial T_e}{\partial \mathbf{r}} - \frac{6\mathbf{R}^{ei}}{5k_B n_e}\right\}. \quad (2.10)$$

Here, for the electron heat conductivity, we have

$$\chi = Cn_e k_B V_T l_{ee}, \quad (2.11)$$

where $V_T = \sqrt{k_B T_e/m_e}$ is the thermal speed of electrons. The C constant equals (75/32) in the Sonin–Laguerre one-polynomial approximation of the Hilbert–Chapman–Enskog method and (375/128) in the approximation of two polynomials, and

$$l_{ee} = m_e^2 V_T^4/\sqrt{\pi}e^4 n_e\Lambda = V_T/\nu_{ee} \quad (2.12)$$

is the mean free path of a thermal electron between collisions with the other electrons. Note that, in a strong pumping field, the electron-ion collision frequency substantially decreases compared with the frequency given by the usual formula

$$\nu_{ei}(V_T) = 4\sqrt{2\pi}e^2 e_i^2 n_i\Lambda/3m_e^2 V_T^3,$$

whereas a high-frequency field has no effect on the electron–electron collision frequency (2.12), because the difference of the speeds of colliding electrons only depends on their thermal motion in the dipole approximation.

The system of equations for the ionic component is simpler:

$$\frac{\partial n_i}{\partial t} + \mathrm{div}(n_i\mathbf{u}^i) = 0, \quad (2.13)$$

$$m_i n_i\left(\frac{\partial \mathbf{u}^i}{\partial t} + \left(\mathbf{u}^i\frac{\partial}{\partial \mathbf{r}}\right)\mathbf{u}^i\right) + \frac{\partial(n_i k_B T_i)}{\partial \mathbf{r}} = -\mathbf{R}^{ei}, \quad (2.14)$$

$$\frac{\partial T_i}{\partial t} + \mathbf{u}^i\frac{\partial T_i}{\partial \mathbf{r}} + \frac{2}{3}T_i\mathrm{div}\mathbf{u}^i = 0. \quad (2.15)$$

In both (2.4) and (2.15), a weak (on the order of the mass ratio between the electron and the ion) effect of energy transfer from electrons to ions in their collisions is ignored. In a strong pumping field, this effect is additionally suppressed [5]. A comparatively weak effect of ionic heat transfer is also ignored in (2.15).

3. Radiation filamentation will be described in terms of the stationary statement of the problem, which considers strengthening of filamentation perturbations in space in the direction of pumping field propagation in a spatially uniform plasma. Accordingly, consider ground state perturbation

$$n_e = n_{e0} + \delta n_e(\mathbf{r}), \quad \mathbf{u}^e = \delta\mathbf{u}^e(\mathbf{r}), \quad T_e = T_{e0} + \delta T_e(\mathbf{r}),$$
$$n_i = n_{i0} + \delta n_i(\mathbf{r}), \quad \mathbf{u}^i = \delta\mathbf{u}^i(\mathbf{r}), \quad T_i = T_{i0} + \delta T_i(\mathbf{r}),$$
$$\langle Q_{IB}\rangle = \langle Q_{IB}^0\rangle + \delta Q_{IB}(\mathbf{r}). \quad (3.1)$$

In addition to Eqs. (2.2)–(2.4) and (2.13)–(2.15), we will take into account the condition of electrical neutrality

$$en_{e0} + e_i n_{i0} = 0, \quad e\delta n_e + e_i\delta n_i = 0. \quad (3.2)$$

We restrict our consideration to the simplest case, when the electron temperature far exceeds the temperature of ions. More exactly, we assume that

$$e_i T_{e0}/|e|T_{i0} \gg 1. \quad (3.3)$$

This is a natural assumption for a plasma heated by high-frequency radiation. Equations (2.2)–(2.4) and (2.13)–(2.15) combined with the electrical neutrality conditions then yield

$$\delta n_e = n_{e0}\left\{-\frac{\delta T_e}{T_{e0}} - \frac{e^2\delta|\mathbf{E}|^2}{4m_e k_B T_{e0}\omega_0^2}\right\}, \tag{3.4}$$

$$-\chi\Delta\delta T_e = \langle\delta Q_{IB}\rangle, \tag{3.5}$$

where

$$\langle\delta Q_{IB}\rangle = \frac{4\pi e^2 e_i^2 n_i n_e\Lambda}{m_e}\left\langle\delta\frac{1}{u_E(t)}\right\rangle. \tag{3.6}$$

Taking into account electromagnetic field filamentation perturbation, we can write

$$\begin{aligned}\mathbf{E}(\mathbf{r}, t) &= \mathbf{E}_0 e^{i\mathbf{k}_0\mathbf{r}}\\ &+ [\delta\mathbf{E}_+(\mathbf{r})e^{i\mathbf{kr}} + \delta\mathbf{E}_-(\mathbf{r})e^{-i\mathbf{kr}}]e^{i\mathbf{k}_0\mathbf{r}}.\end{aligned} \tag{3.7}$$

The wave vector is normal to the direction of pumping field propagation, $(\mathbf{k}_0\mathbf{k}) = 0$. Equation (3.7) allows us to write

$$\begin{aligned}\delta|\mathbf{E}|^2 &= \mathbf{E}_0[\delta\mathbf{E}_+^*(\mathbf{r})e^{-i\mathbf{kr}} + \delta\mathbf{E}_-^*(\mathbf{r})e^{i\mathbf{kr}}]\\ &+ \mathbf{E}_0^*[\delta\mathbf{E}_+^*(\mathbf{r})e^{i\mathbf{kr}} + \delta\mathbf{E}_-^*(\mathbf{r})e^{-i\mathbf{kr}}],\end{aligned} \tag{3.8}$$

$$\begin{aligned}\left\langle\delta\frac{1}{u_E(t)}\right\rangle &= -\frac{2m_e\omega_0}{|e|}\\ &\times\langle[2|\mathbf{E}_0|^2 - \mathbf{E}_0^2\exp\{-2i(\omega_0 t - \mathbf{k}_0\mathbf{r})\}\\ &- \mathbf{E}_0^{*2}\exp\{2i(\omega_0 t - \mathbf{k}_0\mathbf{r})\}]^{-3/2}\end{aligned} \tag{3.9}$$

$$\begin{aligned}&\times(\mathbf{E}_0[\delta\mathbf{E}_+^*(\mathbf{r})e^{-i\mathbf{kr}} + \delta\mathbf{E}_-^*(\mathbf{r})e^{i\mathbf{kr}}]\\ &+ \mathbf{E}_0^*[\delta\mathbf{E}_+(\mathbf{r})e^{i\mathbf{kr}} + \delta\mathbf{E}_-(\mathbf{r})e^{-i\mathbf{kr}}]\\ &- \mathbf{E}_0[\delta\mathbf{E}_+(\mathbf{r})e^{i\mathbf{kr}} + \delta\mathbf{E}_-(\mathbf{r})e^{-i\mathbf{kr}}]\exp\{-2i(\omega_0 t - \mathbf{k}_0\mathbf{r})\}\\ &- \mathbf{E}_0^*[\delta\mathbf{E}_+^*(\mathbf{r})e^{-i\mathbf{kr}} + \delta\mathbf{E}_-^*(\mathbf{r})e^{i\mathbf{kr}}]\exp\{2i(\omega_0 t - \mathbf{k}_0\mathbf{r})\}\rangle.\end{aligned}$$

In conformity with Eqs. (3.5)–(3.9), we have

$$\delta T_e = \langle\delta Q_{IB}\rangle/\chi k^2, \tag{3.10}$$

$$\delta n_e = n_{e0}\left\{-\frac{\langle\delta Q_{IB}\rangle}{\chi k^2 T_{e0}} - \frac{e^2\delta|\mathbf{E}|^2}{4m_e k_B T_{e0}\omega_0^2}\right\}. \tag{3.11}$$

In writing the equations describing spatial amplification of electromagnetic field perturbations, we will take into account that, according to Eqs. (3.8)–(3.11),

$$\delta n_e = \delta n_+ e^{i\mathbf{kr}} + \delta n_- e^{-i\mathbf{kr}}. \tag{3.12}$$

Here, $\delta n_- = \delta n_+^*$. According to Eq. (3.12), reduced perturbation field equations have the form

$$\left(2i\mathbf{k}_0\frac{\partial}{\partial\mathbf{r}} - k^2\right)\delta\mathbf{E}_{+(-)} = \frac{\omega_{Le}^2\delta n_{+(-)}}{n_{e0}}\mathbf{E}_0, \tag{3.13}$$

where $\omega_{Le} = \sqrt{4\pi e^2 n_{e0}/m_e}$ is the electronic Langmuir frequency.

4. To most clearly describe the special features of a strongly collisional plasma, which is responsible for parametric instability of thermal nonlinearity, in a high-power radiation field, consider the simplest case of circular pumping polarization. A general treatment of elliptical polarization is given in Appendix 1, and the particular case of linear polarization important for practical applications requires a special consideration, see Appendix 2.

We assume that

$$\mathbf{k}_0 = (0, 0, k_0), \quad \mathscr{E}_0 = (\mathscr{E}_{0x}, \mathscr{E}_{0y}, 0).$$

In addition,

$$\begin{aligned}\mathscr{E}_{0x} &= \frac{E_0}{\sqrt{2}}\sin(\omega_0 t - k_0 z),\\ \mathscr{E}_{0y} &= \frac{E_0}{\sqrt{2}}\cos(\omega_0 t - k_0 z).\end{aligned} \tag{4.1}$$

Accordingly, $E_{0x} = iE_0/\sqrt{2}$, $E_{0y} = E_0/\sqrt{2}$, and

$$\begin{aligned}\delta|\mathbf{E}|^2 &= \frac{E_0}{\sqrt{2}}([-i\delta E_{+x} + \delta E_{+y} + i\delta E_{-x}^* + \delta E_{-y}^*]e^{i\mathbf{kr}}\\ &+ [i\delta E_{+x}^* + \delta E_{+x}^* - i\delta E_{-x} + \delta E_{-y}]e^{-i\mathbf{kr}}),\end{aligned} \tag{4.2}$$

$$\left\langle\delta\frac{1}{u_E(t)}\right\rangle = -\frac{m_e\omega_0\delta|\mathbf{E}|^2}{\sqrt{2}|e|E_0^3}. \tag{4.3}$$

These equations allow Eq. (3.6) to be rewritten as

$$\delta Q_{IB} = -\frac{e^2 n_{e0}}{4m_e\omega_0^2}\nu(E)\delta|\mathbf{E}|^2, \tag{4.4}$$

where

$$\nu(E) = \frac{8\sqrt{2}\pi e^2 e_i^2 n_{i0}\Lambda}{m_e^2 V_E^3} = \frac{6\sqrt{\pi}V_T^3}{V_E^3}\nu_{ei}(V_T). \tag{4.5}$$

Here,

$$V_E = |e|E_0/m_e\omega_0. \tag{4.6}$$

V_E is the amplitude of the rate of electron oscillations in the pumping field. In the strong field limit, the condition

$$V_E \gg V_T \tag{4.7}$$

is satisfied. According to (2.9), the absorbed heat in a circularly polarized field is given by

$$Q_{IB}^{(0)} = \frac{e^2 n_{e0}}{m_e \omega_0^2} \frac{1}{2} \nu(E) E_0^2, \tag{4.8}$$

which allows (4.4) to be written in the form

$$\langle \delta Q_{IB} \rangle = \frac{\partial \langle Q_{IB}^{(0)} \rangle}{\partial (E_0^2)} \delta |\mathbf{E}|^2. \tag{4.9}$$

Using these equations makes it possible to rewrite (3.11) as

$$\frac{\delta n_e}{n_{e0}} = -\frac{\delta |\mathbf{E}|^2}{4E_0^2} P, \tag{4.10}$$

where

$$\begin{aligned} P &= \frac{V_E^2}{V_T^2}\left\{1 - \frac{\nu_{ee}(V_T)\nu(E)}{Ck^2 V_T^2}\right\} = \frac{V_E^2}{V_T^2} - \frac{2\langle Q_{IB}^{(0)} \rangle}{T_{e0} k^2 \chi} \\ &= \frac{V_E^2}{V_T^2} + \frac{4}{T_{e0} k^2 \chi} \frac{\partial \langle Q_{IB}^{(0)} \rangle}{\partial \ln (E_0^2)}. \end{aligned} \tag{4.11}$$

Next, Eqs. (3.13), which describe spatial changes in the filamentation perturbation field, will be used to obtain

$$\begin{aligned} &\left(2ik_0 \frac{d}{dz} - k^2\right)(-i\delta E_{+x} + \delta E_{+y}) \\ &= \left(-2ik_0 \frac{d}{dz} - k^2\right)(i\delta E_{-x}^* + \delta E_{-y}^*) \\ &= -\frac{\omega_{Le}^2}{4c^2} P[-i\delta E_{+x} + \delta E_{+y} + i\delta E_{-x}^* + \delta E_{-y}^*]. \end{aligned} \tag{4.12}$$

It follows from (4.12) that field perturbations change by the $\exp(Gz)$ exponential law, where the spatial amplification coefficient G (filamentation instability increment) is given by

$$\begin{aligned} G^2(k) &= \frac{1}{4k_0^2} \\ &\times \left[-k^4 + \frac{\omega_{Le}^2 V_E^2}{2c^2 V_T^2}\left\{k^2 - \frac{\nu_{ee}(V_T)\nu(E)}{CV_T^2}\right\}\right]. \end{aligned} \tag{4.13}$$

Perturbation field amplification characteristic of filamentation instability occurs when the right-hand side of (4.13) is positive and when a $G(k) > 0$ solution is possible. Assuming this condition to be met, we easily find that the spatial amplification coefficient $G(k)$ reaches a maximum at

$$k^2 = k_{\max}^2 = V_E^2 \omega_{Le}^2 / 4V_T^2 c^2. \tag{4.14}$$

Accordingly, the equation

$$\begin{aligned} G_{\max}^2 &= G^2(k_{\max}) \\ &= \frac{\omega_{Le}^2 V_E^4}{64c^2 k_0^2 V_T^4}\left\{\frac{\omega_{Le}^2}{c^2} - \frac{16\nu_{ee}(V_T)\nu(E)}{CV_E^2}\right\}, \end{aligned} \tag{4.15}$$

determines the maximum amplification coefficient value.

Whereas the wave vector (4.12) corresponding to a maximum spatial increment does not depend on dissipation, the increment of spatial filament amplification itself is determined by dissipation. Consider the contribution of thermal dissipation described by (4.13) and (4.15). In a weak pumping field, when the condition inverse to (4.7) is satisfied, heat effects favor filamentation instability. If, however, condition (4.7) is met, we have the opposite situation. To make this clear, consider the expression in braces in the right-hand side of (3.11). The second term there describes the contribution of the ponderomotive force. According to (4.4), this force as usual decreases the electron density as the intensity of radiation $\delta|\mathbf{E}|^2$ increases. In a weak field, the thermal action of radiation on a plasma also decreases the electron density. Indeed, at an almost constant pressure, an increase in the intensity of radiation increases the electron temperature, and this results in a decrease in the electron density. It follows from (2.9) that, under high-power heating radiation conditions, an increase in the intensity of radiation decreases the amount of heat absorbed by electrons. For this reason, radiation intensity growth causes cooling of electrons. At a constant pressure, this results in an increase in the electron density, which is responsible for the negative sign of the second term in braces in (4.11) and, accordingly, the negative sign of a similar term in (4.13). To summarize, we revealed the reason hindering filamentation in the thermal mechanism of inverse bremsstrahlung heating. In the high-power radiation limit, this mechanism can prevent radiation self-focusing in a plasma. It, however, unambiguously follows from (4.13) and (4.15) that the influence of such a mechanism hindering filamentation decreases as the intensity of pumping increases. This follows from the equation

$$\left(\frac{V_E}{V_T}\right)_{\text{th}}^5 = \frac{96\sqrt{\pi}}{C} \frac{c^2}{\omega_{Le}^2 l_{ee}(V_T) l_{ei}(V_T)}, \tag{4.16}$$

which determines the intensity threshold of high-power heating radiation above which filamentation is possible. Equation (4.16) corresponds to vanishing of the right-hand side of (4.15). Under these conditions,

$$l_{ei}(V_T) = \frac{V_T}{\nu_{ei}(V_T)} = \frac{3m_e^2 V_T^4}{4\sqrt{2\pi} e^2 e_i^2 n_i \Lambda} \tag{4.17}$$

is the mean free path of a thermal electron between its collisions with ions.

Clearly, threshold condition (4.16) corresponds to the strong pumping field condition when

$$\lambda_L \approx c/\omega_{Le} \gg \sqrt{l_{ee}(V_T)l_{ei}(V_T)} \approx l_{ee}(V_T); \qquad (4.18)$$

that is, when the London depth λ_L of the high-frequency skin layer of the plasma substantially exceeds the mean free path of electrons with respect to electron-electron and electron-ion collisions.

The wave vector (4.14) value at the threshold pumping field intensity given by Eq. (4.16) corresponds to the size of a filament which appears slightly above the threshold. The mean free path is then given by

$$l_{ee}(V_T)(k_{\max})_{\text{th}} \sim \left(\frac{l_{ee}}{\lambda_L}\right)^{3/5}\left(\frac{l_{ee}(V_T)}{l_{ei}(V_T)}\right)^{1/5}. \qquad (4.19)$$

The smallness of this value compared with unity, which corresponds to the feasibility of applying our strongly collisional description, is again determined by inequality (4.18).

5. In the preceding section, we characterized the peculiar influence of heat effects on the nonlinear electromagnetic field dependence of the electron density, which determines parametric instabilities, for circular polarization. The same will be done for the general case of elliptical polarization in Appendix 1 and for the particular case of linear polarization (an important case which requires a special consideration) in Appendix 2.

Let us analyze the conditions under which our analysis is applicable. First, consider inequality (4.18). For the mean free path of electrons to be small compared with the London length, the inequality

$$n_e \gg 3\times 10^{15}\Lambda^{-2}T_e^4, \qquad (5.1)$$

should be satisfied. Here, T_e [eV] is the electron temperature, and n_e [cm^{-3}] is the number of electrons in one cubic centimeter. The condition of a comparatively high plasma density and a comparatively low temperature implied by Eq. (5.1) should not conflict with the condition of plasma quasi-ideality. The corresponding condition of weakness of interelectronic interactions is fulfilled when the Coulomb logarithm is large,

$$\Lambda = \ln(r_D/r_{\min}) \gg 1, \qquad (5.2)$$

where r_D and $r_{\min}$ are the maximum and minimum electron target parameters for weak momentum transfer [6]. The inequality

$$r_D = \sqrt{\frac{k_B T_e}{4\pi e^2 n_e}} \gg \frac{e^2}{k_B T_e} = r_{\min} \qquad (5.3)$$

can be rewritten as

$$3\times 10^{19}T_e^3 \gg n. \qquad (5.4)$$

Two inequalities (5.1) and (5.4) for the number density of electrons are satisfied simultaneously if $\Lambda^2\times 10^4 \gg T_e$.

To summarize the preceding, note that our treatment of the filamentation phenomenon based on an anomalous plasma kinetics from [4] has led us to theoretically establish a new thermally controlled effect of suppression of electromagnetic radiation filamentation. In contrast to the weak field limit, in which an increase in radiation intensity increases heating, in a high-power radiation field, when the rate of electron oscillations in the pumping field exceeds its thermal speed, an increase in the intensity of radiation causes a decrease in the electron temperature. This effect can most simply be demonstrated by writing the spatial amplification coefficient (increment) of the filament perturbation field in the form

$$G = \frac{1}{2k_0} \times\left\{-k^4 + \frac{\omega_{Le}^2}{2c^2}\left[\frac{k^2V_E^2}{V_T^2} + \frac{4}{\chi T_e}\frac{\partial\langle Q_{IB}^{(0)}\rangle}{\partial(\ln V_E^2)}\right]\right\}^{1/2}. \qquad (5.5)$$

This equation unambiguously follows from Eqs. (4.4) and (4.8)–(4.11) for circularly polarized radiation filamentation, from Eqs.(A.1.14), (A.1.18), (A.1.21), and (A.1.22) for the general case of elliptical polarization if thermal motion is ignored, and, lastly, from Eq. (A.2.13) for linearly polarized radiation filamentation. In the usual weak field limit, when $\langle Q_{IB}^{(0)}\rangle \approx E_0^2$, we have

$$\frac{\partial\langle Q_{IB}^{(0)}\rangle}{\partial(\ln V_E^2)} = \langle Q_{IB}^{(0)}\rangle > 0. \qquad (5.6)$$

In contrast, in a strong field, when the relation $Q_{IB}^{(0)} \approx E_0^{-1}$ can be used, we obtain

$$\frac{\partial Q_{IB}^{(0)}}{\partial(\ln V_E^2)} \approx -\frac{1}{2}Q_{IB}^{(0)}. \qquad (5.7)$$

Equations (5.5)–(5.7) serve to qualitatively distinguish between strong and weak pumping field conditions. The corresponding difference arises because of the dependence of radiation absorbed by a plasma on radiation intensity, which is essential to our consideration.

At the same time, the total decrease in the heating effect caused by an increase in the intensity of high-power radiation decreases competition between the thermal and ponderomotive nonlinearity mechanisms. This makes filamentation possible after surpassing threshold (4.16). It follows that the anomaly of the thermal mechanism suppressing radiation filamentation can be observed in a fairly wide high-power pumping field intensity range ($V_T < V_E < V_{E,\text{th}}$) if condition (5.1) is satisfied; that is, if the mean free path of electrons is substantially smaller than the London length. The

region of such intensities in practical applications is determined by the condition

$$q_T \equiv 2.7 \times 10^{12} \frac{T}{\lambda_\mu^2} \le q_r \le q_T Z^{2/5} \xi, \tag{5.8}$$

where q_r [W/cm^2] is the density of the pumping radiation energy flux, λ_μ is the pumping radiation wave length in micron units, and

$$\xi = \left(4\Lambda^2 \frac{n_e}{T_e^4} \times 10^{-14}\right)^{2/5}. \tag{5.9}$$

For instance, if $n_e = 10^{21}$ cm^{-3} and $T_e = 30$ eV, then $\xi \approx 16.5$. Even if $Z \sim 1$, this ensures meeting condition (5.8). Simultaneously, conditions (5.1) and (5.4) are also satisfied.

To summarize, we established the possibility of anomalous parametric plasma nonlinearity manifestations resulting from suppression of filamentation instability. The phenomenon is explained by a decrease in the heat absorbed by a plasma as a result of the inverse bremsstrahlung effect as the intensity of high-power pumping field radiation grows.

ACKNOWLEDGMENTS

This work was financially supported by the Russian Foundation for Basic Research (project no. 99-02-18075) and the International Scientific-Engineering Center (grant 1253).

APPENDIX 1

Let the pumping field be elliptically polarized, when $\mathbf{k}_0 = (0, 0, k_0)$, $\mathscr{E}_0 = (\mathscr{E}_{0x}, \mathscr{E}_{0y}, 0)$, and

$$\begin{aligned} \mathscr{E}_{0x} &= \varepsilon_x E_0 \cos(\omega_0 t - \mathbf{k}_0 \mathbf{r}), \\ \mathscr{E}_{0y} &= -\varepsilon_y E_0 \sin(\omega_0 t - \mathbf{k}_0 \mathbf{r}), \end{aligned} \tag{A.1.1}$$

where ε_x and ε_y are the components of the vector of pumping field transverse polarization satisfying the conditions $\varepsilon_x^2 + \varepsilon_y^2 = 1$ and $\varepsilon_x > \varepsilon_y \ge 0$. Taking into account Eqs. (2.1) and (2.4), we obtain $E_{0x} = \varepsilon_x E_0$ and $E_{0y} = -i\varepsilon_y E_0$. We can therefore write (cf. [8])

$$\begin{aligned} &2^{3/2} E_0^3 [2|\mathbf{E}_0|^2 - \mathbf{E}_0^2 \exp\{-2i(\omega_0 t - \mathbf{k}_0 \mathbf{r})\} \\ &\quad - \mathbf{E}_0^{*2} \exp\{2i(\omega_0 t - \mathbf{k}_0 \mathbf{r})\}]^{-3/2} \\ &= \{1 - \rho^4 \cos[2(\omega_0 t - \mathbf{k}_0 \mathbf{r})]\}^{-3/2} \\ &= A_0^{3/2}(\rho^2) + 2\sum_{m=0}^{\infty} A_m^{3/2}(\rho^2) \cos[2m(\omega_0 t - \mathbf{k}_0 \mathbf{r})], \end{aligned} \tag{A.1.2}$$

where

$$\rho^4 = 1 - 4\varepsilon_x^2 \varepsilon_y^2 = (\varepsilon_x^2 - \varepsilon_y^2)^2. \tag{A.1.3}$$

Further, we will need two coefficients of the Fourier transform of (A.1.2),

$$A_0^{3/2}(\rho^2) = \frac{2}{\pi(1-\rho^2)\sqrt{1+\rho^2}} E\left(\sqrt{\frac{2\rho^2}{1+\rho^2}}\right), \tag{A.1.4}$$

$$\begin{aligned} A_1^{3/2}(\rho^2) &= \frac{2\sqrt{1+\rho^2}}{\pi\rho^2(1-\rho^4)} \\ &\times \left\{E\left(\sqrt{\frac{2\rho^2}{1+\rho^2}}\right) + (1-\rho^2) K\left(\sqrt{\frac{2\rho^2}{1+\rho^2}}\right)\right\}. \end{aligned} \tag{A.1.5}$$

Here, $E(z)$ and $K(z)$ are the complete elliptical integrals.

Equations (A.1.1)–(A.1.5) allow Eq. (3.9) to be rewritten in the form

$$\begin{aligned} \left\langle \delta \frac{1}{u_E(t)} \right\rangle &= -\frac{m\omega_0}{2^{1/2}|e|E_0^3} \\ &\times \{ e^{i\mathbf{kr}} ([\mathbf{E}_0 \delta\mathbf{E}_-^* + \mathbf{E}_0^* \delta\mathbf{E}_+] A_0^{3/2}(\rho^2) \\ &- A_1^{3/2}(\rho^2)[\mathbf{E}_0 \delta\mathbf{E}_+ + \mathbf{E}_0^* \delta\mathbf{E}_-^*]) \\ &+ e^{-i\mathbf{kr}} ([\mathbf{E}_0 \delta\mathbf{E}_-^* + \mathbf{E}_0^* \delta\mathbf{E}_-] A_0^{3/2}(\rho^2) \\ &- A_1^{3/2}(\rho^2)[\mathbf{E}_0 \delta\mathbf{E}_- + \mathbf{E}_0^* \delta\mathbf{E}_+^*]) \}. \end{aligned} \tag{A.1.6}$$

Using these results and Eqs. (3.8) and (3.11), we can, according to (3.12), write

$$\delta n_+ = -n_{e0} \frac{e^2}{4m_e^2 \omega_0^2 V_T^2} \{\boldsymbol{\alpha}_0 \delta\mathbf{E}_-^* + \boldsymbol{\alpha}_0^* \delta\mathbf{E}_+\}. \tag{A.1.7}$$

Here,

$$\begin{aligned} \boldsymbol{\alpha}_0 &= \mathbf{E}_0 - \frac{\nu(E) k_B n_{e0}}{\chi k^2} \\ &\times \{\mathbf{E}_0 A_0^{3/2}(\rho^2) - \mathbf{E}_0^* A_1^{3/2}(\rho^2)\}, \end{aligned} \tag{A.1.8}$$

where

$$\nu(E) = \frac{8\sqrt{2}\pi e^2 e_i^2 n_{i0} \Lambda}{m_e^2 V_E^3}.$$

Equations (3.13) and (A.1.7) give

$$\begin{aligned} &\left(2ik_0 \frac{d}{dz} - k^2\right) \delta\mathbf{E}_+ \\ &= -\frac{\omega_{Le}^2}{c^2} \frac{e^2 \mathbf{E}_0}{4m_e^2 V_T^2} (\boldsymbol{\alpha}_0 \delta\mathbf{E}_-^* + \boldsymbol{\alpha}_0^* \delta\mathbf{E}_+), \end{aligned} \tag{A.1.9}$$

$$\left(-2ik_0\frac{d}{dz}-k^2\right)\delta\mathbf{E}_-^* = -\frac{\omega_{Le}^2}{c^2}\frac{e^2\mathbf{E}_0^*}{4m_e^2V_T^2}(\boldsymbol{\alpha}_0\delta\mathbf{E}_-^* + \boldsymbol{\alpha}_0^*\delta\mathbf{E}_+). \quad \text{(A.1.10)}$$

Hence, the dependence of field perturbations is ~exp(*Gz*), where the spatial amplification coefficient is given by the equation

$$1+\frac{\omega_{Le}^2}{c^2}\frac{e^2}{4m_e^2\omega_0^2V_T^2}\times\left(\frac{\boldsymbol{\alpha}_0^*\mathbf{E}_0}{2ik_0G-k^2}-\frac{\boldsymbol{\alpha}_0\mathbf{E}_0^*}{2ik_0G+k^2}\right) = 0. \quad \text{(A.1.11)}$$

Because

$$\boldsymbol{\alpha}_0\mathbf{E}_0^* = \boldsymbol{\alpha}_0^*\mathbf{E}_0 = E_0^2\left\{1-\frac{\nu(E)k_Bn_{e0}}{\chi k^2}[A_0^{3/2}(\rho^2)-\rho^2A_1^{3/2}(\rho^2)]\right\} \quad \text{(A.1.12)}$$

and

$$A_0^{3/2}(\rho^2)-\rho^2A_1^{3/2}(\rho^2) = \frac{2}{\pi\sqrt{1+\rho^2}}K\left(\sqrt{\frac{2\rho^2}{1+\rho^2}}\right), \quad \text{(A.1.13)}$$

the solution to Eq. (A.1.11) can be written in the form

$$G = \pm\frac{1}{2k_0}\left\{-k^4+\frac{\omega_{Le}^2k^2V_E^2}{2c^2V_T^2}\times\left[1-\frac{2\nu(E)k_Bn_{e0}}{\pi\chi k^2\sqrt{1+\rho^2}}K\left(\sqrt{\frac{2\rho^2}{1+\rho^2}}\right)\right]\right\}^{1/2}. \quad \text{(A.1.14)}$$

Clearly, the filament spatial amplification coefficient is maximum when the wave vector is given by Eq. (4.14). Then

$$G_{\max}^2 = \frac{V_E^4\omega_{Le}^2}{64V_T^4c^2k_0^2}\left\{\frac{\omega_{Le}^2}{c^2}-\frac{16\nu_{ee}(V_T)\nu(E)}{CV_E^2}\times\frac{2}{\pi\sqrt{1+\rho^2}}K\left(\sqrt{\frac{\rho^2}{1+\rho^2}}\right)\right\}. \quad \text{(A.1.15)}$$

In the $\rho^2 = 0$ limit, which corresponds to circular pumping polarization, Eq. (4.11) follows from Eq. (A.1.14). In the opposite limit of a low circular polarization degree *A*, when

$$A^2 = 1-\rho^4 \ll 1 \quad \text{(A.1.16)}$$

and

$$\frac{2}{\pi\sqrt{1+\rho^2}}K\left(\sqrt{\frac{\rho^2}{1+\rho^2}}\right)\approx\frac{\sqrt{2}}{\pi}\ln\frac{8}{A}\approx 0.45\ln\frac{8}{A}, \quad \text{(A.1.17)}$$

the role played by a decrease in the electron temperature as the intensity of radiation grows increases logarithmically, which prevents filamentation. As a result, an additional factor equal to Eq. (A.1.17) appears in Eq. (4.14), which determines the filamentation threshold.

This increase in the filamentation threshold determined by polarization corresponds with the polarization dependence of plasma heating by a high-power pumping radiation established in [4], when the effective frequency of collisions (4.6), which characterizes absorption of elliptically polarized pumping field, is given by

$$\nu_{\text{eff}}(E) = \frac{8\sqrt{2}\pi e^2e_i^2n_{i0}\Lambda}{m_e^2V_E^3}\times\frac{2}{\pi\sqrt{1+\rho^2}}K\left(\sqrt{\frac{\rho^2}{1+\rho^2}}\right). \quad \text{(A.1.18)}$$

In the (A.1.16) limit, this equation takes the form

$$\nu_{\text{eff}}(E) = \frac{16e^2e_i^2n_{i0}\Lambda}{m_E^2V_E^3}\ln\frac{8}{A}. \quad \text{(A.1.19)}$$

The last formula is inapplicable if

$$A < 8(V_T/V_E), \quad \text{(A.1.20)}$$

when the influence of thermal motion on radiation absorption in a plasma should be taken into account in describing the effects under consideration [8].

Lastly, let us write the generalization of Eqs. (4.4) and (4.8) that follows from our analysis,

$$\langle\delta Q_{IB}\rangle = -\frac{e^2n_{e0}}{4m_e\omega_0^2}\nu_{\text{eff}}(E)\delta|\mathbf{E}|^2, \quad \text{(A.1.21)}$$

$$\langle Q_{IB}^{(0)}\rangle = \frac{e^2n_{e0}}{2m_e\omega_0^2}\nu_{\text{eff}}(E)E_0^2. \quad \text{(A.1.22)}$$

In particular, it follows from these equations that the filament spatial amplification coefficient can be written in form (5.5).

APPENDIX 2

We cannot completely ignore the influence of thermal motion of electrons on the inverse bremsstrahlung absorption of a strong pumping field in the particular

case of a linearly polarized pumping field $\mathscr{E}_0 = (\mathscr{E}_0, 0, 0)$, where

$$\mathscr{E}_0 = E_0 \cos(\omega_0 t - \mathbf{k}_0 \mathbf{r}).$$

Indeed, the rate of electron oscillations then equals $\mathbf{u}_E = (u_E, 0, 0)$, where

$$u_E(t) = -V_E \sin(\omega_0 t - \mathbf{k}_0 \mathbf{r})$$

can vanish at a certain time moment. Equation (A.1.2) then becomes meaningless at $\rho^2 = 1$.

In view of this, the heat absorbed by a plasma in a strong pumping field will be written as (cf. [4])

$$Q_{IB} = \frac{4\pi e^2 e_i^2 n_i \Lambda}{m_e} \int d\mathbf{u} \frac{F_0(\mathbf{u})}{|\mathbf{u} + \mathbf{u}_E(t)|}. \tag{A.2.1}$$

We then have the absorbed heat perturbation

$$\delta Q_{IB} = \delta u_E \frac{4\pi e^2 e_i^2 n_i \Lambda}{m_e} n_e \frac{i}{\pi} \int_0^\infty q dq \exp\left(-\frac{1}{2} q^2 V_T^2\right) \times \int_{-1}^{+1} x dx \exp[iqx u_E(t)]. \tag{A.2.2}$$

This equation takes into account that $F_0(\mathbf{u})$ is a Maxwell distribution. In (A.2.1),

$$\delta u_E = \frac{ie}{2 m_e \omega_0} \times (\delta E \exp[-i(\omega_0 t - \mathbf{k}_0 \mathbf{r})] - \delta E^* \exp[i(\omega_0 t - \mathbf{k}_0 \mathbf{r})]), \tag{A.2.3}$$

which corresponds to the field perturbation

$$\delta \mathscr{E} = \frac{1}{2} \{ \delta E \exp[-i(\omega_0 t - \mathbf{k}_0 \mathbf{r})] + \delta E^* \exp[i(\omega_0 t - \mathbf{k}_0 \mathbf{r})] \}. \tag{A.2.4}$$

Using the expansion in Bessel functions $J_n(z)$

$$\exp[iqx u_E(t)] = \sum_{n=0}^{\infty} J_n(qxV_E) \exp[-in(\omega_0 t - \mathbf{k}_0 \mathbf{r})]$$

and the equality $\delta|\mathbf{E}|^2 = E_0(\delta E + \delta E^*)$, we obtain from (A.2.4)

$$\langle \delta Q_{IB} \rangle = \frac{e^2 n_e}{m_e \omega_0^2} \frac{2 e^2 e_i^2 n_{i0} \Lambda}{m_e^2 V_E} \delta |\mathbf{E}|^2 I, \tag{A.2.5}$$

where

$$I = \frac{d}{dV_E} \int_{-1}^{+1} dx \int_0^\infty dq \exp\left(-\frac{1}{2} V_T^2 q^2\right) J_0(qxV_E). \tag{A.2.6}$$

Using the equation

$$\int_0^\infty dz \exp(-y^2 z^2) J_0(pz) = \frac{\sqrt{\pi}}{2y} \exp\left(-\frac{p^2}{8y^2}\right) I_0\left(\frac{p^2}{8y^2}\right)$$

and the asymptotic relation for the Bessel function of an imaginary argument

$$I_0(z) \approx \frac{1}{\sqrt{2\pi z}} e^z,$$

we obtain

$$I \approx -\frac{2}{V_E^2} \ln \frac{V_E}{2V_T} \tag{A.2.7}$$

in the strong field limit. It follows that the absorbed field perturbation averaged over fast oscillations can be written as

$$\langle \delta Q_{IB} \rangle = -\frac{1}{4} \nu_{\text{eff}}(E) \frac{e^2 n_{e0}}{m_e \omega_0^2} \delta |\mathbf{E}|^2. \tag{A.2.8}$$

Here,

$$\nu_{\text{eff}}(E) = \frac{16 e^2 e_i^2 n_{i0} \Lambda}{m_e^2 V_E^3} \ln \frac{V_E}{2V_T} \tag{A.2.9}$$

is the nonlinear effective frequency of collisions, which characterizes pumping field absorption:

$$\langle Q_{IB}^{(0)} \rangle = \frac{1}{2} \nu_{\text{eff}}(E) \frac{e^2 n_{e0}}{m_e \omega_0^2} E_0^2. \tag{A.2.10}$$

Equation (A.2.8) can, with logarithmic accuracy, be written in a form similar to Eq. (4.9) obtained for circularly polarized pumping,

$$\langle \delta Q_{IB} \rangle = \frac{\langle \partial Q_{IB} \rangle}{\partial (E_0^2)} \delta |\mathbf{E}|^2. \tag{A.2.11}$$

Lastly, applying (3.11) yields

$$\frac{\delta n_e}{n_{e0}} = -\frac{e^2 \delta |\mathbf{E}|^2}{4 m_e^2 \omega_0^2 V_T^2} \left\{ 1 - \frac{\nu_{\text{eff}}(E) \nu_{ee}(V_T)}{C k^2 V_T^2} \right\}, \tag{A.2.12}$$

which describes density perturbations. We again see that heat effects suppress the ponderomotive effect, which generally causes radiation filamentation. If (A.2.12) is written as

$$\frac{\delta n_e}{n_{e0}} = -\frac{\delta |\mathbf{E}|^2}{4 m_e^2 \omega_0^2 V_E^2} \left\{ \frac{V_E^2}{V_T^2} + \frac{4}{k^2 \chi T_{e0}} \frac{\partial \langle Q_{IB}^{(0)} \rangle}{\partial \ln (E_0^2)} \right\}, \tag{A.2.13}$$

we can write Eq. (5.5), because Eq. (A.2.13) is similar to the equation that follows from (4.10) and (4.11).

REFERENCES

1. A. G. Litvak, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **11**, 1433 (1968); B. G. Eremin and A. G. Litvak, Pis'ma Zh. Éksp. Teor. Fiz. **13**, 603 (1971) [JETP Lett. **13**, 430 (1971)].
2. W. L. Kruer, Comments Plasma Phys. Control. Fusion **9**, 33 (1985).
3. V. P. Silin, *Parametric Action of High Power Radiation on Plasma* (Nauka, Moscow, 1973).
4. V. P. Silin, Zh. Éksp. Teor. Fiz. **111**, 478 (1997) [JETP **84**, 262 (1997)].
5. A. Ya. Polishchuk and J. Meyer-Ter-Vehn, Phys. Rev. E **49**, 663 (1994).
6. V. P. Silin, *An Introduction to the Kinetic Gas Theory* (Fiz. Inst. Ross. Akad. Nauk, Moscow, 1998).
7. V. P. Silin, Pis'ma Zh. Éksp. Teor. Fiz. **67**, 313 (1998) [JETP Lett. **67**, 329 (1998)].
8. V. P. Silin, Zh. Éksp. Teor. Fiz. **47**, 2254 (1964) [Sov. Phys. JETP **20**, 1510 (1964)].

Translated by V. Sipachev

Journal of Experimental and Theoretical Physics, Vol. 92, No. 4, 2001, pp. 626–633.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 4, 2001, pp. 719–727.

PLASMA, GASES

Heating of Deuterium Clusters by a Superatomic Ultra-Short Laser Pulse

V. P. Kraĭnov[a, *] and M. B. Smirnov[b, **]

[a]*Moscow Physicotechnical Institute,*
Institutskiĭ proezd 9, Dolgoprudnyĭ, Moscow oblast, 141700 Russia
[b]*Institute of Molecular Physics, Russian Research Centre Kurchatov Institute,*
pl. Kurchatova 1, Moscow, 123182 Russia
**e-mail: krainov@cyberax.ru*
***e-mail: smirnov@imp.kiae.ru*
Received November 13, 2000

Abstract—The mechanisms of heating of the electronic component of large deuterium clusters by a superatomic ultra-short laser pulse field are considered. During pulse rise, the so-called "vacuum heating" plays the determining role. Electrons escaping from a cluster into the vacuum with a low energy return back in a time equal to the period of the laser under laser field action. The returning electrons have a higher energy (on the order of the vibrational energy in the laser radiation field), which causes cluster heating. As the laser field increases, the electronic temperature largely grows at the expense of decreasing the Coulomb potential energy of electron repulsion because of a decrease in the number of electrons. The dynamics of above-barrier cluster ionization at the leading edge of a superatomic laser pulse is calculated. The results are discussed in the light of recent experiments aimed at creating desktop sources of monoenergetic neutrons formed as a result of the fusion of deuterium nuclei in a cluster plasma. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Recent experimental studies of photoionization of atomic and molecular clusters by femtosecond laser pulses showed that exciting large clusters comprising thousands atoms makes it possible to obtain superheated microplasma with an electron temperature up to several keV. Of great interest are deuterium clusters because of a possibility of creating a plasma with a kinetic energy of deuterium atoms sufficient for tunnel nuclear fusion to occur when two such nuclei collide [1, 2]. Deuterium nuclei acquire a kinetic energy of several keV as a result of a Coulomb explosion of clusters after removing all electrons from them by a laser field.

This work is concerned with the mechanism of heating of the electronic component of deuterium clusters at the leading edge of a superatomic laser pulse, when the radius of the cluster slightly grows during very fast detachment of electrons from parent deuterium molecules in the cluster (so called inner ionization) and further escape of these electrons from the cluster itself (so-called outer ionization). We suggest two mechanisms of this heating. According to the first mechanism, an electron with a low kinetic energy that escapes from the cluster under laser field action can return back in a fraction of the period of the laser, this time with an energy on the order of the vibrational energy in the laser radiation field. After the return, this electron collides with other electrons, and the thermalization of the electronic component occurs. The temperature of this component gradually grows as radiation at the leading edge of the laser pulse becomes increasingly intense.

The other mechanism of heating of the electronic component is related to the conversion of the Coulomb potential energy of electrons that remain within a cluster ion at a given time moment of outer ionization into their kinetic energy. This mechanism is based on the model of multiple ionization of a large Thomas–Fermi cluster by a strong electromagnetic field; the model was developed by us in [3]. In this model, electrons of a large cluster ion occur within a sphere of radius $R' < R$, where R is the radius of the ionic component. Within this sphere, a plasma of electrons and positively charged ions is neutral, and an electric field is absent. The concentration of electrons in the cluster ion does not change during outer ionization, and only radius R' of the electronic component decreases. Such a plasma approach to the problem requires that the Debye radius of screening of the Coulomb field of electrons

$$r_D = \sqrt{\frac{T}{4\pi N}}$$

be smaller than radius R'. Here, T is the electron temperature and N is the concentration of electrons. We will, as a rule, use the atomic system of units, in which the charge and the mass of the electron and the Planck constant equal one.

We will not discuss the mechanism of heating of cluster electrons related to the induced inverse bremsstrahlung

effect. In the presence of a laser field, electrons scattered on atomic ions largely absorb field energy rather than emit [4]. This mechanism is known for clusters of heavy elements (for instance, xenon atoms) [5], where multiply charged atomic ions are formed, and the inverse bremsstrahlung effect is substantial. In this work, we only discuss singly charged deuterium ions, for which the probability of this effect is low.

We also do not consider heating of electrons as a result of excitation of the Mie collective dipole resonance (surface plasmon) [5]. The resonance arises when the laser field frequency coincides with the Mie frequency. In the system under discussion, the laser field frequency is much lower than the Mie frequency.

Let us turn to the problem of the penetration of a laser field into a cluster plasma. At a low ionization degree, a laser field freely penetrates through the whole cluster ion. This follows from numerical calculations of xenon atom clusters [6], in which the $5p^6$ and $5s^2$ outer shells were largely ionized. At a high ionization degree of heavy element clusters, a laser field penetrates a very short distance into the cluster because of absorption caused by the induced inverse bremsstrahlung effect in elastic scattering of free cluster electrons on multiply charged atomic cluster ions [7]. For clusters of deuterium molecules considered in this work, this effect is insubstantial, see above. The penetration depth related to the excitation of plasma oscillations [8] has the form

$$\delta = \frac{c}{\omega\sqrt{|\varepsilon|}}. \quad (1)$$

Here, ω is the laser radiation frequency, and the permittivity caused by conduction electrons is

$$\varepsilon = 1 - \frac{\omega_p^2}{\omega^2} < 0,$$

where ω_p is the plasma frequency,

$$\omega_p = \sqrt{4\pi N} = 8.4 \text{ eV}.$$

A typical laser radiation frequency is much lower, and this is the reason why screening occurs. We will use the $\omega = 1.55$ eV frequency in our calculations, which corresponds to a 800 nm light wave length [1, 2]. The depth of field penetration then exceeds 300 Å; that is, it is substantially larger than the diameter of deuterium clusters.

In a cluster, deuterium molecules attract each other by van der Waals forces and form a dielectric liquid. The concentration of deuterium nuclei in this liquid (equal to the concentration of electrons) is $N = 5.15 \times 10^{22}$ cm^{-3} [9], as in a macroscopic deuterium liquid at a temperature below its boiling point; that is, N is of the order of the concentration of atoms in metals. In agreement with this concentration value, the mean distance between deuterium molecules equals 3.4 Å, whereas the distance between two atoms in the deuterium molecule is much smaller and only amounts to 0.7 Å. We assume that a large cluster comprising thousands deuterium molecules has a spherical shape (although experimentally, this was only proved for large metal atom clusters).

2. INNER IONIZATION

At a certain point during superatomic laser pulse increase, inner ionization of deuterium molecules in the cluster begins. All electrons are detached from parent nuclei and become collectivized (but still do not escape from the cluster), intramolecular bonds dissociate, and naked deuterium nuclei are formed.

Such an ionization is an above-barrier process. It continues for a short time of one-two femtoseconds (a similar picture is observed in the ionization of xenon clusters by a superatomic laser pulse, see [6]). Indeed, suppose that the intensity of the laser at a pulse maximum is $I = 5 \times 10^{16}$ W/cm^2. The corresponding maximum field is $F = 1.17$ au (we assume linear field polarization). According to the numerical calculations [10], above-barrier ionization of atomic hydrogen (deuterium) in 3 fs (that is, approximately during the period of the laser that generates in the visible region at 800 nm) requires a much weaker field $F = 0.08$ au. A similar value, $F = 1/16$ au, follows from the Bethe rule [11]. The same estimate is valid with hydrogen or deuterium molecules.

On the other hand, the calculations performed in [12] show that tunnel ionization can play the determining role in complete inner ionization at still lower fields because of a spatial distribution of laser intensity. This is not very important for describing subsequent outer ionization.

For a typical Gaussian pulse 35 fs wide [1, 2], the field strength envelope (in atomic units) depends on time t (in femtoseconds) as

$$F(t) = 1.17\exp(-t^2/780) \quad (2)$$

and field $F = 0.08$ au corresponds to the $t = -45$ fs time point. Here and in outer ionization calculations, time $t = 0$ corresponds to a maximum field value.

Intramolecular bonds, naturally, disappear during 1–2 fs at the very beginning of laser pulse rise, and the spherical deuterium cluster becomes a completely ionized dense neutral plasma comprising free electrons and deuterium nuclei. For instance, a cluster with $R = 25$ Å (this is a typical experimental value [1, 2]) contains

$$n = N(4\pi/3)R^3 = 3370$$

electrons. The dielectric cluster therefore becomes metallic!

3. OUTER IONIZATION

Consider outer ionization, when electrons escape from the surface of the cluster outward. We assume this ionization to be also field (cold) and above-barrier; that is, the laser field is superatomic not only for the inner but also for the outer ionization. Thermal ionization (the vaporization of heated electrons from the surface

of the cluster) described by the Richardson–Dushman equation is insignificant under these conditions because of the fast occurrence of the ionization. Competing with outer above-barrier ionization is tunnel ionization at a weaker field, which we neglect. Of course, this neglect can also prove incorrect in the light of the results obtained in [12] for atomic ionization, and this problem requires further inquiries.

The Bethe condition [11], which we apply to the classical above-barrier ionization of electrons from the cluster, has the simple form

$$F(t) = \frac{E_Z^2}{4Z}. \quad (3)$$

Here, $Z = Z(t)$ is the charge of the cluster ion at point t and E_Z is the potential of its further ionization equal to the Coulomb potential of the corresponding cluster ion,

$$E_Z = \frac{Z(t)}{R}, \quad (4)$$

where R is the radius of the cluster determined by deuterium nuclei (we assume that Coulomb repulsion between deuterium nuclei does not cause a significant increase in this radius at the leading edge of the laser pulse, see the corresponding estimates given below).

According to Eq. (3), at time t determined from this relation, $Z(t)$ electrons escape from the cluster.

The remaining $n - Z$ electrons of the cluster ion occur within a sphere of a smaller radius R' determined by the condition of electrical neutrality of this sphere (see Introduction); that is, by the condition

$$\frac{n-Z}{n} = \left(\frac{R'}{R}\right)^3. \quad (5)$$

The region between R and R' only contains deuterium nuclei. It follows from Eqs. (3) and (4) that the number of escaped electrons, Z, is related to the laser field amplitude at a given time moment by the simple equation

$$Z(t) = 4R^2F(t), \quad t < 0. \quad (6)$$

According to this equation, all 3370 electrons of a deuterium cluster with $R = 25$ Å escape outward in field $F = 0.38$ au at $t = -29.5$ fs [see Eq. (1)]. From a cluster of radius 50 Å, 26 960 electrons that this cluster contains escape in field $F = 0.75$ au at $t = -18.5$ fs. Similar estimates can easily be obtained for other maximum laser pulse intensities.

Of course, this approach is valid at a fairly high peak strength F, when the ionization is above-barrier; that is, when the condition

$$F > \frac{n}{4R^2} = \frac{\pi NR}{3} \quad (7)$$

is satisfied. At a fixed peak laser field value, this condition bounds the deuterium cluster radius R from above.

4. VACUUM HEATING OF CLUSTER ELECTRONS

An electron escaping from the $r = R'$ surface of the cluster ion with a certain energy comes under the influence of the laser radiation field and can be returned by this field in a fraction of the period of the laser depending on the laser field phase φ. The electron then acquires energy on the order of electron vibrational energy. This is the so-called "vacuum heating" mechanism suggested by Brunel [13] to describe interaction of laser radiation with metallic surfaces. The electron vibrational energy at a given time averaged over the laser radiation period is

$$U_p = \frac{F^2(t)}{4\omega^2}. \quad (8)$$

The one-dimensional Newton equation for electron motion in a linearly polarized field along the field polarization axis has the form

$$\frac{d^2x}{dt^2} = F(t)\cos(\omega t + \varphi). \quad (9)$$

Integrating Eq. (9) yields the current electron velocity (time in the laser pulse envelope is treated as a parameter)

$$\frac{dx}{dt} = v + \frac{F(t)}{\omega}[\sin(\omega t + \varphi) - \sin\varphi]. \quad (10)$$

Here, v is the initial electron velocity at the moment of its escape from the cluster ion. The coordinate of the electron again becomes zero at time t when

$$x(t) = \frac{F(t)}{\omega^2}[\cos\varphi - \cos(\omega t + \varphi)] + \left[v - \frac{F(t)}{\omega}\sin\varphi\right]t = 0. \quad (11)$$

It follows that determining the return time requires solving the transcendental equation

$$\left[\frac{v}{v_F} - \sin\varphi\right]\varphi_0 = \cos(\varphi + \varphi_0) - \cos\varphi. \quad (12)$$

Here, the return phase $\varphi_0 = \omega t$ and the field velocity

$$v_F = F(t)/\omega$$

are introduced.

A solution to Eq. (12) is shown in Fig. 1 as the dependence of φ_0 on the initial phase φ at various initial velocity values, $v/v_F = 0$, 0.1, 0.2, …, 1.2, and 1.25. At $v = 0$, this solution is known for the corresponding problem of the theory of tunnel ionization of atoms by a strong low-frequency field [14]. The figure shows that, at each velocity value, the return only occurs in a limited interval of initial laser radiation phases φ. At a fairly high velocity $v \geq 1.25v_F$ (the central closed curve with the smallest area in Fig. 1), electrons do not at all

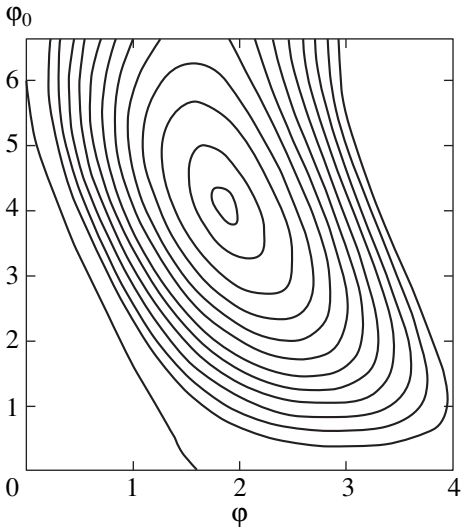


Fig. 1. Dependence of electron return time t (via phase $\varphi_0 = \omega t$) on initial laser field φ according to transcendental equation (12) at various escaping electron velocities v expressed in field velocity $v_F = F(t)/\omega$ units. The left-most curve corresponds to $v = 0$. Next follow curves with $v/v_F = 0.1$, 0.2, 0.3, …, 1.1, 1.2, and 1.25. The lowest return time at a given initial laser field phase φ corresponds to the first return, the next value, to the second return, etc.

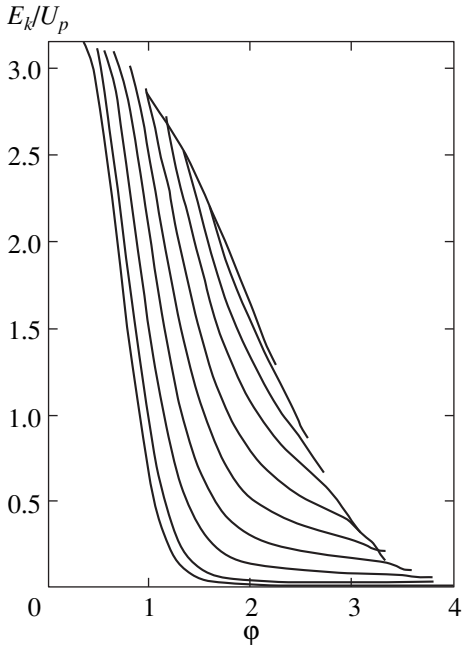


Fig. 2. Dependence of the kinetic energy of an electron returning to the cluster (in vibrational energy U_p units) on the initial laser field phase φ according to Eq. (13) at the same escaping electron velocities as in Fig. 1. Velocities increase from the left-most curve to the right.

return into the cluster but fly away to infinity at velocity v. It should, of course, be borne in mind in the calculations that an electron can escape in both positive and negative axis x directions.

In this approach, an electron is considered a classical particle, and electron wave packet spreading after the ionization is ignored. From the quantum-mechanical point of view, this corresponds to the neglect of the transverse electron velocity with respect to the laser field polarization vector compared with its longitudinal velocity. The presence of a transverse velocity decreases the probability of the return of the electron back to the cluster. Compared with the corresponding problem of electron return in tunnel ionization of an atom, the favorable factor is the large size of the cluster. In addition, as in the atomic problem, of the greatest importance are electrons that escape with low energies, see below. Transverse velocities are then fairly small compared with longitudinal velocities. We will estimate them later on.

The curves shown in Fig. 1 allow us to calculate the kinetic energy of the electron at the moment of its return back to the cluster as a function of initial phase φ,

$$E_k = \frac{1}{2}\left(\frac{dx(t)}{dt}\right)^2 = 2U_p\left[\frac{v}{v_F} + \sin(\varphi + \varphi_0) - \sin\varphi\right]^2. \quad (13)$$

The calculation results are plotted in Fig. 2 for the same electron velocities as in Fig. 1. Note that, according to Corkum [14], a maximum kinetic energy value for the ionization of atoms is $E_k = 3.17U_p$. This value is attained at a zero initial velocity and phase $\varphi = 17°$.

Assuming this phase to take on random values allows Eq. (13) to be averaged uniformly with respect to the phase. The result is shown in Fig. 3, where the mean energy of returning electrons is plotted as a function of the initial electron velocity at which it escapes

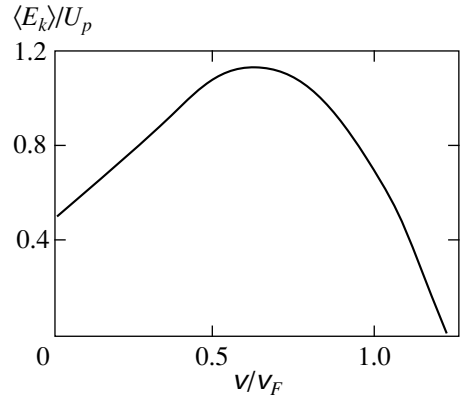


Fig. 3. Dependence of the mean kinetic energy of electrons returning to the cluster (in vibrational energy U_p units) on their velocity v at the moment of escape from the cluster expressed in field velocity $v_F = F(t)/\omega$ units.

from the $r = R'$ cluster surface. This energy accounts for a considerable fraction of the vibrational energy, and, therefore, the Brunel mechanism [13] of so-called vacuum heating can in principle cause heating of electrons remaining inner the cluster ion at the leading edge of a laser pulse and should be taken into account in the energy balance for the outer ionization of the cluster.

Clearly, if the escaping electron energy $v^2/2$ is subtracted from the returning electron energy, the result may be both positive (electron component heating) and negative (electron component cooling). For this reason, predominant initial velocities v of electrons escaping during the outer ionization play a very important role. We can use the results obtained for the tunnel and above-barrier ionizations of atoms [15]. The distribution of escaping electrons with respect to their longitudinal velocities (along the laser field polarization vector) has the form of a Gaussian curve with a maximum at a zero velocity,

$$dZ \propto \exp\left(-\frac{v^2\gamma^3}{3\omega}\right)dv. \tag{14}$$

Here, we introduced the Keldysh parameter [16]

$$\gamma = \frac{\omega\sqrt{2E_Z}}{F(t)} = \omega\sqrt{\frac{8R}{F(t)}}. \tag{15}$$

We used Eqs. (4) and (6) to derive the equation for this parameter. Strictly, Eq. (14) is valid for the tunnel ionization. For the above-barrier ionization, the tunnel exponent is replaced by the Airy function (see the corresponding distributions in [17]). We, however, wish to determine the characteristic longitudinal velocities rather than the energy distribution of escaping electrons per se, and longitudinal velocities have similar symbolic estimates for the tunnel and above-barrier ionizations [17]. According to Eq. (14), the typical longitudinal velocity important for escaping electrons can be estimated as

$$v = \sqrt{\frac{3\omega}{\gamma^3}}, \quad \frac{v}{v_F} = \frac{\sqrt{3}}{[F(t)]^{1/4}(8R)^{3/4}}. \tag{16}$$

At a laser pulse maximum ($t = 0$), Eq. (16) yields $v/v_F = 0.02$, whereas at $t = -45$ fs ($F = 0.08$ au), we find $v/v_F = 0.04$.

Equation (14) is valid if the Keldysh parameter γ is smaller than or close to unity. For the cluster and laser field parameters used above as typical examples, this parameter value is on the order of one.

As far as the characteristic transverse velocity $v_\perp$, which determines wave packet spreading, is concerned, its estimates for the above-barrier and tunnel ionizations also coincide, namely [17],

$$v_\perp \propto \frac{\sqrt{F(t)}}{(2E_Z)^{1/4}} \ll v. \tag{17}$$

The conclusion can therefore be drawn that a large majority of electrons escape from clusters at low velocities compared with the field velocity (as in atoms in the tunnel ionization). According to Fig. 3, the mean kinetic energy of an electron flying back into the cluster therefore equals

$$E_k \approx U_p/2.$$

It follows that the laser radiation energy absorbed by cluster ion electrons per unit time is given by the simple equation

$$\frac{d\varepsilon}{dt} = \frac{F^2(t)}{8\omega^2}\frac{dZ}{dt}. \tag{18}$$

Here, dZ/dt is the number of electrons escaping from the cluster per unit time. A part of these electrons return in a fraction of the laser period back into the cluster and heat electrons that remain there, whereas another part irreparably fly away to infinity. This process continues until electrons still remain in the cluster. As mentioned above, the process terminates before $t = 0$, when the laser pulse field is maximum. The electron wave packet spreading mentioned above can only slightly decrease the factor in Eq. (18).

5. ENERGY BALANCE

Energy (18) is spent to heat the electronic component and change the potential energy of electrons (no energy exchange between the electronic and ionic components occurs during the ultra-short laser pulse). A part of the energy is lost with electrons emitted in the outer ionization. The energy balance has the form [7]

$$\frac{d\varepsilon}{dt} - \left[\frac{3}{2}T - \frac{Z}{R}\right]\frac{dZ}{dt} = \frac{d}{dt}\left\{\frac{3}{2}(n - Z(t))T(t) + \frac{3}{5}\frac{(n - Z(t))^2}{R'}\right\}. \tag{19}$$

This equation takes into account that cluster ion electrons are uniformly distributed within a sphere of radius $R' < R$ determined by Eq. (5) (see Introduction). The first term in the right-hand side of Eq. (19) is the change in the kinetic energy of electrons that remain in the cluster ion at a given time moment per unit time. The second term is the rate of changes in the Coulomb potential energy of electrons.

The second term in the left-hand side of Eq. (19) is the energy of electrons that escape from the cluster ion per unit time. The $3T/2$ value is the mean energy of an escaping electron inner the cluster ion, and Z/R is the energy lost by the electron when it flies from the cluster outward.

The $T(t)$ value is the electron temperature. Of course, establishing this temperature requires effective

electron-electron collisions inner the cluster. Such collisions do occur if the free path of electrons is smaller than R'. Let us estimate the free path by the formulas of the theory of collisions in a plasma [18]. The electron temperature present in these formulas will be replaced by the mean kinetic energy of electrons returning to the cluster, $E_k = 3T/2$,

$$l = v_e \tau_{ee} = \sqrt{2E_k}\frac{3(2E_k/3)^{3/2}}{4\sqrt{2\pi}N\ln\Lambda}. \quad (20)$$

According to the calculations discussed above, the kinetic energy of an electron is estimated here as $E_k = U_p/2$, $v_e = \sqrt{2E_k}$ is the velocity of an electron inner the cluster, the typical value of the Coulomb logarithm $\ln\Lambda$ can be set equal to 10, and N is the concentration of electrons inner the cluster (this concentration does not change in the outer ionization because electrons are contracted into a sphere of a smaller radius). At a typical laser pulse field $F = 0.25$ au, Eq. (18) yields $l = 4$ Å. It follows that electrons frequently collide with each other inner the cluster, which ensures a Maxwell temperature distribution. An exception is the final stage of outer ionization, when l strongly increases, whereas R' decreases. These considerations justify not only the use of Eq. (19) as the energy balance equation but also the electron rescattering mechanism itself.

Substituting Eqs. (5) and (18) into Eq. (19) eventually yields the following equation for determining the electron temperature:

$$(n-Z)\frac{dT}{dt} = \frac{2}{3R} \times\left[Z + \frac{Z^2}{128R^4\omega^2} + n^{1/3}(n-Z)^{2/3}\right]\frac{dZ}{dt}. \quad (21)$$

Here, the number of escaped electrons $Z(t)$ is given by Eq. (6).

The second term in the right-hand side of Eq. (21) responsible for the energy that returns to the cluster during vacuum heating is small compared with the first term. Indeed, the ratio between these terms is

$$\frac{Z}{128R^4\omega^2}.$$

This ratio reaches a maximum when $Z = n$. For a cluster of radius 25 Å, it equals $0.08 \ll 1$.

In view of these considerations, vacuum heating can be ignored, and heating of electrons that remain inner the cluster ion can be assumed to be fully determined by the Coulomb potential energy of electrons.

Ignoring the vacuum heating term, let us rewrite Eq. (21) in the form

$$(n-Z)\frac{dT}{dZ} = \frac{2}{3R}[Z + n^{1/3}(n-Z)^{2/3}]. \quad (22)$$

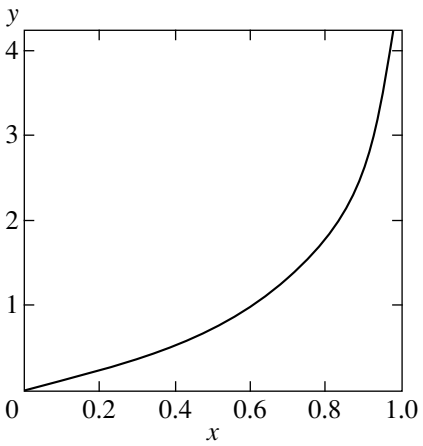


Fig. 4. Universal dependence of the electron temperature of a cluster ion (in $y = 3RT/2n$ units) on the ion charge (in $x = Z/n$ units).

The integration of Eq. (22) yields the universal relation between the electron temperature and the charge of the cluster ion. Let us introduce the notation

$$y = \frac{3RT}{2n}, \quad x = \frac{Z}{n}.$$

On the assumption that the temperature is zero in the absence of outer ionization, we obtain

$$y = -\ln(1-x) - x + \frac{3}{2}[1-(1-x)^{2/3}]. \quad (23)$$

This dependence is plotted in Fig. 4, where x varies from 0 to 1. The time dependence of the electron temperature is determined by Eqs. (23) and (6), which gives the charge of the cluster ion as a function of time.

For instance, at $x = Z/n = 1/2$, Eq. (23) yields the electron temperature

$$T_{1/2} \approx \frac{n}{2R}.$$

In particular, for a cluster of radius 25 Å, $T_{1/2} = 0.97$ keV.

According to Eq. (22), the complete electromagnetic energy absorbed by one cluster equals its initial Coulomb energy (if the contribution of vacuum heating to the total energy is ignored in agreement with the above estimates):

$$\delta E = \frac{3n^2}{5R}. \quad (24)$$

The mean temperature of escaped electrons under the conditions when all of them escape from the cluster does not depend on the intensity of laser radiation (of course, if this radiation is superatomic) and equals

$$\langle T\rangle = \frac{2n}{5R}. \quad (25)$$

Note that this mean temperature is close to the $T_{1/2}$ temperature, at which a half of all electrons escape from the cluster. In particular, for a 25 Å cluster, the mean electron temperature is 0.78 keV, and for a 50 Å cluster, this temperature equals 3.1 keV. This value is independent of the peak laser radiation intensity because, in a stronger laser field, the middle part of the laser pulse "does not work."

Consider a cluster of large size, suppose, of radius 80 Å (such a cluster contains 110000 electrons). According to Eq. (6), laser field (1) is only capable of causing above-barrier ionization of 107000 electrons. It follows that the effectiveness of outer cluster ionization decreases as the size of the cluster increases (at a given laser radiation intensity).

6. CONCLUSION

After the escape of all electrons from a deuterium cluster of radius 25 Å in a field with a 5×10^{16} W/cm^2 peak intensity at time $t = -29$ fs, a sphere comprising positively charged deuterium nuclei is formed. The concentration of nuclei in this sphere is $N = 5.15 \times 10^{22}$ cm^{-3}. At this moment, a Coulomb explosion begins, and the whole potential Coulomb energy of this sphere, $n^2/2R$, converts into the kinetic energy of deuterium nuclei. The mean kinetic energy of one nucleus equals

$$E_d = n/2R = 0.95 \text{ keV},$$

and the maximum energy is

$$n/R = 1.9 \text{ keV}.$$

The cluster begins to expand fairly rapidly, and its radius increases to three times the initial radius is subsequent 30 fs; that is, by the time the laser pulse attains a maximum. Note that the expansion of a cluster of xenon atoms proceeds simultaneously with outer ionization because of the multiple character of the ionization [6].

As a result, a deuterium plasma is formed with an $N' = 10^{19}$ cm^{-3} mean concentration of electrons and nuclei [2]. The time of collisions of deuterium nuclei with each other can be estimated by the formula for plasma collisions [18]

$$\tau_{ii} = \frac{3\sqrt{M}(2E_d/3)^{3/2}}{4\sqrt{2\pi}N'\ln\Lambda}. \quad (26)$$

Here, M is the reduced mass of the deuterium nucleus in a collision of two nuclei with each other and E_d is its kinetic energy. Substituting the values cited above gives an estimate of 1–3 ns for the collision time. This leads us to conclude that there is certainly no time for a Maxwell distribution to be established in this plasma. Most likely, the kinetic energy distribution of deuterium nuclei should be considered uniform from zero to the $2E_d$ maximum value.

Clearly, in the absence of high-energy deuterons, the nuclear fusion reaction

$$d + d = {}^3\text{He} + n$$

virtually does not occur, because the cross section of the tunnel reaction is exceedingly small; it equals 10^{-37} cm^2 at an $E_d = 2$ keV deutron energy [19]. For the yield of neutrons to be appreciable, deutrons should have energies exceeding 10 keV.

It is likely that, in the experiments described in [1, 2], where neutron yields up to 10^4 per laser pulse were observed, there were collective mechanisms of energy transfer from electrons to deuterium nuclei with the formation of high-energy nuclei. One of such mechanisms can be related to a process similar to ambipolar diffusion, when electrons escaping from the cluster drag a part of deuterium nuclei along with them by Coulomb attraction forces. Diverse instabilities in a deuterium plasma can also be responsible for the appearance of high-energy deuterons. For instance, in [2], the mean kinetic energy of deuterons equaled 12 keV! Lastly, the mechanism suggested by A.D. Sakharov can operate. According to Sakharov, outer ionization proceeds very rapidly, and there occurs recoil-induced initial partial contraction of the ionic sphere resulting from the escape of all electrons. An analysis of these mechanisms requires special consideration.

ACKNOWLEDGMENTS

This work was financially supported by the Russian Foundation for Basic Research (project nos. 99-02-17810 and 01-02-16056). The authors thank Prof. T. Ditmire for valuable advice and the participants of the Moscow Seminar on Multiphoton Processes (headed by Prof. N.B. Delone) for discussions.

REFERENCES

1. T. Ditmire *et al.*, Nature (London) **398**, 489 (1999).
2. J. Zweiback *et al.*, Phys. Rev. Lett. **85**, 3640 (2000).
3. M. B. Smirnov and V. P. Kraĭnov, Zh. Éksp. Teor. Fiz. **115**, 2014 (1999) [JETP **88**, 1102 (1999)].
4. M. V. Fedorov, *Atomic and Free Electrons in a Strong Light Field* (World Scientific, Singapore, 1997).
5. T. Ditmire *et al.*, Phys. Rev. A **53**, 3379 (1996).
6. I. Last and J. Jortner, Phys. Rev. A **62**, 013201 (2000).
7. V. P. Kraĭnov and M. B. Smirnov, Usp. Fiz. Nauk **170**, 969 (2000).

8. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 8: *Electrodynamics of Continuous Media* (Nauka, Moscow, 1982; Pergamon, New York, 1984).
9. *Handbook of Chemistry and Physics*, Ed. by D. R. Lide (CRC Press, London, 1998–1999), 79th ed.
10. D. Bauer and P. Mulser, Phys. Rev. A **59**, 569 (1999).
11. H. A. Bethe and E. E. Salpeter, *Quantum Mechanics of One- and Two-Electron Atoms* (Academic, New York, 1957; Fizmatgiz, Moscow, 1960).
12. M. V. Ammosov and N. B. Delone, Laser Phys. **7**, 79 (1997).
13. F. Brunel, Phys. Rev. Lett. **59**, 52 (1987).
14. P. B. Corkum, Phys. Rev. Lett. **71**, 1994 (1993).
15. N. B. Delone and V. P. Krainov, *Multiphoton Processes in Atoms* (Springer-Verlag, Berlin, 2000), 2nd ed.
16. L. V. Keldysh, Zh. Éksp. Teor. Fiz. **47**, 1945 (1964) [Sov. Phys. JETP **20**, 1307 (1964)].
17. V. P. Krainov, J. Opt. Soc. Am. B **14**, 425 (1997).
18. V. P. Silin, Zh. Éksp. Teor. Fiz. **47**, 2254 (1964) [Sov. Phys. JETP **20**, 1510 (1964)].
19. E. Teller, *Fusion* (Academic, New York, 1981).

Translated by V. Sipachev

Journal of Experimental and Theoretical Physics, Vol. 92, No. 4, 2001, pp. 634–642.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 4, 2001, pp. 728–737.

SOLIDS
Structure

The Observation of the Contribution of Longitudinal Susceptibility to the Frequency of the Soft Magnetoresonance Mode in $SmFeO_3$

V. D. Buchel'nikov[a, *], N. K. Dan'shin[b, **], Yu. I. Nepochatykh[b], and V. G. Shavrov[c, *]**

[a]*Chelyabinsk State University, Chelyabinsk, 454021 Russia*
[b]*Donetsk Physicotechnical Institute, National Academy of Sciences of Ukraine, Donetsk, 340114 Ukraine*
[c]*Institute of Radio Engineering and Electronics, Russian Academy of Sciences, ul. Mokhovaya 11, Moscow, 101999 Russia*
**e-mail: buche@cgu.chel.su*
***e-mail: danshin@host.dipt.donetsk.ua*
****e-mail: shavrov@mail.cplire.ru*

Received August 9, 2000

Abstract—The behavior of the spectrum of the soft magnetoresonance mode in samarium orthoferrite was experimentally and theoretically studied in the region of the $\Gamma_4 \longleftrightarrow \Gamma_{24}$ orientation phase transition induced by an external magnetic field. The special features of this behavior can be explained by the relative contributions of interaction between ferrite subsystems and longitudinal susceptibility. It is also shown that the contribution of longitudinal susceptibility to the gap of the soft magnetoresonance mode in samarium orthoferrite can also be substantial in low fields; that is, in the vicinity of spontaneous orientation phase transitions. This is explained by the occurrence of spontaneous orientation phase transitions in samarium orthoferrite at high temperatures, at which longitudinal susceptibility is comparable in magnitude with transverse susceptibility. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The first convincing evidence of a contribution of longitudinal susceptibility to resonance frequencies of ordered magnets was obtained in [1] for the example of weak ferromagnets $YFeO_3$ and $DyFeO_3$, which are rare-earth metal orthoferrites. The selection of these compounds and the method used to detect the corresponding effect were not fortuitous. By then, the static properties of rare-earth metal orthoferrites had been studied fairly completely and reliably established in a large number of theoretical and experimental works. In particular, magnetic phase transitions suitable for purposes of such a study had been found. The corresponding data constitute a necessary prerequisite for constructing a theory of the dynamic properties of these magnets and for correctly interpreting the corresponding dynamic experiments. The detailed data on the crystal and magnetic structure of rare-earth metal orthoferrites, which are also used in this work, can be found in monograph [2], which allows us only briefly to characterize the object of study. Most rare-earth metal orthoferrites contain two magnetic subsystems, iron *d* and rare-earth metal *f* subsystems, and crystallize in space group D_{2h}^{16}. At temperature T_N (600–700 K for various rare-earth metal orthoferrites), iron spins are ordered to form a weakly canted antiferromagnetic (weakly ferromagnetic) structure F_z, G_x corresponding to the Γ_4 irreducible representation of the space group specified above. Here, F_z and G_x are the components of the ferromagnetic and antiferromagnetic vectors $\mathbf{F} = \mathbf{M}_1 + \mathbf{M}_2$ and $\mathbf{G} = \mathbf{M}_1 - \mathbf{M}_2$ ($\mathbf{M}_1$ and $\mathbf{M}_2$ are the magnetizations of iron sublattices). The rare-earth metal subsystem is in the paramagnetic state at T_N. Only in some of the rare-earth metal orthoferrites, this subsystem undergoes magnetic ordering as a result of *f–f* exchange interactions as temperature decreases to $T < 10$ K. On the other hand, because of *f–d* interactions, there occurs ordering of rare-earth metal ions induced by the iron subsystem, and ferro- and antiferromagnetic vectors of their own can also be put in correspondence to these ions. Strictly, the static magnetic structure of rare-earth metal orthoferrites is described by four *d* and four *f* sublattices. These eight sublattices can, however, be reduced to four sublattices (two *d* and two *f*) for describing the dynamics of acoustic modes studied in this work. The presence of two interacting magnetic subsystems in rare-earth metal orthoferrites is responsible for the occurrence of various orientation phase transitions. The general reason for their appearance is anisotropic and temperature-dependent *f–d* interaction. The contribution of rare-earth metal ions to the thermodynamic potential of the magnetic subsystem increases as temperature lowers. Because the magnetic properties of rare-earth metal ions are exceedingly anisotropic, this results in a spontaneous change in the orientation of magnetization of

iron sublattices with respect to the crystallographic axes; that is, in an orientation phase transitions. In most rare-earth metal orthoferrites, such transitions largely occur as smooth rotations of the **F** and **G** vectors in the *ac* plane of the crystal. As a result of temperature lowering, a magnet spontaneously transforms from one symmetrical phase, $\Gamma_4(F_z, G_x)$, to another, $\Gamma_2(F_x, G_z)$, through the $\Gamma_{24}(F_{x,z}, G_{x,z})$ canted phase; that is, the $\Gamma_4 \longrightarrow \Gamma_{24} \longrightarrow \Gamma_2$ "cascade" of phase transitions occurs. The T_1 and T_2 temperatures of the onset and completion of vector rotations correspond to second-order orientation phase transition points and are specific to each rare-earth metal orthoferrite. When temperature increases, phase transitions occur in the reverse order. The $\Gamma_4 \longleftrightarrow \Gamma_{24}$ and $\Gamma_{24} \longleftrightarrow \Gamma_2$ transitions can separately be induced by applied magnetic fields, **H** || **c** and **H** || **a**, respectively.

It is shown in the theoretical model developed in [1] that the occurrence of an "order–order" transition in external magnetic field **H** is a necessary condition for observing the contribution of longitudinal susceptibility to the dynamics of magnets. According to [1], this contribution results in the appearance of a gap in the spectrum of spin waves precisely at induced orientation phase transition points. A visual test suggested in [1] can conveniently be used in experiments and allows conclusions to be drawn both on the presence and magnitude of the corresponding effect. In a first approximation, the gap in the spectrum of an experimentally observed softening magnetoresonance mode is given by

$$\nu_0 = \frac{\gamma}{2\pi}\left(\frac{\chi_\parallel}{\chi_\perp}\right)^{1/2} H_{\mathrm{tr}},$$

where γ is the gyromagnetic ratio; $\chi_\parallel$ and $\chi_\perp$ are the longitudinal and transverse susceptibilities, respectively; and H_{tr} is the field of the phase transition. It follows from this equation that soft mode activation increases when either temperature T (because $\chi_\parallel/\chi_\perp \propto T$) or applied magnetic field grows. Simultaneously increasing transition temperature T_{tr} and field H_{tr} induces the $\Gamma_2 \longleftrightarrow \Gamma_{24}$ orientation phase transition. We selected this transition to test the validity of the theory suggested in [1]. For this transition, external parameters act in the same direction on the width of the gap. In the experiments described in [1], the gap was observed at fairly high values of both temperature, $T > 100$ K, and field, $H > 60$ kOe.

The theory [1] did not take into account the mechanisms that inevitably lead to the appearance of an energy gap as a result of dynamic interactions between various vibrational subsystems of a magnet (ordered spin, paramagnetic, magnetoelastic, and dipole subsystems) and between vibrations of these subsystems and longitudinal magnetization oscillations of rare-earth metal orthoferrite sublattices [3, 4]. If interactions of various subsystems of a magnet are ignored, gap ν_0 in the model [1] should vanish as $H, T \longrightarrow 0$. Gap ν_0, however, never disappears in experiments. This raises the question of what role is played by longitudinal susceptibility in the formation of a resonance frequency at ultimately small values of external parameters or, at least, one of them. Of interest for practical applications is the case of $H \longrightarrow 0$, because, along with field-induced orientation phase transitions, spontaneous orientation phase transitions occur in ordered magnets no less frequently. For instance, orientation phase transitions of similar structures can be induced in rare-earth metal orthoferrites by both field and temperature at $H = 0$. At present, various orientation phase transitions were observed in the absence of a field in eight (of 15 known) rare-earth metal orthoferrites. In essence, our task was to elucidate the role played by longitudinal susceptibility in the formation of the dynamics of an ordered magnet in the vicinity of spontaneous orientation phase transitions. In this respect, the present work differed from work [1] aimed at studying field-induced orientation phase transitions. It is made clear in the discussion that follows that precisely samarium orthoferrite $SmFeO_3$ of all rare-earth metal orthoferrites plays the key role in fulfilling the task specified above.

2. STATEMENT OF THE PROBLEM. EXPERIMENTAL METHOD AND MEASURING TECHNIQUE

Whereas reorientation onset and completion temperatures T_1 and T_2 are specific to each compound and substantially different for different rare-earth metal orthoferrites, their T_N iron ion ordering temperatures are virtually equal ($T_N \approx 600$ K for $SmFeO_3$). This provides a unique possibility, inherent only in rare-earth metal orthoferrites, of selecting a sequence of compounds of the same class with orientation phase transitions of the same structure but with spontaneous reorientation temperatures that differ by more than an order of magnitude. For good reason, the effects under discussion are more correct to analyze for a series of precisely one-type magnets rather than for substituted or isomorphous substances. Such a series is convenient to characterize by the dimensionless relative temperature of spontaneous reorientation $\tau_{SR} = T_{SR}/T_N$, where $T_{SR} = (T_1 + T_2)/2$ [4, 5]. All rare-earth metal orthoferrites can then be arranged in a series between two limiting values of this parameter, $\tau_{SR} = 0$ and $\tau_{SR} = 1$. These τ_{SR} values correspond to the (also limiting) $\chi_\parallel/\chi_\perp \approx 0$ and $\chi_\parallel/\chi_\perp = 1$ ratio values, respectively, at $T = 0$ and $T = T_N$.

If the whole series of rare-earth metal orthoferrites is considered, at least two questions arise. First, whether or not in all (if not in all, then in what) rare-earth metal orthoferrites, the longitudinal susceptibility in the $H \longrightarrow 0$ limit plays a noticeable role in the formation of the soft mode frequency. Secondly, what role is played by longitudinal susceptibility in the vicinity of the $\Gamma_4(F_z, G_x) \longleftrightarrow \Gamma_{24}(F_{x,z}, G_{x,z})$ orientation phase transition, where the field and the temperature of the

transition change along the line of this second-order orientation phase transition in opposite directions and, therefore, give opposite contributions to the effects caused by longitudinal susceptibility.

The contribution of longitudinal susceptibility to the dynamics will be estimated by the values of the $\partial\nu_0/\partial H$ ($H \longrightarrow 0$) derivatives and $\partial\nu_0/\partial T$ ($T \longrightarrow T_{SR}$) derivatives related to the former by the phase diagram. The measurements performed earlier for several rare-earth metal orthoferrites $RFeO_3$ (R = Yb, Tm, and Er) [6, 7] covered the range $\tau_{SR} = 0.01$–0.15 but gave no indications that these derivatives differed from zero up to 10 kOe fields. Clearly, this excludes the possibility of detecting the effects under discussion also in $NdFeO_3$, whose τ_{SR} equals 0.17 [8]. All the other rare-earth metal orthoferrites except $SmFeO_3$ (R = Tb, Dy, Ho, etc.) in which spontaneous reorientation occurs have still lower τ_{SR} parameter values. Samarium orthoferrite is the only rare-earth metal orthoferrite in which the Γ_2 phase exists at temperatures above room temperature, and τ_{SR} amounts to approximately 0.8. The estimate obtained in [1] from the temperature dependence of the $\chi_\parallel/\chi_\perp$ ratio shows that this ratio is about 0.7 in $SmFeO_3$; that is, it is close to its maximum limiting value $\chi_\parallel/\chi_\perp = 1$. Previous attempts at reproducing the spectrum of the soft magnetoresonance mode of $SmFeO_3$ by traditional methods were, however, unsuccessful because of a number of technical and physical difficulties, the overcoming of which required additional efforts.

Traditional measuring cells for low-temperature measurements could not be used in experiments with $SmFeO_3$, whose spontaneous phase transition temperatures T_1 and T_2 exceeded 400 K. This required the development of wide-band waveguide cells with controlled heating that would allow temperature to be stabilized and scanned in a fairly narrow range.

It is known that the paramagnetic subsystem of rare-earth metal ferrites (the *f* subsystem) can cause substantial damping of oscillations of ordered *d* subsystem spins. Damping increases as temperature grows. For this reason, it can result in a substantial broadening of magnetoresonance lines in $SmFeO_3$ and, eventually, weakening of the effects under consideration. This proved to be the principal difficulty of experiments with this magnet. To overcome it, special attention was given to the quality of the sample. We selected a high-quality crystal about 1 mm^3 in volume whose shape was difficult to identify with any geometric figure.

The sample was glued at its *ac* plane in the center of a piston that short-circuited a rectangular waveguide with oscillations of the H_{10} type. We selected the "less favorable" $\Gamma_4(F_z, G_x) \longleftrightarrow \Gamma_{24}(F_{x,z}, G_{x,z})$ transition for measurements, which allowed us to answer both questions raised above in an experiment with a single sample mounting. The mutual orientation of the **F** vectors and the magnetic component of a microwave field **h** optimal for the observation of absorption lines is $\mathbf{F} \perp \mathbf{h}$. The **F** vector and **c** axis were therefore directed along the narrow waveguide cross section. In addition, the accuracy of the orientation of the field along the corresponding crystal axes is of fundamental importance in measurements of this kind. The orientation phase transition under consideration is induced by field $\mathbf{H} \parallel \mathbf{c}$. In our experiments, the required accuracy of orienting **H** along **c** (within approximately 20 minutes of arc) was attained by the traditional method based on a high sensitivity of the spectrum of a soft magnetoresonance mode to deviations of the field from crystal axes in the *ac* plane [1].

Measurements were performed on the direct amplification spectrometer that was used in all our previous experiments and operated in the reflected power mode. Special measures were, however, taken to increase its sensitivity and resolution by more than an order of magnitude. The most important step of modernization was the use of a modulation mode in which magnetic field was modulated but temperature scanned. The external magnetic field remained constant, and its value was set as a parameter in each particular case. As a result, we recorded the derivative of the absorption line with respect to temperature. Outside the immediate vicinity of orientation phase transitions (approximately ±3% from T_{tr}), the half-width of resonance lines along the temperature axis was on average not larger than 10 K.

The use of the improved procedure allowed us to attain virtually the same accuracy of measurements as in similar experiments on the other rare-earth metal orthoferrites [1, 6, 7]. This in the first place refers to the accuracy of determining energy gap values at transition points (the error did not exceed ±2.5 GHz for gaps 27–50 GHz wide). Our ultimate goal was not only to determine the width of gaps but to measure its dependence on external field and temperature, and the range of gap measurements had to ensure that gap variations be substantially larger than the error of measurements. It will be shown that this requirement was satisfied by temperature variations in the range 400–490 K and by using fields up to 12 kOe.

3. EXPERIMENTAL RESULTS

An analysis of the totality of resonance experiments in the vicinity of field-induced orientation phase transitions in rare-earth metal orthoferrites and isomorphous weak ferromagnet Fe_3BO_6 [9, 10] revealed one remarkable feature of the spectrum of the soft mode important for the purposes of this work. The experimental points in the temperature and field dependences of soft mode resonance frequencies lay very close to straight lines on both sides of the orientation phase transition line. This allows reliable results to be obtained by linearly extrapolating experimental dependences, if necessary. The *H*–*T* phase diagrams of rare-earth metal orthoferrites exhibited similar behavior; they also had the form of straight lines, at least, for $H \longrightarrow 0$.

Examples of the recorded derivatives of absorption lines with respect to temperature obtained in a 6 kOe field at various frequencies are shown in Fig. 1. The left and right lines in each record refer to absorption in the Γ_{24} and Γ_4 phases, respectively. Circles correspond to the centers (maxima) of absorption lines. The totality of such records obtained at many radiation frequencies were used to reproduce the temperature-field dependences of soft magnetoresonance modes. As follows from Fig. 1, the temperatures of the maxima of the right and left absorption lines approach each other as the frequency decreases. Theoretically, they should coincide at the frequency which is the energy gap in the spectrum of homogeneous spin waves. The temperature of the maximum of the absorption line for this frequency is the temperature of the orientation phase transition. However practically, the gap in $SmFeO_3$ can only be determined by extrapolation (see above), because comparatively broad resonance lines overlap each other when the system approaches the orientation phase transition point (see record *3* in Fig. 1), and determining the positions of the centers of the resonance lines becomes problematic. This prevents a direct reproduction the magnetoresonance spectrum in the immediate vicinity of the orientation phase transition, although in reality, the width of this "inaccessible" region is comparatively small and does not exceed ±2% of T_{tr}.

The temperature-field dependences of soft mode frequencies in the vicinity of the induced $\Gamma_4(F_z, G_x) \longleftrightarrow \Gamma_{24}(F_{x,z}, G_{x,z})$ orientation phase transition at two external field values, 4 and 6 kOe, are shown in Fig. 2. Similar dependences were obtained at the other external field H values (0.1, 2, 8, 10, and 12 kOe). In these experiments, experimental values lie close to straight lines. This allows us to reliably perform linear extrapolations of the temperature dependences of resonance frequencies on the sides of the $\Gamma_4(F_z, G_x)$ and $\Gamma_{24}(F_{x,z}, G_{x,z})$ phases up to their intersection at the transition point. As a whole, the spectrum of the soft mode assumes the shape of a "frequency wedge" whose cusp indicates the T_{tr} induced transition temperature (such a wedge was observed in all earlier measurements of this kind). In addition to T_{tr}, the position of the intersection point unambiguously determines the absolute energy gap ν_0 value in the spectrum of spin waves in a given external field. Processing energy gap values obtained in this way by the method of least squares shows that the corresponding values are best described by a linear temperature dependence with the slope $\partial\nu_0/\partial T = -0.47$ GHz/K. Figure 2 shows that gap ν_0 increases almost twofold (approximately by 25 GHz) as temperature lowers from 460 to 410 K. This increase is an order of magnitude larger than the error of energy gap measurements. Note that the $\partial\nu_0/\partial T$ derivative is negative (Fig. 2).

The H–T phase diagram of $SmFeO_3$ in field **H** ∥ **c** constructed based on the data given in Fig. 2 is shown in Fig. 3. This diagram corresponds to the field-induced

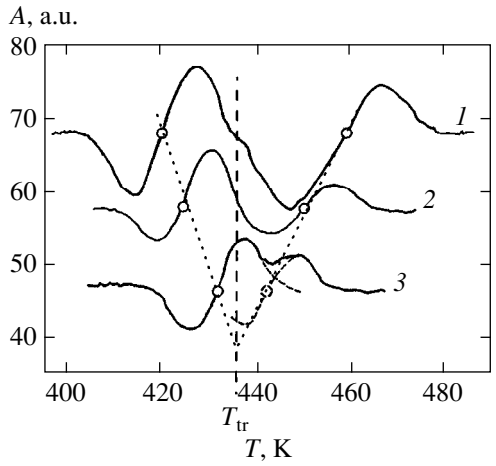


Fig. 1. Absorption signal derivatives with respect to temperature in a 6 kOe magnetic field recorded at frequencies ν = 68.22 (*1*), 56.07 (*2*), and 47.4 (*3*) GHz; *A* is the amplitude of absorption signal derivative in arbitrary units; open circles are resonance absorption maxima on the temperature axis; portions of resonance lines that cannot be observed experimentally because of their interference near the orientation phase transition point are shown by dashed lines for record *3*.

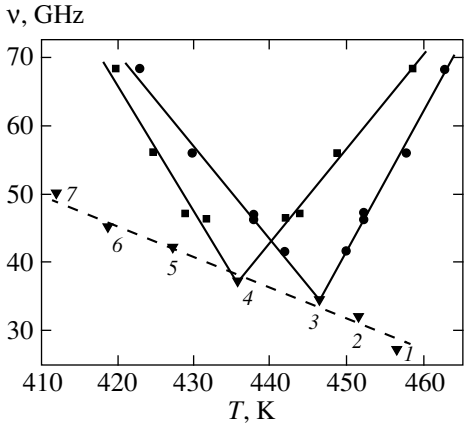


Fig. 2. Temperature dependences of soft mode frequencies at points of Γ_{24}–Γ_4 reorientation completion in $SmFeO_3$ in field **H** ∥ **c** (fields (●) 4 kOe and (■) 6 kOe) and (▼) temperature dependence of energy gap under the same conditions in external fields H (kOe): (*1*) 0.1, (*2*) 2.0, (*3*) 4.0, (*4*) 6.0, (*5*) 8.0, (*6*) 10.0, and (*7*) 12.0.

$\Gamma_4(F_z, G_x) \longleftrightarrow \Gamma_{24}(F_{x,z}, G_{x,z})$ orientation phase transition, and the extrapolation of the line of this transition to $H = 0$ gives the temperature of the spontaneous orientation phase transition of the same structure, $T_1 = (458 \pm 3)$ K. Note that the T_1 and T_2 values and, to a lesser extent, T_N in rare-earth metal orthoferrites depend on the procedure for the preparation of their single crystals, purity of the starting material, and, ultimately, the quality of samples used in actual measurements. For this reason, if all temperature parameters are taken from various sources rather than measured for the same sample, calculations inevitably include the corre-

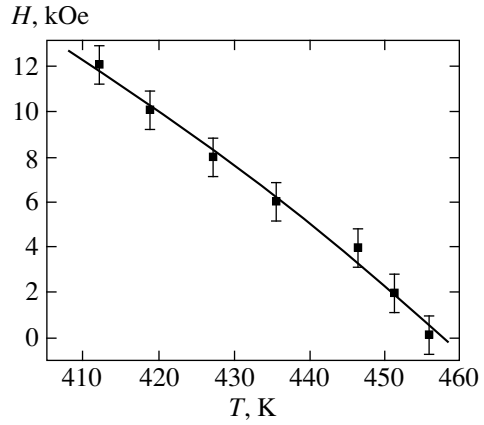


Fig. 3. Phase H–T diagram of $SmFeO_3$ in field **H** || **c**.

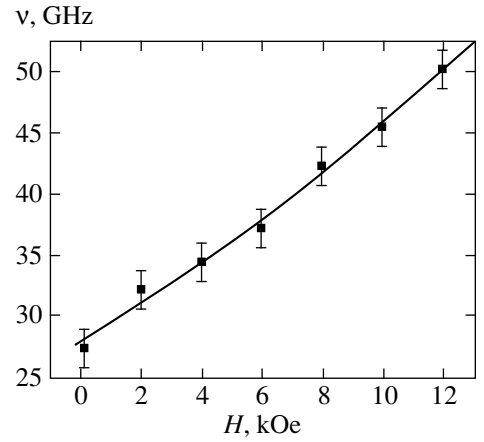


Fig. 4. Field dependence of soft mode gaps at points of Γ_{24}–Γ_4 spin reorientation completion in $SmFeO_3$ in field **H** || **c**.

sponding errors and give parameter estimates rather than parameter values. We also do not claim that our results are anything more than mere estimates. For this reason, taking into account that $\Delta T = T_1 - T_2$ is always much smaller than T_N, we assumed that $T_1 = T_2$ to obtain the $\tau_{SR} \approx 0.8$ estimate for $SmFeO_3$ (see above).

The field dependence of the energy gap, also constructed based on the data given in Fig. 2, is shown in Fig. 4. When $H \longrightarrow 0$, this dependence is characterized by the $\partial\nu_0/\partial H \approx 1.52$ GHz/kOe derivative. The data corresponding to $T \longrightarrow T_1$ (Fig. 2) or $H \longrightarrow 0$ (Fig. 4) can be used to determine the "starting" energy gap for the $\Gamma_4(F_z, G_x) \longleftrightarrow \Gamma_{24}(F_{x,z}, G_{x,z})$ spontaneous orientation phase transition. This energy gap equals (27.7 ± 2.5) GHz, which is substantially smaller than the value obtained in our earlier work [3] (about 40 GHz) with the use of the traditional procedure (by temperature variations in zero field) characterized by lower sensitivity and resolution. On the other hand, the gap value may, as mentioned above, be noticeably influenced by the quality of the initial single crystal used to prepare the sample. In [11], even a method for indirectly estimating uniformity of rare-earth metal ferrite crystals by gap sensitivity to external field orientation was suggested; according to [11], more magnetically perfect crystals have narrower gaps.

Attention should be given to at least three features of the results obtained in this work. First, as expected, the energy gap in the spectrum of $SmFeO_3$ spin waves substantially depends on temperature and magnetic field even in the $H \longrightarrow 0$ limit; that is, in the vicinity of the spontaneous orientation phase transition. Samarium orthoferrite is the only rare-earth metal orthoferrite that exhibits this effect. The results of this work in combination with those reported earlier for the other rare-earth metal orthoferrites and isomorphous Fe_3BO_6 actually put a stop to discussing possibilities of observing longitudinal susceptibility contributions to gaps in the vicinity of spontaneous orientation phase transitions in the other rare-earth metal ferrites. Indeed, on the spontaneous reorientation temperature scale, there is no rare-earth metal orthoferrites between $NdFeO_3$ and $SmFeO_3$ with suitable τ_{SR} values. In a further study of the problem for rare-earth metal orthoferrites, the gap can only be filled by mutual substitution of rare-earth metal ions with competing anisotropy, for instance, by substituting Tm^{3+} for Sm^{3+}.

According to one more important observation made in this work, the energy gap increases as temperature lowers (we have already mentioned that the $\partial\nu_0/\partial T$ derivative is negative). At first sight, this contradicts the theory suggested in [1], according to which temperature (longitudinal susceptibility) and field act in one direction and increase the energy gap. In reality, the observed gap widening under the conditions of differently directed external parameter actions is the result of competing contributions of T and H into the gap value. The field contribution to ν_0 is positive and larger in magnitude, which is responsible for the resultant gap increase observed experimentally. Note that the incommensurateness of the temperature and field contributions to such a dynamic characteristic as the gap is predetermined by the static parameters of the magnet reflected by its H–T phase diagram. If the H–T phase diagram is qualitatively and quantitatively similar to that shown in Fig. 3, the temperature and field contributions will always be incommensurate, and the effect will apparently be a field effect even when $H \longrightarrow 0$. Note that the point in question is longitudinal susceptibility-induced gap changes rather than the gap value itself. Observing gap changes, however, requires fulfilling the conditions at which this effect can in principle be observed. In our view, this is only possible if $\tau_{SR} \geq 0.4$–0.5, considering the present-day level of experimental studies.

Lastly, the third point to be mentioned is that gap ν_0 begins to increase in a field starting with its fairly large value at $H = 0$. This means that the results of this work cannot be correctly described by either the theory sug-

gested in [1] or the theory suggested in [3]. A more general theoretical model is required to simultaneously and most completely take into account interactions between various magnet vibrational subsystems and the contributions of both precession and longitudinal magnetization oscillations to the dynamics of a magnet. Fundamentals of such a theory were developed and summarized in [4]. We will use this theory to interpret the experimental results obtained in this work.

4. THEORY

Samarium orthoferrite is an orthoferrite with a non-Kramers rare-earth metal ion (with an odd number of 4*f* electrons). The phenomenological free energy density of such ferrites is given by Eqs. (3.1)–(3.7) from [3]. Because it follows from experiment that the influence of longitudinal susceptibility on the spectrum of coupled vibrations is strong in samarium orthoferrite, we should abandon the approximation of equal and constant magnetizations of the *d* and *f* subsystem sublattices in describing its dynamics (condition (3.5) in [3]). For this purpose, the equation for free energy *F* density should be augmented by the term

$$\frac{1}{2}A_1\mathbf{G}^2 + \frac{1}{4}A_2\mathbf{G}^4 + \frac{1}{2}D'\mathbf{G}^2\mathbf{F}^2 \qquad (1)$$

(see the equation for free energy density of a two-sublattice antiferromagnet [4]).

Consider the equilibrium state corresponding to the Γ_4 phase. The corresponding samarium orthoferrite equilibrium parameter values are found by minimizing the free energy. The minimization results are as follows:

$$\begin{gathered} F_x = F_y = f_x = f_y = G_y = G_z \\ = c_x = c_y = c_z = u^0_{ik} = 0 \quad (i \neq k), \\ F_z = F_0, \quad G_x = G_0, \quad f_z = f_0, \\ F_0 = \chi_\perp(2M_0H - dG_0), \quad f_0 = \frac{\mu_z H + B_x G_0}{\lambda_3'}. \end{gathered} \qquad (2)$$

The modulus of the antiferromagnetic vector of the *f* subsystem is found from the equation

$$\begin{gathered}[A_1 + A_2G_0^2 + D'F_0^2 + 2(B_{11}u^0_{xx} + B_{12}u^0_{yy} + B_{13}u^0_{zz})]G_0 \\ = NB_xf_0 - dF_0,\end{gathered}$$

and the equilibrium components of the deformation tensor u^0_{ii} are given by Eqs. (3.8a) from [3]. In Eq. (2), $\chi_\perp = (A + D'G_0^2)^{-1}$ is the transverse susceptibility of samarium orthoferrite. The other denotations coincide with those used in [3]. As distinguished from [3], ground state (2) is written here using the condition of smallness of the isotropic exchange coupling constant (*a*) between the *d* and *f* subsystems from the outset.

An analysis of the matrix of the second derivatives of free energy density, which should be positive semidefinite, yields the following stability condition for phase Γ_4:

$$\begin{gathered} K_{ac}G_0^2 + \frac{NB_xG_0\mu_zH}{\lambda_3'} \\ + 2M_0H\chi_\perp(2M_0H\eta - dG_0) \geq 0, \end{gathered} \qquad (3)$$

where $\eta = 1 - \chi_\parallel/\chi_\perp$ and $\chi_\parallel = [A + (D + D')G_0^2]^{-1}$ is the longitudinal magnetic susceptibility of samarium orthoferrite. The other denotations coincide with those used in [3]. The equality sign in Eq. (3) determines the orientation phase transition point from the collinear Γ_4 phase to the canted Γ_{24} phase.

The dynamic properties of the Γ_4 samarium orthoferrite phase will be described with the use of the Landau–Lifshitz, elasticity, and Maxwell equations, see Eqs. (3.13)–(3.18) in [3]. In the Landau–Lifshitz equation, relaxation terms (3.14) [3] will be replaced by relaxation terms of a more general form, which describe both transverse and longitudinal relaxations in the *f* and *d* subsystems [12]:

$$\begin{gathered} \mathbf{R}(\mathbf{x}, \mathbf{y}, \mathbf{F}_x, \mathbf{F}_y) = -\Lambda_0\mathbf{F}_x \\ - \Lambda_\perp\{(x^2 + y^2)\mathbf{F}_x + 2(\mathbf{x}\cdot\mathbf{y})\mathbf{F}_y\} - (\Lambda_\perp - \Lambda_\parallel) \\ \times\{\mathbf{x}(\mathbf{x}\cdot\mathbf{F}_x) + \mathbf{x}(\mathbf{y}\cdot\mathbf{F}_y) + \mathbf{y}(\mathbf{x}\cdot\mathbf{F}_y) + \mathbf{y}(\mathbf{y}\cdot\mathbf{F}_x)\}, \end{gathered} \qquad (4)$$

where $\Lambda_{0,\perp,\parallel}$ are the relaxation parameters ($\Lambda_{0,\perp,\parallel} = \Lambda_{f0,\perp,\parallel}$ for the *f* subsystem and $\Lambda_{0,\perp,\parallel} = \Lambda_{d0,\perp,\parallel}$ for the *d* subsystem). For the *f* subsystem, **x**, **y** in Eq. (4) equals **f**, **c** in the Landau–Lifshitz equation for ferromagnetic vector **f**, and **x**, **y** = **c**, **f** in the Landau–Lifshitz equation for antiferromagnetic vector **c**, and $\mathbf{F}_x = \partial F/\partial\mathbf{x}$. For the *d* subsystem, **f**; **c** in Eq. (4) should be replaced by **F** and **G**, respectively.

It has been proved in [4] that, if longitudinal relaxation is taken into account and the assumption of equal and constant magnetization moduli of the *d* and *f* sublattices is abandoned, the classical Landau–Lifshitz equations remain applicable to the problem of describing the dynamic properties of magnets (also see [13, 14]).

After the linearization of the system of coupled equations near equilibrium (2), we obtain the following dispersion equation for coupled harmonic waves in the Γ_4 phase that propagate along the *z* axis:

$$\begin{gathered} (1-\eta)\omega^5 + i[\Lambda_f(1-\eta)(\omega_{1f} + \omega_{2f}) \\ + \Lambda_d\omega_E(2-\eta)]\omega^4 - [(\omega_d^2 + \omega_f^2)(1-\eta) + \Lambda_d^2\omega_E^2 \\ + \Lambda_f^2\omega_{1f}\omega_{2f}(1-\eta) + \Lambda_f\Lambda_d(1-\eta)(\omega_{1f} + \omega_{2f})\omega_E \\ \times(2-\eta)]\omega^3 - i\Big\{\Lambda_f(1-\eta)(\omega_{1f} + \omega_{2f})(\omega_d^2 - \omega_E\omega_{fd}')\end{gathered}$$

$$+\Lambda_d\omega_E(2-\eta)\left[\omega_d^2+\omega_f^2-\omega_{\rm dip}r-\omega_H(\omega_H+\omega_D)\right.$$
$$\left.+\frac{\omega_E\omega_{me5}\omega_{5k}^2}{\omega^2-\omega_{5k}^2}\right]\Big\}\omega^2+\Big\{\omega_f^2(1-\eta)$$
$$\times\left[\omega_d^2+\frac{\omega_E\omega_{me5}\omega_{5k}^2}{\omega^2-\omega_{5k}^2}-\omega_E\omega_{fd}\right]+\Lambda_d^2\omega_E^2\omega_f^2$$
$$+\Lambda_f^2\omega_{1f}\omega_{2f}(1-\eta)\left[\omega_d^2+\frac{\omega_E\omega_{me5}\omega_{5k}^2}{\omega^2-\omega_{5k}^2}-\omega_E\omega_{fd}\right]$$
$$+\Lambda_f\Lambda_d(\omega_{1f}+\omega_{2f})\omega_E\left[\omega_d^2-\omega_{\rm dip}r-\omega_H(\omega_H+\omega_D)\right.$$
$$\left.+\frac{\omega_E\omega_{me5}\omega_{5k}^2}{\omega^2-\omega_{5k}^2}-\omega_E\omega'_{fd}\right]\Big\}\omega-i\Lambda_d\omega_E\omega_f^2\left[\omega_d^2-\omega_E\omega_{\rm dip}r\right.$$
$$\left.-\omega_H^2(1-\eta)+r\omega_L(\omega_H+\omega_D)(\eta\omega_H+\omega_D)/\omega_E\right.$$
$$\left.+\frac{\omega_E\omega_{me5}\omega_{5k}^2}{\omega^2-\omega_{5k}^2}-\omega_E\omega_{fd}\right]=0, \quad (5)$$

where

$$\Lambda_f=\Lambda_{f0}+\Lambda_{f\perp}f_0^2,\quad \Lambda_d=\Lambda_{d0}+\Lambda_{d\perp}G_0^2,$$
$$\omega_{1,2f}=\frac{gN\lambda'_{1,2}}{M_B},\quad \omega_E=\frac{g\chi_\perp^{-1}}{M_0},\quad \omega_f=\frac{gNf_0\sqrt{\lambda'_1\lambda'_2}}{M_B},$$
$$\omega_d=[\omega_E(\omega_{S0}+\omega_{me5}+\omega_{fd}+\omega_{\rm dip}r)+\omega_H^2(1-\eta)]^{1/2},$$
$$\omega_{S0}=g\left[K_{ac}G_0^2+\frac{NB_x\mu_zG_0H}{\lambda'_3}\right.$$
$$\left.+2M_0H\chi_\perp(2\eta M_0H-dG_0)\right],$$
$$\omega_{me5}=\frac{gB_{55}^2G_0^4}{M_0c_{55}},\quad \omega_{fd}=\frac{gNB_z''^2G_0^2}{\lambda'_1M_B},$$
$$\omega_H=2gH,\quad \omega_{\rm dip}=\frac{\omega_L(\omega_H+\omega_D)^2}{\omega_E^2},\quad \omega_D=-\frac{gdG_0}{M_0},$$
$$\omega_L=16\pi gM_0,\quad \omega'_{fd}=\frac{gNB_z''^2G_0^2}{M_0(\lambda'_1+\lambda'_2)},$$
$$\omega_{5k}=s_5k,\quad r=\left(1-\frac{\nu^2k^2}{\varepsilon\omega^2}\right)^{-1}.$$

The other denotations coincide with those used in [3]. For simplicity of deriving the dispersion equation, the assumption $\lambda_7=\lambda_8=B'_z=0$ made in [3] was augmented by the assumption $\mu_{xy}=\mu_{yx}=\mu_y=\mu_x=0$.

For a comparison with experiment, we will only give solutions to dispersion equation (5) corresponding to elementary excitations of magnetic subsystems when $k\longrightarrow 0$. These solutions are

$$\omega_{1,2}^2=\frac{1}{2}\{\omega_d^2+\omega_f^2\pm[(\omega_d^2-\omega_f^2)^2+4\omega_f^2\omega_E\omega_{fd}]^{1/2}\}, \quad (6)$$

$$\omega_3^2=-i\Lambda_d\omega_E[\omega_d^2-\omega_E\omega_{\rm dip}-\omega_H^2(1-\eta)-\omega_E\omega_{fd}$$
$$+\omega_L(\omega_H+\omega_D)(\eta\omega_H+\omega_D)/\omega_E]$$
$$\times\{(1-\eta)[\omega_d^2-\omega_E\omega_{fd}]\}^{-1}. \quad (7)$$

The first two branches correspond to precession and the third one, to relaxation oscillations of magnetic subsystems. For simplicity, damping of precession oscillations was not taken into account in Eq. (6).

5. A COMPARISON OF THEORY AND EXPERIMENT. DISCUSSION

According to the experimental results, spontaneous orientation phase transitions occur in samarium orthoferrite at high temperatures, $T>400$ K. As is known [3], the *f* subsystem parameters and interactions of the *d* and *f* subsystems do not exceed several kelvins in temperature units. It follows that, for samarium orthoferrite in the region of orientation phase transitions, the approximations

$$f_0\ll 1,\quad \lambda'_i\approx T \quad (8)$$

are fairly accurate. Inequality (3) can then be used to obtain the following equation for field $H_{\rm tr}$ at which induced orientation phase transitions occur [the equality sign in Eq. (3)]:

$$H_{\rm tr}=\frac{\omega_E}{8gM_0\eta}\left\{\left[\left(\frac{2M_0\omega_D}{\omega_E}+\frac{\mu_zNG_0B_x}{T}\right)^2\right.\right.$$
$$\left.\left.-\frac{16gM_0\eta K_{ac}G_0^2}{\omega_E}\right]^{1/2}-\left(\frac{2M_0\omega_D}{\omega_E}+\frac{\mu_zNG_0B_x}{T}\right)\right\}. \quad (9)$$

The anisotropy constant K_{ac} tends to zero in the region of a spontaneous orientation phase transition [2], and, at low external fields, Eq. (9) can therefore be considerably simplified,

$$H_{\rm tr}=-\frac{M_0K_{ac}G_0^2}{2M_0\omega_D/\omega_E+\mu_zNG_0B_x/T}. \quad (10)$$

Assuming that the second term in the denominator of Eq. (10) is small compared with the first one at high

temperatures ($T > 400$ K) allows the equation for the transition field to be further simplified to

$$H_{\rm tr} = -\frac{M_0 K_{ac} G_0^2}{2M_0\omega_D/\omega_E}. \quad (11)$$

According to [3], the K_{ac} anisotropy constant linearly depends on temperature in rare-earth metal orthoferrites in the region of spontaneous transitions. The induced orientation phase transition field (11) should then also linearly depend on temperature. Such a dependence is indeed observed in experiments (Fig. 3). Unfortunately, because of the absence of exact experimental values for the quantities present in Eqs. (9)–(11), we cannot perform a more rigorous comparison of theory and experiment. To further simplify Eqs. (6) for precession spin oscillation frequencies, consider two limiting cases.

At $\omega_d \ll \omega_f$, solutions to Eqs. (6) have the form

$$\begin{aligned}\omega_1^2 &= \omega_d^2 + \omega_E\omega_{fd},\\ \omega_2^2 &= \omega_f^2 - \omega_E\omega_{fd},\end{aligned} \quad (12)$$

whereas at $\omega_d \gg \omega_f$, they become

$$\begin{aligned}\omega_1^2 &= \omega_d^2 + \frac{\omega_f^2\omega_E\omega_{fd}}{\omega_d^2},\\ \omega_2^2 &= \omega_f^2\left(1 - \frac{\omega_E\omega_{fd}}{\omega_d^2}\right).\end{aligned} \quad (13)$$

It follows from Eqs. (12) and (13) that the soft mode in samarium orthoferrite can be either a *d* subsystem (12) or an *f* subsystem (13) mode depending on the ratio between the vibrational frequencies of the *f* and *d* subsystems. According to [3], rare-earth metal modes are not resolved experimentally in rare-earth metal orthoferrites because they are strongly damped (*f* mode damping is on the order of magnitude of its frequency). It is therefore natural to suggest that the mode measured experimentally for samarium orthoferrite is a *d* subsystem mode, that is, the condition $\omega_d \ll \omega_f$ is fulfilled, and the soft mode in the region of the $\Gamma_4 \longleftrightarrow \Gamma_{42}$ orientation phase transition is a *d* subsystem mode. At the induced $\Gamma_4 \longleftrightarrow \Gamma_{42}$ phase transition point [the equality sign in Eq. (3)], the soft mode (12) frequency is then given by

$$\nu_d = [\omega_E(\omega_{me5} + \omega_{\rm dip}) + \omega_{H_{\rm tr}}^2(1-\eta)]^{1/2}/2\pi. \quad (14)$$

As not all samarium orthoferrite parameters were measured experimentally, let us approximately estimate the ω_E and $\omega_{\rm dip}$ frequencies from [3, 15]: $\omega_E \approx 2 \times 10^{14}$ s^{-1} and $\omega_{\rm dip} \approx 8 \times 10^7$ s^{-1}. The magnetoelastic frequency in rare-earth metal orthoferrites is usually several times or even an order of magnitude lower than the dipole frequency [3]. It follows that, at the spontaneous orientation phase transition point ($H = 0$), the soft mode frequency (14) of the *d* subsystem approximately equals $\nu_d = \sqrt{\omega_E\omega_{\rm dip}}/2\pi \approx 20$ GHz. This value is close to the $\nu_d \approx 27.7$ GHz experimental value (Fig. 4).

According to Eq. (14), the characteristic temperature dependence of the soft mode frequency shown in Fig. 2 can be explained by the temperature dependences of the $H_{\rm tr}$ induced orientation phase transition field and longitudinal susceptibility. We assume that longitudinal susceptibility weakly depends on temperature in the temperature range 400–450 K. The linear temperature dependence of frequency (14) should then be explained by a similar temperature dependence of transition field (11). It also follows from Eq. (14) that the soft mode frequency increases with increasing the phase transition field. This conclusion is also in agreement with the experimental dependence (Fig. 4).

Consider the behavior of the relaxation branch of quasi-spin oscillations in the region of an orientation phase transition. At the orientation phase transition point ($\omega_{S0} = 0$), the equation for the relaxation branch (7) frequency has the form

$$\begin{aligned}\omega_3^2 &= -i\Lambda_d\omega_E\\ &\times[\omega_E\omega_{me5} + \omega_L(\omega_H + \omega_D)(\eta\omega_H + \omega_D)/\omega_E]\\ &\times\{(1-\eta)[\omega_E(\omega_{me5} + \omega_{\rm dip}) + \omega_H^2(1-\eta)]\}^{-1}.\end{aligned} \quad (15)$$

It follows from Eq. (15) that the relaxation branch has activation determined by magnetoelastic and dipole contributions at the orientation phase transition point. Unlike the corresponding frequency in [1], frequency (15) does not vanish at the orientation phase transition point even in the absence of magnetoelastic coupling. Relaxation branch activation is then determined by coupling between magnetic subsystem oscillations and electromagnetic waves. At the spontaneous orientation phase transition point at $H = 0$, relaxation branch activation in the absence of magnetoelastic coupling is determined by the Dzyalochinski interaction.

As the ω_1, ω_2, and ω_3 branches (6), (7) are activation at the orientation phase transition point, quasi-elastic modes with $k \longrightarrow 0$ play the role of soft modes. This situation was described in detail in [4] for a two-sublattice antiferromagnet.

6. CONCLUSION

To summarize, we experimentally and theoretically proved that all characteristics of the spectrum of spin oscillations observed for samarium orthoferrite in the region of the $\Gamma_4(F_z, G_x) \longleftrightarrow \Gamma_{24}(F_{x,z}, G_{x,z})$ orientation phase transition induced by an external magnetic field can be explained by the ratio between the contributions of orthoferrite subsystem interactions and longitudinal susceptibility. We for the first time showed that the contribution of longitudinal susceptibility to the soft magnetoresonance mode gap in samarium orthoferrite can be substantial in low fields; that is, in the vicinity of sponta-

neous orientation phase transitions. This is explained by the occurrence of spontaneous orientation phase transitions in samarium orthoferrite at high temperatures, at which longitudinal susceptibility is commensurate with transverse susceptibility.

ACKNOWLEDGMENTS

The authors are deeply indebted to the referee for valuable comments.

This work was financially supported by the Ukrainian Foundation for Basic Research and the Ministry of Education of the Russian Federation (project no. 97-0-7.0-11).

REFERENCES

1. A. M. Balbashev, Yu. M. Gufan, N. Yu. Marchukov, and E. G. Rudashevskiĭ, Zh. Éksp. Teor. Fiz. **94** (4), 305 (1988) [Sov. Phys. JETP **67**, 821 (1988)].
2. K. P. Belov, A. K. Zvezdin, A. M. Kadomtseva, and R. Z. Levitin, *Orientational Transition in Rare-Earth Magnets* (Nauka, Moscow, 1977).
3. V. D. Buchel'nikov, N. K. Dan'shin, L. T. Tsymbal, and V. G. Shavrov, Usp. Fiz. Nauk **166**, 585 (1996) [Phys. Usp. **39**, 547 (1996)].
4. V. D. Buchel'nikov, N. K. Dan'shin, L. T. Tsymbal, and V. G. Shavrov, Usp. Fiz. Nauk **169**, 1049 (1999).
5. N. K. Dan'shin and Yu. I. Nepochatykh, Fiz. Nizk. Temp. **24**, 353 (1998) [Low Temp. Phys. **24**, 267 (1998)].
6. N. K. Dan'shin and G. G. Kramarchuk, Fiz. Nizk. Temp. **19**, 888 (1993) [Low Temp. Phys. **19**, 632 (1993)].
7. N. K. Dan'shin, Fiz. Nizk. Temp. **20**, 353 (1994) [Low Temp. Phys. **20**, 281 (1994)].
8. S. N. Barilo, A. P. Ges', A. M. Guretskiĭ, *et al.*, Fiz. Tverd. Tela (Leningrad) **33**, 621 (1991) [Sov. Phys. Solid State **33**, 354 (1991)].
9. É. Arutyunyan, K. N. Kocharyan, R. M. Martirosyan, *et al.*, Zh. Éksp. Teor. Fiz. **98**, 712 (1990) [Sov. Phys. JETP **71**, 398 (1990)].
10. N. K. Dan'shin, Yu. I. Nepochatykh, and V. F. Shkar', Zh. Éksp. Teor. Fiz. **109**, 639 (1996) [JETP **82**, 341 (1996)].
11. F. B. Hagedorn and E. M. Gyorgy, Phys. Rev. **174**, 540 (1968).
12. A. A. Mukhin and A. S. Prokhorov, Tr. Inst. Obshch. Fiz. Akad. Nauk SSSR **25**, 162 (1990).
13. V. D. Buchel'nikov and V. G. Shavrov, Pis'ma Zh. Éksp. Teor. Fiz. **60**, 534 (1994) [JETP Lett. **60**, 548 (1994)].
14. V. D. Buchel'nikov and V. G. Shavrov, Zh. Éksp. Teor. Fiz. **106**, 1756 (1994) [JETP **79**, 951 (1994)].
15. A. K. Zvezdin, V. M. Matveev, A. A. Mukhin, and A. I. Popov, *Rare-Earth Ions in Magnetially-Ordered Crystals* (Nauka, Moscow, 1985).

Translated by V. Sipachev

Journal of Experimental and Theoretical Physics, Vol. 92, No. 4, 2001, pp. 643–651.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 4, 2001, pp. 738–748.

SOLIDS
Structure

Soft-Potential Model and Homogeneous Width of Spectral Lines of Impurity Centers in Molecular Amorphous Media

Yu. G. Vainer*, M. A. Kol'chenko, and R. I. Personov
Institute of Spectroscopy, Russian Academy of Sciences, Troitsk, Moscow oblast, 142190 Russia
**e-mail: vainer@isan.troitsk.ru*
Received September 18, 2000

Abstract—A study is made to analyze the possibility of using the soft-potential model in optical investigations of disordered molecular systems with impurities. A procedure is suggested for calculating the temperature dependence of the homogeneous width of a phononless line in amorphous media with impurities within the soft-potential model. A calculation is performed of the temperature dependence of the width of a phononless line (optical dephasing) in an amorphous system of polymethyl methacrylate (PMMA) with an addition of tetra-tert-butylterrylene (TBT) using the parameters of this system known from the literature. Calculations are performed of the contributions to the width of a phononless line due to the interaction of an impurity with tunneling two-level systems, with thermally activated barrier crossings in double-well potentials, and with quasilocal modes of the matrix. The model calculation results are compared with the experimental data on the photon echo for TBT/PMMA, measured by us in the temperature range from 0.3 to 20 K. It is found that the soft-potential model describes qualitatively correctly the temperature behavior of the homogeneous width of a phononless line. In the temperature range of $T < 2$ K, where the main contribution to optical dephasing is associated with tunneling two-level systems, the predicted values of phononless line width agree well with the experimental data. At higher temperatures, some difference is observed between the prediction and experimental data, which may be due to the effect of impurity on the formation of quasilocal oscillation of the matrix. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The results of numerous investigations of amorphous materials for the last three decades have demonstrated that, at low temperatures, these materials exhibit a number of properties that differ considerably from the characteristics of analogous materials of ordered structure. As early as in 1971, Zeller and Pohl [1] have found that, at temperatures below 1 K, amorphous materials exhibit a linear temperature dependence of heat capacity and a quadratic temperature dependence of thermal conductivity, as distinct from a cubic dependence for both characteristics of the same materials in the crystalline state. Later, other low-temperature singularities of amorphous materials were found, distinguishing these materials from crystals. A number of theoretical models were suggested for describing the anomalous behavior of disordered condensed media at low temperatures. The widest acceptance was gained by a model based on the concept of "tunneling" two-level systems, suggested by Anderson *et al.* [2] and, independently, by Phillips [3]. Within this model, the properties of amorphous media at temperatures below several kelvins are due to the presence in such media (in addition to phonons) of characteristic low-energy excitations associated with transitions (tunneling) of groups of atoms or molecules between two local minima formed on the surface of the potential energy of disordered media. As distinct from phonons, which are collective excitations, the excitations in two-level systems are localized. The density of states of two-level systems as a function of their energy is almost constant and, at $T \leq 1$ K, considerably exceeds the density of phonon states. In spite of its simplicity, the model of tunneling two-level systems describes well most of the effects observed in experiments with amorphous materials at temperatures below several kelvins (see, for example, the monographs [4–6] and the references cited there).

Amorphous materials exhibit a number of universal anomalous properties which distinguish these materials from ordered substances at higher temperatures as well. Such properties include, for example, the additional (to acoustic phonons and two-level systems) contribution to the heat capacity at temperatures from several kelvins to several tens of kelvins [7], the presence of a plateau in the curve of temperature dependence of thermal conductivity in the region of 10 K [8], a linear decrease in the sound velocity with increasing temperature in the region above several kelvins [9], the presence of the so-called boson peak in the low-frequency Raman scattering [7], and a number of other effects. In so doing, the model of two-level systems proves inadequate for describing the properties listed above. This fact, as well as the results of investigations of inelastic scattering of neutrons in amorphous materials [10, 11], point to the existence of other low-energy excitations in

these materials, in addition to phonons and two-level systems.

In order to describe the anomalous properties of amorphous materials in a wider temperature range, the soft-potential model has been suggested and is used extensively at present (see the paper by Karpov *et al.* [12]). According to this model, quasilocal low-energy excitations of two more types exist in amorphous media, in addition to phonons and tunneling two-level systems. By analogy with two-level systems, these additional excitations are interpreted as motions of groups of atoms or molecules in local minima of potential surface. These are, firstly, relaxation systems in which (as distinct from two-level systems) thermally activated barrier crossings between two wells occur rather than simply tunnel transitions. Secondly, these are quasilocal low-frequency harmonic oscillators which, within the soft-potential model, are realized in single-well potentials. The density of states of harmonic oscillators is close to zero at low temperatures and increases with energy. According to the soft-potential model, it is the relaxation systems and harmonic oscillators that are responsible for the anomalous properties of amorphous materials at temperatures above several kelvins. The potentials, in which two-level and relaxation systems and harmonic oscillators are realized, are "soft" in the sense that external stresses transform them readily to one another. The soft-potential model reproduces the results of the standard model of tunneling two-level systems at low temperatures and, at the same time, describes well numerous phenomena in glasses at higher temperatures (up to tens of degrees and more) [13–16].

It is known that valuable information about the dynamics of condensed media may be obtained from the optical spectra of impurity atoms or molecules (impurity centers), minor amounts of which are introduced for this purpose into the material being studied. The dynamic processes in the matrix show up especially clearly in homogeneous broadening of phononless lines in the spectra of impurity centers. The investigation of the temperature dependence of the phononless line width may enable one, for example, to obtain important information about the mechanisms of optical dephasing in the system being studied. Such investigations of amorphous media became possible only as a result of development of selective spectroscopy techniques; these techniques help eliminate substantial inhomogeneous broadening of spectra and reveal narrow phononless lines. Numerous researchers used the method of burning stable spectral holes [17, 18] and photon echo [19, 20] to find that, at low temperatures (below 4–5 K), the measured width of phononless line in amorphous media was one–two orders of magnitude greater than that in crystals and exhibited a temperature dependence differing considerably from that in the case of crystals [21]. In amorphous media at low temperatures, this phononless line width usually varies by the law

$$\Gamma \propto T^{\alpha}, \text{ where } 1 \leq \alpha \leq 2,$$

while in the case of crystals at the same temperatures one must expect a power dependence with $\alpha = 7$. In order to interpret these facts in analyzing the spectra of impurity glasses and polymers, extensive use is made at present of a theory based on the model of tunneling two-level systems, which describes well the behavior of the line width at temperatures below several kelvins (see, for example, the review [22] and the references cited there). However, this model proves inadequate for describing the line broadening at higher temperatures (above 4–5 K). Therefore, in order to describe experimental data on line broadening at the above-identified temperatures, the interaction of an impurity center with some quasilocal oscillation at a frequency of 10–30 cm^{-1} is usually included in treatment [23, 24]. In so doing, the very existence and nature of this oscillation are not clear, as a rule, because this oscillation is not usually observed in spectra in the explicit form. At the same time, the soft-potential model naturally includes such excitations and yields concrete predictions of the density of their states. However, as far as we know, the soft-potential model was not used until very recently in optical investigations for describing the temperature broadening of a phononless line (the sole exception is the study [25]: see below). Therefore, it is of doubtless interest to check the validity of this model for describing the processes of optical dephasing in impurity amorphous materials.

Recently, we have used the photon echo method to obtain data on temperature broadening of a phononless line (optical dephasing) in the spectra of impurity molecular systems in a fairly wide (0.4 to 50 K) temperature range [26, 27]. In this paper, we will treat in more detail the experimental data on the temperature behavior of the width of a phononless line in one of impurity amorphous systems such as polymethyl methacrylate (PMMA) with an addition of tetra-tert-butylterrylene (TBT) and analyze these data using the soft-potential model. Our main objective will be to ascertain the possibility of using the soft-potential model to interpret the experimental data. The above-identified system has been selected because all of the parameters required for model calculations are known for this system. Because in this study we have analyzed the data obtained in a relatively wide temperature range, we have included in the treatment all mechanisms of dephasing, which, in our opinion, may show up in the temperature range being investigated. We have treated the contributions made by tunnel transitions and barrier crossings in double-well potentials, as well as the contribution made by quasilocal low-frequency harmonic oscillators in single-well potentials. For comparison with the experiment, we have calculated the temperature dependence of the phononless line width, proceed-

ing only from the parameters of the system being analyzed known from the literature (instead of approximating the experimental data by way of fitting of the respective parameters, as this was done, for example, in [25]). All of the quantities required for calculations were borrowed from the literature, where they were obtained from experimental data using other physical methods.

Before turning to treatment of the calculation results and their comparison with experimental data, we will treat briefly the salient points of the soft-potential model and derive some relations for the homogeneous width of phononless line within this model.

2. SALIENT POINTS OF THE SOFT-POTENTIAL MODEL

According to the soft-potential model [12] (see also [28] and the review [29]), additional (to ordinary phonons) quasilocal low-frequency excitations in amorphous media are described by the Hamiltonian of anharmonic oscillators

$$\hat{H} = -\frac{\hbar^2}{2M}\frac{d^2}{dx^2} + V(x) \tag{1}$$

with the effective mass M and potential energy in the form of the fourth-degree polynomial

$$V(x) = E_0\left\{\eta\left(\frac{x}{a}\right)^2 + \xi\left(\frac{x}{a}\right)^3 + \left(\frac{x}{a}\right)^4\right\}. \tag{2}$$

Here, E_0 is an energy of the order of the binding energy of glass-forming atoms (molecules), x is the generalized coordinate, and a is the characteristic length of the order of the interatomic distance. The values of the dimensionless parameters η and ξ are random, which is due to the structure inhomogeneity of the medium. It is assumed in the model that the distribution of the above-identified parameters has the form [30]

$$P(\eta, \xi) = |\eta| P_0(\eta, \xi); \tag{3}$$

thus, $|\eta|, |\xi| \ll 1$. The function $P_0(\eta, \xi)$ is usually assumed to be a constant. The characteristic scale of the dimensionless quantity η defining the height of the barrier between wells in potential (2) is described by the small parameter of the model,

$$\eta_L = (\hbar^2/2ma^2E_0)^{1/3} \approx 10^{-2},$$

where m is the average mass of glass-forming atoms. The scale of energies is defined by the quantity W, which characterizes the spectrum of levels in potential (2) at $\eta = \xi = 0$,

$$W = E_0\eta_L^2 = k_B T_c,$$

where k_B is the Boltzmann constant and T_c is the characteristic temperature which, for different materials, is in the range from 2 to 10 K.

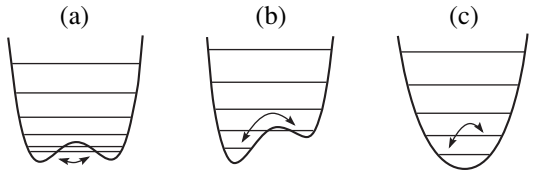


Fig. 1. A sketch of energy levels in a soft double-well potential given by Eq. (1) for different values of the parameters η and ξ, and the types of possible transitions: (a) region of two-level systems, with $\eta = -4\eta_L$ and $\xi = 0.2\sqrt{\eta_L}$; (b) region of relaxation systems, with $\eta = -4\eta_L$ and $\xi = 3\sqrt{\eta_L}$; (c) region of harmonic oscillators, with $\eta = 4\eta_L$ and $\xi = \sqrt{\eta_L}$.

The form of potential (2) is defined by the relative magnitude of the parameters η and ξ. It may be both double-well and single-well. If the asymmetry of a double-well potential is much less than the distance between levels in a single well, two lower levels in potential (2) form a two-level system (Fig. 1a) with the distance between the levels [31] of

$$E = \sqrt{J^2 + A^2}, \tag{4}$$

where

$$J \approx W\exp\left[-\frac{\sqrt{2}}{3}\left(\frac{|\eta|}{\eta_L}\right)^{3/2}\right]$$

is the tunnel splitting and

$$A = \frac{W}{\sqrt{2}}\frac{|\xi|}{\sqrt{\eta_L}}\left(\frac{|\eta|}{\eta_L}\right)^{3/2}$$

is the asymmetry (difference between the energies of two minima). The height V of the barrier between the wells is defined by the expression

$$V = \frac{W}{4}\left(\frac{\eta}{\eta_L}\right)^2. \tag{5}$$

According to the soft-potential model, the distribution function of the above-identified two-level systems with respect to the parameters E and p has the form [13]

$$P_{TLS}(E, p) = \left(\frac{2}{9}\right)^{1/3}\frac{P_0\eta_L^{5/2}}{W}\frac{1}{p\sqrt{1-p}}\ln^{-2/3}\frac{W}{E\sqrt{p}}, \tag{6}$$

where $p = (J/E)^2$.

It follows from the foregoing data that the concept of two-level systems in the soft-potential model is close to that in the standard model of tunneling two-level systems. For example, the distribution given by Eq. (6) differs from the respective distribution function adopted in the standard model (see, for example, [5]) by the logarithmic factor alone.

In the cases when the asymmetry of a double-well potential proves to exceed considerably the distance between levels in a single well (Fig. 1b), the main contribution to kinetic phenomena will be made by thermally activated barrier crossings between wells, i.e., relaxation systems. The distribution of relaxation systems with respect to asymmetry (which, in this case, coincides with energy, as follows from Eq. (4)) and barrier heights has the form [32]

$$P_{RS}(E, V) = \frac{P_0 \eta_L^{5/2}}{W^{5/4} V^{3/4}}. \quad (7)$$

The relaxation rate of relaxation systems is defined by the expression

$$R(V) = R_0 \exp(-V/k_B T), \quad (8)$$

where R_0 is a parameter of the order of 10^{12} to 10^{13} s^{-1} [33].

One can readily see that, with positive values of

$$\eta > \frac{9}{32}\xi^2,$$

potential (2) is a single-well one, and only harmonic oscillators may be excited in this potential. In so doing, if $\eta \gg \eta_L$, the anharmonicity is fairly low, and the oscillation is almost harmonic, with the distance between levels of

$$E = 2W\sqrt{\eta/\eta_L}$$

and the density of states $n(E) \propto E^4$ [30]. With energies above

$$E_b = 2.2W/A^{1/6},$$

where $A = 0.169(W/k_B T_g)$ and T_g is the vitrification temperature, the interaction between harmonic oscillators acquires considerable importance. In this case, quasilocal excitation can no longer be regarded as independent. New delocalized harmonic modes are characterized by the density of states $n(E) \propto E$ [16]. According to the soft-potential model, the rearrangement of the density of states of harmonic oscillators as a result of their interaction is responsible for the emergence of a boson peak in spectra of Raman scattering at frequencies $\omega \approx \omega_b = E_b/\hbar$. In accordance with [34], we will treat E_b as the upper limit of the energy of soft modes.

3. BASIC RELATIONS DESCRIBING OPTICAL DEPHASING IN THE SOFT-POTENTIAL MODEL

In order to perform a quantitative analysis of the temperature dependence of the times of optical dephasing using the soft-potential model, one must have appropriate expressions for the phononless line width, derived within this model. Until recently, however, no theoretical treatment of optical dephasing of electronic transitions of impurity centers in an amorphous matrix were performed within the soft-potential model. As far as we know, such treatment has been performed recently only by Garcia and Fernandez [25] (see also [35]). They introduced a model Hamiltonian for an impurity interacting with two-level systems and with harmonic oscillators, within the soft-potential model. Garcia and Fernandez derived expressions for the phononless line width of impurity in an amorphous matrix in two limiting cases, namely, for the interaction of impurity with two-level systems and with harmonic oscillators. In so doing, the expression which described broadening caused by two-level systems was derived in the so-called slow modulation limit. This approximation is valid when the time of experiment considerably exceeds the characteristic times of the processes being studied in an amorphous medium (for example, in experiments in the burning of stable spectral holes). In the case of photon echo used to obtain the experimental data analyzed below, the characteristic time of experiment is defined by the lifetime T_1 of excited state and by the time T_2' of pure optical dephasing of impurity molecules; this latter time is, as a rule, in a wide range of times of dynamic processes occurring in an amorphous medium. Therefore, in our case, it would be incorrect to employ the formula derived by Garcia and Fernandez [25] for the calculation of optical dephasing caused by two-level systems. In addition, it must be emphasized that Garcia and Fernandez [25] did not treat the contributions to broadening due to relaxation systems. For a qualitative analysis of experimental data within the soft-potential model, we performed a more general treatment and derived expressions for the temperature dependence of the times of dephasing caused both by tunnel transitions in two-level systems and by barrier crossings in relaxation systems. Our inference is based on using the formula describing the decay curve of photon echo in amorphous media within the standard model of two-level systems and stochastic model of spectral line width [36]. We assume that the analytical expression derived in the latter study, which allows only for the interaction with tunneling two-level systems, may be further used to describe the broadening caused by transitions in relaxation systems (with due regard for the difference in the form of distribution and in the relaxation rate between the two-level and relaxation systems), as was done by us.

According to the above-identified model [36], the signal intensity of a two-pulse photon echo as a function of delay τ is described by the expression [36]

$$I(\tau) = [\exp(-\tau/T_1)\exp(-\Phi(\tau))]^2, \quad (9)$$

where T_1 is the lifetime of excited state and

$$\Phi(\tau) = \frac{2\pi^3}{3} B\rho \times \left\langle \frac{A}{ER} \operatorname{sech}^2\left(\frac{E}{2k_B T}\right) F\left(\tanh\left(\frac{E}{2k_B T}\right), R\tau\right)\right\rangle. \quad (10)$$

In the latter expression, B is the constant of two-level system–chromophore interaction, ρ is the density of states of two-level systems, R is the rate of transitions in a double-well potential, and

$$F(\xi, x) = 2e^{-x} \times \int_0^x dx' I_0[\xi(x-x')]x'[I_0(x') + I_1(x')], \quad (11)$$

where $I_0(z)$ and $I_1(z)$ are modified Bessel functions of the first and second kind. The averaging in expression (10) indicated by angle brackets is performed by the parameters of double-well potentials.

Note that the decay curves of signals of photon echo, calculated by expression (9), generally speaking, are not exponential; therefore, the resultant decay curves give an ambiguous value of the time T_2 of optical dephasing. We determined the time of optical dephasing by way of approximating the calculated decay curves of photon echo signal by the closest exponential dependence with the effective time T_2, in the same manner as this was done by Naumov *et al.* [37, because this method is as close as possible to the procedure usually employed in calculating the time of optical dephasing by the experimentally obtained curves of photon-echo decay.

In the case of a single-phonon mechanism of relaxation of a two-level system, the relaxation rate is defined by the expression [38]

$$R(E) = CJ^2 E \coth(E/2k_BT),$$

where

$$C = \frac{1}{2\pi\rho'\hbar^4}\sum_\sigma \frac{\Lambda_\sigma^2}{v_\sigma^5}$$

has the meaning of the constant of two-level system–phonon interaction, ρ' is the density of the medium, and Λ_σ and v_σ denote the elastic constant and the rate of propagation of elastic vibrations in the medium for waves of different polarizations, respectively. We calculated the contribution to optical dephasing due to interaction between the impurity and two-level systems as was done by Geva and Skinner [36]; in our case, however, the averaging over the parameters of two-level systems was performed using distribution (6) predicted by the soft-potential model rather than within the standard model. One can use the foregoing distribution and relation (4) to reduce expression (10) to

$$\Phi^{TLS}(\tau) = \frac{2\pi^3}{3}B\int_{E_{\min}}^{E_{\max}} dE \times \int_{p_{\min}}^1 dp \frac{\sqrt{1-p}}{pR_{\max}(E)}P_{TLS}(E, p)\operatorname{sech}^2\left(\frac{E}{2k_BT}\right) \times F\left(\tanh\left(\frac{E}{2k_BT}\right), pR_{\max}(E)\tau\right). \quad (12)$$

Here,

$$R_{\max}(E) = CE^3\coth(E/2k_BT)$$

is the maximum relaxation rate of two-level systems, and

$$p = R/R_{\max}(E) = (J/E)^2.$$

As in [36], we assumed

$$E_{\min} = J_{\min} \approx 0, \quad E_{\max} \approx 20k_BT,$$
$$p_{\min} = (J_{\min}/E)^2 \approx 0.$$

We used for analysis experimental data obtained in a fairly wide temperature range; therefore, we investigated also the possible importance of the contribution made to dephasing by two-phonon transitions in two-level systems. The relaxation rate of two-level systems for two-phonon processes is given by [39]

$$R_{(2ph)} = C'\left(\frac{J}{E}\right)^2 T^6 EF(E, T),$$

where

$$C' = \frac{k_B^6}{\pi^3(\rho')^2\hbar^7}\sum_{\tau,\sigma}\frac{\Lambda_{\tau,\sigma}^2}{v_\tau^5 v_\sigma^5}$$

is the constant of two-level system–phonon interaction for two-phonon transitions and

$$E(E, T) = \frac{1}{140}\left(\left(\frac{E}{2k_BT}\right)^2 + \pi^2\right) \times \left(\left(\frac{E}{2k_BT}\right)^4 - \pi^2\left(\frac{E}{2k_BT}\right)^2 + \frac{10}{3}\pi^4\right)\coth\left(\frac{E}{2k_BT}\right).$$

In this case, expression (10) takes the form

$$\Phi^{2ph}(\tau) = \frac{2\pi^3}{3}B\int_{E_{\min}}^{E_{\max}} dE \times \int_{\rho_{\min}}^1 dp \frac{\sqrt{1-p}}{pR_{(2ph)\max}(E)}P_{TLS}(E, p)\operatorname{sech}^2\left(\frac{E}{2k_BT}\right) \times F\left(\tanh\left(\frac{E}{2k_BT}\right), pR_{(2ph)\max}(E)\tau\right), \quad (13)$$

where

$$R_{(2ph)\max}(E) = C'T^6EF(E,T)\coth\left(\frac{E}{2k_BT}\right).$$

We used the resultant expression to calculate the contribution made to dephasing by two-phonon processes.

We will dwell in more detail on the derivation of the expression for line broadening caused by interaction with relaxation systems, because, to the best of our knowledge, no calculations were previously performed of the contribution made to optical dephasing by barrier crossings in double-well potentials. In view of expression (8) for the relaxation rate of relaxation systems, as well as of the fact that, for relaxation systems, the asymmetry $A \approx E$, one can readily see that all parameters, with respect to which the averaging is performed in (10), are functions of V and E; therefore, the averaging in (10) may be represented as

$$\langle Q(E,V)\rangle = \int_{E_{\min}}^{E_{\max}} dE \int_0^{V_{\max}} dV Q(E,V)\frac{P_{RS}(E,V)}{\rho}, \quad (14)$$

where

$$Q(E,V) = \frac{A}{ER}\operatorname{sech}^2\left(\frac{E}{2k_BT}\right)F\left(\tanh\left(\frac{E}{2k_BT}\right), R\tau\right).$$

In this expression, according to [36], one can assume that

$$E_{\min} \approx 0, \quad E_{\max} \approx 20k_BT,$$

and $V_{\max}$ is the maximum barrier height which corresponds to the minimum observed relaxation rate of double-well potentials and is defined by the time of experiment (see Eq. (8)). In our case, $V_{\max} \approx 10k_BT$. The distribution $P_{RS}(E, V)$ is normalized to the density of states of relaxation systems.

As a result of averaging, expression (10) is reduced to

$$\Phi^{PC}(\tau) = \frac{2\pi^3}{3}B\int_{E_{\min}}^{E_{\max}} dE \int_0^{V_{\max}} dV \frac{1}{R_0\exp(-V/k_BT)} \times P_{RS}(E,V)\operatorname{sech}^2\left(\frac{E}{2k_BT}\right) \times F\left(\tanh\left(\frac{E}{2k_BT}\right), R_0\exp\left(-\frac{V}{k_BT}\right)\tau\right). \quad (15)$$

We used this expression to calculate the contribution made to the phononless line width by interaction with relaxation systems.

In order to calculate the contribution to the phononless line broadening caused by the interaction with harmonic oscillators which, unlike two-level and relaxation systems, are phonon-type excitations, we have used the analytical expression derived by Garcia and Fernandez [25],

$$\Delta\Gamma^{HO}(T) = \frac{P_0\eta_L^{5/2}}{3\sqrt{2}\pi\rho' W^4\hbar^3}V_c\sum_\sigma\frac{\Lambda_\sigma^2}{v_\sigma^5}(k_BT)^7 \times \int_0^{E_b/k_BT} dx\frac{x^6e^x}{(e^x-1)^2}, \quad (16)$$

where ρ' is the density of the medium and V_c is the average volume of localization of harmonic oscillator. It is interesting to note that this expression is similar to the formula derived by McCumber and Sturge [40] and describing, within the Debye model, the temperature broadening of phononless line in crystals, which is caused by acoustic phonons. However, within the latter model, the energy distribution of phonons has the form

$$n(E) \propto E^2,$$

which is different from our case. In addition, it is not E_b that appears in the upper limit of the integral in the above-mentioned formula, as in our case, but the Debye temperature T_D, which is usually much (approximately by an order of magnitude) higher.

4. RESULTS OF MODEL CALCULATIONS AND COMPARISON WITH EXPERIMENT

The calculations were performed using expressions (9), (12), (13), (15), and (16). The required values of parameters were borrowed from the literature and are listed in the table. Note that all of these values were found from experimental data obtained by other physical methods. In our calculations, we assumed the value of R_0 in expression (8) to be equal to the value of Debye frequency

$$\omega_D = \frac{k_BT_D}{\hbar}$$

for PMMA, borrowed from [41].

Figure 2 gives the results of calculations (performed within the soft-potential model) of the temperature dependence of the inverse time of dephasing $1/\pi T_2'$ (equal to the homogeneous width of phononless line less the radiation width $1/2\pi T_1$) for a TBT/PMMA system. Given in the same figure are the experimental data on photon echo obtained by us for this system. These data were obtained in a wide temperature range, which could be done thanks to the use of two varieties of the photon echo method, namely, picosecond two-pulse echo at low temperatures ($T < 4.2$ K) and incoherent photon echo at temperatures $T > 4.2$ K. A detailed description of the experimental facility, sample preparation techniques, and measurement procedure is given in [26, 27]. The experimental curve has a characteristic form with a slow increase in the line width at tempera-

tures below 2–3 K and its rapid rise at higher temperatures. This behavior of the temperature dependence points to the presence of at least two different mechanisms of dephasing. Usually, the low-temperature part of the curve is explained by the contribution made to dephasing by tunnel transitions in two-level systems, and the high-temperature part is attributed to dephasing caused by quasilocal low-frequency modes. One can see in the figure that a fairly good agreement between the prediction curve (curve *5*) and experimental data is observed at low temperatures ($T < 2$ K), while a marked disagreement is observed at high temperatures.

According to our calculation results, the main contribution to optical dephasing at temperatures $T < 1$ K is made by single-phonon transitions in two-level systems. At higher temperatures, the contribution by relaxation systems becomes appreciable. The behavior of the temperature dependence of this latter contribution to optical dephasing at temperatures $T > 1$ K is analogous to that of the contribution by two-level systems, but its value proves to be approximately three times less than the value of the latter. Thus, in the temperature range investigated by us, it is hard to distinguish between the contributions made to dephasing by two-level and relaxation systems because of the almost identical behavior of their temperature dependence curves. The calculation results indicate that, at temperatures $T > 2$–3 K, the temperature dependence of the phononless line width is largely defined by the contribution made by harmonic oscillators. One can further see in the figure that the contribution to optical dephasing by two-phonon transitions in two-level systems is insignificant in the entire temperature range being analyzed and may be ignored. Indeed, at temperatures $T < 1$ K, this contribution is more than three orders of magnitude less than the contribution by single-phonon transitions in two-level systems and, at higher temperatures ($T > 20$ K), though exceeding the respective contribution by single-phonon transitions, the contribution by two-phonon transitions is much less significant than the contribution of harmonic oscillators.

One can see in the figure that, at temperatures above W/k_B (which, in our case, is 2.5 K), when the contribution by quasilocal phonons starts to prevail, a marked disagreement is observed between the model calculation results (curve *5* in Fig. 2) and experimental data. We assume that a possible reason for such disagreement is the inaccurate value of the average volume of localization V_c of harmonic oscillator used in these calculations. According to Duval *et al.* [45], the average size of the localization region of harmonic oscillator depends on its frequency ω_0,

$$d = S\frac{v}{\omega_0 c}, \tag{17}$$

where S is the geometric factor and v is the sound velocity inside the soft mode localization region. The size of the localization region of the mode corresponding to the maximum of frequency distribution of harmonic oscillators in pure PMMA, obtained from the data on neutron scattering and low-frequency Raman scattering, is 39.5 Å [43]. It was this value that we used in estimating V_c in our calculations. In view of this, note that, as was demonstrated experimentally in [27], at $T > 2$–3 K, the properties of impurity may have a considerable effect on the temperature behavior of the homogeneous line width. For example, two systems including one and the same matrix (frozen ethanol) but different chromophores (resorufin and zinc tetraphenyl porphin) exhibited a considerable difference in the values of effective frequency of quasilocal oscillation (29 and 12 cm^{-1}, respectively) [27]. Therefore, one can assume that the introduction of an impurity into the matrix may cause the parameters of quasilocal oscillation in the neighborhood of the impurity molecule to vary depending on the properties of this impurity. As a result, the effective value of the average volume of the localization region of harmonic oscillator in a doped system may be less than that of the respective volume for an analogous system without impurity. For example, as is shown in Fig. 2, an adequate agreement between the model curve and the respective experimental curve may be attained if the average size of the localization region of harmonic oscillator is taken to be 19 Å (curve *6*).

Note that in our analysis we ignored the contribution to dephasing made by acoustic phonons, because the density of states of such phonons in the temperature range being analyzed (which lies below the Debye temperature) must be less than the density of states of

Parameters of PMMA matrix and TBT/PMMA system

Parameters			Method of measurement
T_1	3.5 ns	[26]	Method of induced lattices
ρ	1180 kg/m^3	[42]	
Λ_t	0.11 eV	[42]	Investigation of ultrasound propagation
Λ_1	0.15 eV		
v_t	1570 m/s		
v_1	3150 m/s		
T_d	256 K	[41]	Measurement of specific heat
E_b	13.4 cm^{-1}	[43]	Raman scattering and inelastic neutron scattering
V_c	6.2×10^4 Å^3		
W/k_B	2.5 K	[34]	Measurement of specific heat and ultrasound propagation
$P_0\eta_L^{5/2}$	1.27×10^{23} m^{-3}		
B*	1.3×10^{-16} K^{-1} s^{-1}	[44]	Single-molecule spectroscopy

*This quantity is the constant of two-level system–chromophore interaction. For the lack of data, we used its value as the constant of relaxation system–chromophore interaction.

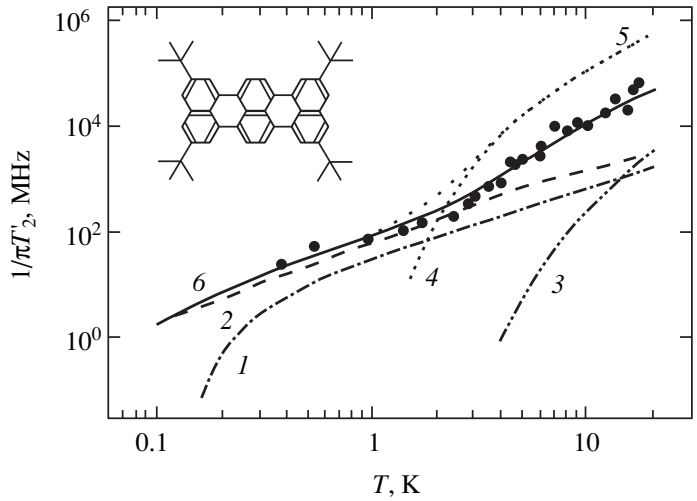


Fig. 2. The results of calculations of the temperature dependence of the inverse time of optical dephasing $1/\pi T_2'$ and the experimental data on photon echo (circles) for a TBT/PMMA system: (*1*) contribution by relaxation systems, (*2*) contribution by single-phonon transitions in two-level systems, (*3*) contribution by two-phonon transitions in two-level systems, (*4*) contribution by harmonic oscillators at $V_c = 6.2 \times 10^4$ Å^3, (*5*) total calculated temperature dependence at $V_c = 6.2 \times 10^4$ Å^3, (6) total dependence at $V_c = 6.9 \times 10^3$ Å^3. The inset gives the structure formula of TBT molecule.

quasilocal oscillation. The results of Gladenkova and Osadko [46] may be regarded as an indirect proof of the validity of this assumption. They used the results of theoretical analysis of spectral line broadening for Eu^{3+} ion in silicate glass to demonstrate that the phonon line broadening due to interaction with acoustic phonons showed up only upon approaching the Debye temperature.

5. CONCLUSION

Calculations performed within the soft-potential model and a comparison of their results with experimental data obtained for a TBT/PMMA system lead one to the following conclusions.

1. The soft-potential model describes qualitatively correctly the temperature behavior of the homogeneous width of a phononless line in a relatively wide temperature range and may be used to advantage in spectral investigations.

2. The results of calculations of the contribution made by relaxation systems to the phononless line width, performed for the first time ever, have demonstrated that the temperature dependence of this contribution is analogous to the dependence on the contribution of tunneling two-level systems (except for the range of $T < 0.3$–0.4 K). The absolute magnitude of the contribution made by relaxation systems is approximately 30% of the contribution made by two-level systems.

3. In the low-temperature region, where the main contribution to optical dephasing is defined by the interaction between an impurity and tunneling two-level systems, the results of calculating the temperature dependence of the phononless line width (in performing this calculation, we employed the soft-potential model and the stochastic model of line width, without using any fitting parameters) are in good agreement with the experimental data.

4. In the high-temperature region, where the main contribution to the phononless line width is made by interaction with harmonic oscillators, the prediction data proved to be appreciably overstated compared with the experimental data. The part played by impurity in the formation of harmonic oscillators is cited as a possible reason for this overstatement.

ACKNOWLEDGMENTS

This study received financial support from the Russian Foundation for Basic Research (project no. 99-02-17927).

REFERENCES

1. R. C. Zeller and R. O. Pohl, Phys. Rev. B **4**, 2029 (1971).
2. P. W. Anderson, B. I. Halperin, and C. M. Varma, Philos. Mag. **2**, 1 (1972).
3. W. A. Phillips, J. Low Temp. Phys. **7**, 351 (1972).
4. *Amorphous Solids: Low Temperature Properties*, Ed. by W. A. Phillips (Springer-Verlag, Berlin, 1981).
5. S. Hunklinger and A. K. Raychaudhuri, in *Progress in Low Temperature Physics*, Ed. by D. F. Brewer (Elseiver, Amsterdam, 1986), Vol. 9.
6. *Tunneling Systems in Amorphous and Crystalline Solids*, Ed. by P. Esquinazi (Springer-Verlag, Berlin, 1998).
7. R. H. Stolen, Phys. Chem. Glasses **11**, 83 (1970).
8. J. J. Freeman and A. C. Anderson, Phys. Rev. B **34**, 5684 (1986).
9. G. Bellessa, Phys. Rev. Lett. **40**, 1456 (1978).

10. U. Buchenau, M. Prager, N. Nucker, *et al.*, Phys. Rev. B **34**, 5665 (1986).
11. U. Buchenau, H. M. Zhou, N. Nucker, *et al.*, Phys. Rev. Lett. **60**, 1318 (1988).
12. V. G. Karpov, M. I. Klinger, and F. N. Ignat'ev, Zh. Éksp. Teor. Fiz. **84**, 760 (1983) [Sov. Phys. JETP **57**, 439 (1983)].
13. U. Buchenau, Yu. M. Galperin, V. L. Gurevich, and H. R. Schober, Phys. Rev. B **43**, 5039 (1991).
14. U. Buchenau, Yu. M. Galperin, V. L. Gurevich, *et al.*, Phys. Rev. B **46**, 2798 (1992).
15. D. A. Parshin, Phys. Rev. B **49**, 9400 (1994).
16. V. L. Gurevich, D. A. Parshin, J. Pelous, and H. R. Schober, Phys. Rev. B **48**, 16318 (1993).
17. B. M. Kharlamov, R. I. Personov, and L. A. Bykovskaya, Opt. Commun. **12**, 191 (1974).
18. A. A. Gorokhovskiĭ, R. K. Kaarli, and L. A. Rebane, Pis'ma Zh. Éksp. Teor. Fiz. **20**, 474 (1974) [JETP Lett. **20**, 216 (1974)].
19. N. A. Kurnit, I. D. Abella, and S. R. Hartmann, Phys. Rev. Lett. **13**, 567 (1964).
20. T. J. Aartsma and D. A. Wiersma, Phys. Rev. Lett. **36**, 1360 (1976).
21. S. Voelker, Ann. Rev. Phys. Chem. **40**, 499 (1989).
22. R. Jankowiak, J. M. Hayes, and G. J. Small, Chem. Rev. **93**, 1471 (1993).
23. B. Jackson and R. Silbey, Chem. Phys. Lett. **99**, 331 (1983).
24. G. Schulte, W. Grond, D. Haarer, and R. Silbey, J. Phys. Chem. **88**, 679 (1988).
25. A. J. García and J. Fernández, Phys. Rev. B **56**, 579 (1997).
26. S. J. Zilker, L. Kador, J. Friebel, *et al.*, J. Chem. Phys. **109**, 6780 (1998).
27. Yu. G. Vainer, M. A. Kol'chenko, A. V. Naumov, *et al.*, J. Lumin. **86**, 265 (2000).
28. D. A. Parshin, Phys. Scr. **49**, 180 (1993).
29. D. A. Parshin, Fiz. Tverd. Tela (St. Petersburg) **36**, 1809 (1994) [Phys. Solid State **36**, 991 (1994)].
30. M. A. Il'in, V. G. Karpov, and D. A. Parshin, Zh. Éksp. Teor. Fiz. **92**, 291 (1987) [Sov. Phys. JETP **65**, 165 (1987)].
31. V. G. Karpov and D. A. Parshin, Zh. Éksp. Teor. Fiz. **88**, 2212 (1985) [Sov. Phys. JETP **61**, 1308 (1985)].
32. D. A. Parshin and S. Sahling, Phys. Rev. B **47**, 5677 (1993).
33. W. Kohler, J. Meiler, and J. Friedrich, Phys. Rev. B **35**, 4031 (1987).
34. M. A. Ramos, L. Gil, A. Bringer, and U. Buchenau, Phys. Status Solidi A **135**, 477 (1993).
35. A. J. García, R. Balda, and J. Fernández, J. Lumin. **83–84**, 417 (1999).
36. E. Geva and J. L. Skinner, J. Chem. Phys. **107**, 7630 (1997).
37. A. V. Naumov, Yu. G. Vainer, and S. J. Zilker, J. Lumin. **86**, 273 (2000).
38. J. Jäckle, Z. Phys. **257**, 212 (1972).
39. P. Doussineau, C. Frenois, R. G. Leisure, *et al.*, J. Phys. (Paris) **41**, 1193 (1980).
40. D. E. McCumber and M. D. Sturge, J. Appl. Phys. **34**, 1682 (1963).
41. R. O. Pohl, in *Amorphous Solids: Low Temperature Properties*, Ed. by W. A. Phillips (Springer-Verlag, Berlin, 1981).
42. J. F. Berret and M. Meißner, Z. Phys. B **70**, 65 (1988).
43. A. Mermet, N. V. Surovtsev, E. Duval, *et al.*, Europhys. Lett. **36**, 277 (1996).
44. E. Geva and J. L. Skinner, J. Chem. Phys. **108**, 8485 (1998).
45. E. Duval, N. García, A. Boukenter, and J. Serughetti, J. Chem. Phys. **99**, 2040 (1993).
46. S. N. Gladenkova and I. S. Osadko, Chem. Phys. Lett. **187**, 628 (1991).

Translated by H. Bronstein

Journal of Experimental and Theoretical Physics, Vol. 92, No. 4, 2001, pp. 652–664.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 4, 2001, pp. 749–762.

SOLIDS
Electronic Properties

The Theory of Tunneling in Normal Metal/*d*-Type Superconductor 2D-Structures

I. A. Devyatov*, D. V. Goncharov, and M. Yu. Kupriyanov
Research Institute of Nuclear Physics, Moscow State University, Moscow, 119899 Russia
**e-mail: idev@pn.sinp.msu.ru*
Received July 27, 2000

Abstract—A sequential theoretical analysis is performed of tunneling in normal metal/*d*-type superconductor structures which contain scattering centers in the interlayer between a normal metal and a superconductor. As a result, it is demonstrated that the presence of a scattering center inside an insulator interlayer leads to partial suppression of previously predicted anomalously high values of conductance in the low-voltage region (zero bias anomaly (ZBA)). In so doing, the inclusion of the "interference" term in the current operator (interference of tunneling through a scattering center with direct potential tunneling) results in the suppression of ZBA. The predicted effect is virtually independent both of the position of the scattering center in the interlayer and of the shape of the resonance curve of scattering (which is Lorentzian in the case of resonance tunneling through the scattering center). © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The body of currently available experimental data present convincing proof of the existence of the *d*-symmetry of the order parameter in high-temperature superconductors (HTSC) [1–4]. With this symmetry, it is assumed that the sign of the order parameter Δ and, consequently, of the quasiclassical Green's functions describing the behavior of quasiparticles depends on the direction of motion of these quasiparticles in the *ab* plane. In particular, when quasiparticles moving along the *a* axis scatter in the direction parallel to the crystallographic axis *b*, the sign reversal must occur. With the angle α between the normal to the HTSC boundary and the crystallographic direction *a* being other than zero, the scattering of quasiparticles by the structure boundaries may be accompanied by just this sign reversal. This leads automatically to several effects, namely, the suppression of the order parameter in the neighborhood of the boundary [5], the formation of a bound electron-hole state with the energy $\varepsilon = 0$ caused by the change of sign of the order parameter [6, 9] and of Andreev states with $\varepsilon < |\Delta|$ [10], and the generation of an isotropic gapless superconducting state of the *s*-type in the presence of diffuse scattering of quasiparticles by the boundary [11].

Such unusual behavior of HTSC must result in a number of singularities on the current-voltage characteristics of both Josephson junctions and a normal metal/insulator/*d*-type superconductor (*N–I–D*) structures. In particular, in the latter case in a model with a δ-functional barrier, there was theoretically proven the existence of anomalies of conductance in the low-voltage region caused by the presence of bound state with $\varepsilon = 0$ (zero bias anomaly (ZBA)) [6–9], as well as of a number of less pronounced singularities caused by Andreev bound states with energies $\varepsilon < |\Delta|$ [10, 11].

The ZBA was observed experimentally in bicrystal substrate junctions [12]. However, all attempts at detecting these singularities in *N–I–D* and *D–I–D* structures of practical importance with an interlayer of metal-oxide materials of semiconductor-type conductance ($Pr_1Ba_2Cu_3O_7$ praseodymium-barium-copper ceramic, $SrTiO_3$ strontium-titanium ceramic, and $La_1Sr_2Ti_3O_7$ lanthanum-strontium-titanium ceramic) have failed. Based on the experimental data of [13–18] obtained for these junctions, one can only state with confidence that the main channel of transport of the normal component of current in these junctions is the resonance tunneling through localized states in the interlayer (see also the review [19]).

Previous theoretical investigations of the processes of resonance tunneling in structures with normal electrodes involved studies within a one-dimensional approximation [20] and studies in view of the three-dimensional behavior of scattering [21, 22]. In so doing, it was demonstrated that the process of resonance tunneling proper effectively involved only a part of the localized states arranged in the middle of a barrier in a layer with a thickness of the order of the effective radius of the localized state. The position of the energy level of such an "effective" localized state might differ from the Fermi energy by a value of the order of this level's half-width [20]. The inclusion of the three-dimensional behavior of tunneling [21, 22] only modified the pre-exponential factors in the correlations between the structure conductance and the barrier thickness. This was due to the fact that the tunneling of quasiparticles occurred predominantly in a narrow cone of angles in the neighborhood of the normal to the boundaries and did not lead to any qualitative changes. The inclusion of the *s*-type superconductivity in one of

the electrodes of the structure [23] did not cause qualitative changes in the pattern of the process, although it resulted in the emergence of a singularity in the averaged differential conductance at $eU \longrightarrow 0$ (U is the voltage across the junction). An attempt at treating the resonant transport in an *S–I–S* structure within the phenomenological 1D-model of resonant coupling between the electrodes was made in [24, 25].

In this paper, we will demonstrate that one cannot restrict oneself to a one-dimensional approximation in the presence of *d*-pairing in the electrodes. Qualitatively, this is associated with the fact that direct tunneling in *N–I–D* structures is likewise a resonant process which leads to the formation of ZBA, but with a singularity half-width much smaller than that obtained in the model with δ-functional barrier. In so doing, the interference of two processes occurring in the low-voltage region brings a partial suppression of ZBA and a sharp enhancement of the process of tunneling via localized states, which proves to involve almost all localized states, irrespective of their location in the interlayer.

2. MODEL OF JUNCTION

We will assume that the tunneling barrier $V(\mathbf{r})$ in the *N–I–D* structure being investigated is the sum of potentials

$$V(\mathbf{r}) = V_{rect} + V_{imp}, \tag{1}$$

in which the first term simulates a two-dimensional square barrier of height V_0 and thickness $2d$,

$$V_{rect}(x) = V_0\theta(|x| - d) \tag{2}$$

(the coordinates are reckoned from the middle of the barrier), and the second term describes the defect

$$V_{imp}(\mathbf{r}) = \begin{cases} -\beta, & |\mathbf{r} - \mathbf{r}_0| \le \rho, \\ 0, & |\mathbf{r} - \mathbf{r}_0| > \rho \end{cases} \tag{3}$$

present at the point $\mathbf{r}_0 = (x_0, y_0)$. Here, $\rho \ll k_0^{-1}$ is the radius defect, and $\hbar k_0$ is the Fermi momentum.

Potential (3) violates the spatial inhomogeneity of the structure; i.e., it brings about the nonconservation of the quasiparticle momentum component (which is parallel to the barrier) in the process of the tunneling of quasiparticles. At $\beta > 0$, potential (3) describes the resonance tunneling (see Appendix A). The negative values of β correspond to the direct, nonresonance, scattering. We will further assume that the density of localized states is low, so that their interference is insignificant, and the barrier thickness $2d$ is relatively great,

$$\kappa_0 d \gg 1. \tag{4}$$

Here, $\hbar\kappa_0 = \sqrt{2m(V_0 - E_f)}$ is the momentum, m is the electron mass, and E_f is the Fermi energy. The validity of condition (4) is essential for effective localization of the quasiparticle wave function on a localized state and agrees with the available experimental data [13–19].

In calculating the transport properties of the *N–I–D* structure, we will assume that the current passing through this structure does not bring about the removal of the HTSC electrode from the state of thermodynamic equilibrium. This condition is automatically valid in *N–I–D* junctions with a wide potential barrier of low penetrability, $D \ll 1$, even in the presence in the barrier of localized states spaced from one another at distances that considerably exceed their effective "transverse" radius $l_\perp = d/\sqrt{\kappa_0 d}$. In the absence of localized states, the corrections to conductance associated with disequilibrium arise only in the second order with respect to penetrability. The presence of localized states leads to the formation in the barrier of spatially narrow tunneling channels with a conductance on the order of quantum conductance and with transverse dimensions of the order of $l_\perp$. With the average distance between localized states exceeding $l_\perp$, the interference of the currents injected via these channels into the HTSC electrode is insignificant, and the nonequilibrium effects are low by virtue of the geometric factor. The foregoing leads one to assume that the voltage U applied to the junction is fixed and decreases only on the potential barrier of the structure.

3. CURRENT TRANSPORT

3.1. General Expression for Current

Within the foregoing assumptions, it is natural to describe the current transport in an *N–I–D* structure in terms of electron-like and holelike excitations [26]. The general expression for current has the form

$$I = \frac{e}{\pi\hbar}\int d\varepsilon\{\varepsilon I_e f(\varepsilon - eU) + I_h(1 - f(-\varepsilon - eU)) + (I_{el} + I_{hl}) f(\varepsilon)\}. \tag{5}$$

The first and second terms in Eq. (5) represent electron currents generated by electrons and holes of normal metal, respectively. Their Fermi distribution is shifted by eU relative to the Fermi distribution $f(\varepsilon)$ of the superconductor. The third term in braces in Eq. (5) is the electron current generated by electron-like and holelike excitations of the superconductor. The structure of the formula for current (5) coincides with the analogous expression of the BTK theory [26]. The only distinction is that, in our case, it is more convenient to calculate the current after the scatterer, on the interface with the superconductor, while in the BTK theory the current was calculated before the scatterer, on the interface with a normal metal.

The current components appearing in expression (5),

$$I_i = I_i^{\text{pot}} + I_i^{\text{int}} + I_i^{\text{res}}, \quad i = e, h, el, hl, \tag{6}$$

$$\frac{I_i^{\text{pot}}}{L_y} = \frac{1}{2\pi}\int dk_y^0\{\left|C_{i\to}^{\text{pot}}(\mathbf{k},\mathbf{k}_0)\right|^2 - \left|C_{i\leftarrow}^{\text{pot}}(\mathbf{k},\mathbf{k}_0)\right|^2\}, \quad (7)$$

$$I_i^{\text{int}} = \frac{1}{\pi}\int dk_y^0 \text{Re}\{C_{i\to}^{\text{pot}}(\mathbf{k},\mathbf{k}_0)(C_{i\to}^{\text{res}}(\mathbf{k},\mathbf{k}_0))^* - C_{i\leftarrow}^{\text{pot}}(\mathbf{k},\mathbf{k}_0)(C_{i\leftarrow}^{\text{res}}(\mathbf{k},\mathbf{k}_0))^*\}, \quad (8)$$

$$I_i^{\text{res}} = \frac{1}{(2\pi)^2}\int\int dk_y dk_y^0 \frac{k_x}{k_x^0} \times \{\left|C_{i\to}^{\text{res}}(\mathbf{k},\mathbf{k}_0)\right|^2 - \left|C_{i\leftarrow}^{\text{res}}(\mathbf{k},\mathbf{k}_0)\right|^2\} \quad (9)$$

are related by the ordinary quantum-mechanical expression

$$\psi_i(\mathbf{r}) = \frac{1}{2\pi}\int dk_y \exp(ik_y y) \times \{\exp(ik_x x)C_{i\to}(\mathbf{k},\mathbf{k}_0) + \exp(-ik_x x)C_{i\leftarrow}(\mathbf{k},\mathbf{k}_0)\} \quad (10)$$

to the Fourier components of scattered electron wave,

$$C_{i\leftrightarrow}(\mathbf{k},\mathbf{k}_0) = 2\pi C_{i\leftrightarrow}^{\text{pot}}(\mathbf{k},\mathbf{k}_0)\delta(k_y - k_y^0) + C_{i\leftrightarrow}^{\text{res}}(\mathbf{k},\mathbf{k}_0), \quad (11)$$

which describe its propagation in a "forward" direction to the superconductor ($C_{i\to}(\mathbf{k},\mathbf{k}_0)$) and in the opposite direction ($C_{i\leftarrow}(\mathbf{k},\mathbf{k}_0)$). In formulas (6)–(10), $k_x^0 = \sqrt{k_0^2 - k_y^{02}}$ and $k_x = \sqrt{k_0^2 - k_y^2}$ are the wave vector components of the initial and scattered (see the next section) electron waves, respectively, which are perpendicular to the *I–D* interface; and L_y in Eq. (7) is the barrier width in the direction perpendicular to the normal to the interfaces. Expressions (8) and (9) preassign the value of current in terms of a single defect. The coefficient of the delta function in Eq. (11) describes a potentially scattered wave, i.e., characterizes the process of tunneling of quasiparticles through a potential barrier containing no localized states. The second term in Eq. (11) describes a resonantly scattered wave and corresponds to resonance tunneling. Note that the term "resonance" for electron wave $C_{i\leftrightarrow}^{\text{res}}(\mathbf{k},\mathbf{k}_0)$ in Eq. (11) and, especially, the terms "interference" and "resonance" for the current components (8) and (9) are fairly arbitrary in our case of a complex structure with a scatterer in the barrier and Andreev reflection on the *S–N* interface (see the discussions in Section 4).

It follows from formulas (5)–(11) that the problem of calculating the current through an *N–I–D* structure in fact reduces to finding the Fourier components of scattered electron waves $C_{i\leftrightarrow}(\mathbf{k},\mathbf{k}_0)$.

3.2. Recurrent Relations for Fourier Components of Electron Waves

In order to calculate the Fourier components of a scattered electron wave $C_{i\leftrightarrow}(\mathbf{k},\mathbf{k}_0)$, it is convenient to use the dynamic model of successive normal and Andreev reflections on the *I–D* interface. This approach was previously used by Belogolovskiĭ *et al.* [27] in studying the processes occurring in *N–I–N–I–S* junctions with conservation of the quasiparticle momentum component parallel to the interface plane. Below, we will generalize this method to the case when such a conservation law is no longer valid because of scattering from localized states in the barrier.

We will treat the incidence of a plane electron wave with the wave vector $\mathbf{k}_0 = (k_x^0, k_y^0)$ and the energy ε from normal metal onto the scatterer. The wave transmitted to the superconductor may be represented as (see Appendix A)

$$\psi_{e\to}^0(\mathbf{r}) = \frac{1}{2\pi}\int dk_y \exp(ik_y y + ik_x x)C_{e\to}^0(\mathbf{k},\mathbf{k}_0), \quad (12)$$

$$C_{e\to}^0 = 2\pi D_0(k_y)\delta(k_y - k_y^0) + C_{e\,\text{res}}^0(\mathbf{k},\mathbf{k}_0), \quad (13)$$

$$C_{e\,\text{res}}^0 = L_e d_\to^e(k_y^0, x_0)\exp(ik_y^0 y_0)\tilde{G}_e(k_y, x_0). \quad (14)$$

The first term in Eq. (13) describes the process of potential scattering by a homogeneous square barrier (2) containing no localized states. The transmission coefficient $D_0(k_y)$ in Eq. (13) is described by the well-known expression

$$D_0(k_y) = -\frac{4ik_x\kappa_x\exp(-2id k_x)}{(\kappa_x - ik_x)^2\exp(2d\kappa_x) - (\kappa_x + ik_x)^2\exp(-2d\kappa_x)}, \quad \kappa_x = \sqrt{\kappa_0^2 + k_y^2}. \quad (15)$$

The second term in Eq. (13) describes the scattering by a defect. The scattering potential violates the spatial homogeneity of the structure. Therefore, in the general case, the scattering process is accompanied by a variation of the transverse electron component. In so doing, the resonance scattering amplitude L_e at $\beta > 0$ in Eq. (3) has the form (see Appendix A)

$$L_e = \frac{2\pi\hbar}{m}\frac{V_0 - E_f}{\varepsilon - \varepsilon_R + i\Gamma^{2D}}, \quad (16)$$

where

$$\Gamma^{2D} = \frac{\Gamma_l^{2D} + \Gamma_r^{2D}}{2}, \quad \Gamma_{l,r}^{2D} = 4\sqrt{\pi}(V_0 - E_f)\frac{k_0\kappa_0}{k_0^2 + \kappa_0^2}\frac{\exp[-2\kappa_0(d \mp x_0)]}{\sqrt{\kappa_0(d \mp x_0)}} \quad (17)$$

is the width of electronic level on a localized state in the 2D case (see Appendix A). The quantity $d_\to^0(k_y^0, x_0)$ in

Eq. (14) is the amplitude of a single electron wave with the wave vector $\mathbf{k}_0$ that propagates from a normal metal under square barrier (2) to a point with the coordinate $\mathbf{r}_0$, at which the localized state is located,

$$d^e_{\rightarrow}(k_y^0, x_0) = \frac{-2ik_x^0\{(\kappa_x^0 + ik_x^0)\exp[-\kappa_x^0(d-x_0)] + (\kappa_x^0 - ik_x^0)\exp[\kappa_x^0(d-x_0)]\}}{(\kappa_x^0 - ik_x^0)^2\exp((2d\kappa_x^0) - (\kappa_x^0 + ik_x^0)^2\exp(-2d\kappa_x^0))}\exp(-ik_x^0 d), \quad (18)$$

and $\tilde{G}_e(k_y, x_0)$ is the Fourier transform of Green's electron function $G_e(\mathbf{r}, \mathbf{r}_0)$ of barrier (2) with the source at point $\mathbf{r}_0$,

$$G_e(\mathbf{r}, \mathbf{r}_0) = \frac{1}{2\pi}\int dk_y \exp[ik_y(y-y_0)]G_e(k_y, x, x_0) = \frac{1}{2\pi}dk_y \exp(ik_y y + ik_x x)\tilde{G}_e(k_y, x_0), \quad (19)$$

$$G_e(k_y, x, x_0) = \frac{2m}{\hbar^2}\frac{\exp[ik_x(x-d)]\{(\kappa_x + ik_x)\exp[-\kappa_x(d+x_0)] + (\kappa_x - ik_x)\exp[\kappa_x(d+x_0)]\}}{(\kappa_x + ik_x)^2\exp(-2d\kappa_x) - (\kappa_x - ik_x)^2\exp(2d\kappa_x)}, \quad (20)$$

$$\tilde{G}_e(k_y, x_0) = G_e(k_y, x, x_0)\exp(-ik_x x)\exp(-ik_y y_0). \quad (21)$$

The electron wave described by Fourier expansion (12) and transmitted through the scatterer will produce, component by component, Andreev reflections in the superconductor [28, 29] and generate a hole wave with the group velocity directed toward the normal metal,

$$\psi^0_{h\leftarrow}(\mathbf{r}) = \frac{1}{2\pi}\int dk_y \exp[i(k_y y + k_x x)]C^0_{h\leftarrow}(\mathbf{k}), \quad (22)$$

$$C^0_{h\leftarrow}(k_x, k_y) = a_e(k_x, k_y)C^0_{e\leftarrow}(k_x, k_y), \quad (23)$$

where $a_e(\mathbf{k})$ denotes the coefficients of Andreev reflection.

The hole wave described by Fourier expansion (22) is reflected potentially and resonantly from the scatterer to generate a hole wave propagating in the direction of superconductor,

$$\psi^1_{h\rightarrow}(\mathbf{r}) = \frac{1}{2\pi}\int dk_y \exp[i(k_y y - k_x x)]C^1_{h\rightarrow}(\mathbf{k}), \quad (24)$$

$$C^1_{h\rightarrow} = r_{h\leftarrow}(k_y)C^0_{h\rightarrow} + L_h d^0_{h\leftarrow}(\mathbf{r}_0)\tilde{G}_h(k_y). \quad (25)$$

The first term in Eq. (25) describes the potential reflection of the hole wave (propagating on the superconductor side) from square potential (2) with the reflection coefficient

$$r_{h\leftarrow}(k_y) = \frac{\exp(2idk_x)(k_x^2 + \kappa_x^2)[\exp(2d\kappa_x) - \exp(-2d\kappa_x)]}{(\kappa_x + ik_x)^2\exp(2d\kappa_x) - (\kappa_x - ik_x)^2\exp(-2d\kappa_x)}. \quad (26)$$

The second term in Eq. (25) describes the resonance reflection from the total potential (1), with the amplitude L_h of resonance hole scattering given by

$$L_h = -\frac{2\pi}{m}\frac{V_0 - E_f}{\varepsilon + \varepsilon_R + i\Gamma^{2D}} = L_e^*(-\varepsilon). \quad (27)$$

The quantity $d_{h\leftarrow}(\mathbf{r}_0)$ in formula (25) is the amplitude of the hole wave (22) that reached the localized state under potential (2),

$$d^0_{h\leftarrow}(\mathbf{r}_0) = \frac{1}{2\pi} \times \int dk_y d_{h\leftarrow}(k_y, x_0)\exp(ik_y y_0)C^0_{h\leftarrow}(\mathbf{k}), \quad (28)$$

$$d_{h\leftarrow}(k_y, x_0) = -\frac{2ik_x\exp(idk_x)\{(\kappa_x + ik_x)\exp[\kappa_x(d+x_0)] + (\kappa_x - ik_x)\exp[-\kappa_x(d+x_0)]\}}{(\kappa_x + ik_x)^2\exp(2d\kappa_x) - (\kappa_x - ik_x)^2\exp(-2d\kappa_x)}, \quad (29)$$

and $\tilde{G}_h(k_y, x_0)$ is the Fourier transform of Green's hole function of homogeneous square barrier (2) with the source at the point $\mathbf{r}_0$,

$$\tilde{G}_h(k_y, x_0) = G_e^*(k_y, x, x_0)\exp(ik_x x - ik_y y_0). \quad (30)$$

The hole wave $\psi^1_{h\rightarrow}(\mathbf{r})$ described by Fourier expansion (24) will be reflected in the Andreev manner [27, 28] from the superconductor and generate an electron wave directed toward the scatterer,

$$\psi^1_{e\leftarrow}(\mathbf{r}) = \frac{1}{2\pi}\int dk_y \exp(ik_y y - ik_x x)C^1_{e\leftarrow}(\mathbf{k}), \quad (31)$$

$$C^1_{e\leftarrow}(\mathbf{k}) = a_h(k_x, k_y)C^1_{h\rightarrow}(\mathbf{k}), \quad (32)$$

$$a_h(k_x, k_y) = a_e(-k_x, k_y). \quad (33)$$

In its turn, the electron wave corresponding to Fourier transform (31) will be reflected normally from the total potential of localized states (1) to generate an electron wave directed deep into the superconductor,

$$\psi^2_{e\rightarrow}(\mathbf{r}) = \frac{1}{2\pi}\int dk_y \exp(ik_y y + ik_x x) C^2_{e\rightarrow}(\mathbf{k}), \quad (34)$$

$$C^2_{e\rightarrow}(\mathbf{k}) = r_{e\leftarrow}(k_y) C^1_{e\leftarrow}(\mathbf{k}) + L_e \tilde{G}_e(k_y, x_0) d^1_{e\leftarrow}(\mathbf{r}_0). \quad (35)$$

In formula (35), $r_{e\leftarrow}(k_y) = r^*_{h\leftarrow}(k_y)$ is the reflection coefficient of the Fourier component of the electron wave described by formula (31) from the square barrier (2), and the quantity $d^1_{e\leftarrow}(\mathbf{r}_0)$ is the amplitude of the electron wave (31) that passed from the superconductor under potential (2) to the localized state,

$$d^1_{e\leftarrow}(\mathbf{r}_0) = \frac{1}{2\pi}\int dk_y \exp(ik_y y_0) d_{e\leftarrow}(k_y, x_0) C^1_{e\leftarrow}(\mathbf{k}), \quad (36)$$

$$d_{e\leftarrow}(k_y, x_0) = d^*_{h\leftarrow}(k_y, x_0). \quad (37)$$

Therefore, we started the treatment with electron wave (12) directed toward the *S*–*N* interface and, after a series of Andreev and normal reflections, we obtained once again an electron wave propagating in the same direction. The process is then repeated for an infinite number of times. In so doing, for the Fourier components of electron and hole waves, formulas (23), (25), (32), and (35) may yield the integral recurrent relations

$$\begin{aligned} C^{2N}_{h\leftarrow}(k) &= a_e(\mathbf{k}) C^{2N}_{e\rightarrow}(\mathbf{k}), \\ C^{2N+1}_{h\rightarrow}(\mathbf{k}) &= r_{h\leftarrow}(k_y) C^{2N}_{h\leftarrow}(\mathbf{k}) + L_h \tilde{G}_h(k_y, x_0) d^{2N}_{h\leftarrow}(\mathbf{r}_0), \\ C^{2N+1}_{e\leftarrow}(\mathbf{k}) &= a_h(\mathbf{k}) C^{2N+1}_{h\rightarrow}(\mathbf{k}), \\ C^{2N+2}_{e\rightarrow}(\mathbf{k}) &= r_{e\leftarrow}(k_y) C^{2N+1}_{e\leftarrow}(\mathbf{k}) + L_e \tilde{G}_e(k_y, x_0) d^{2N+1}_{e\leftarrow}(\mathbf{r}_0), \end{aligned} \quad (38)$$

in which the superscript N indicates the number of Andreev reflections. Formulas (13) and (14) preassign the initial conditions for recurrent relations (38).

3.3. Solutions for Fourier Components of Electron Waves

The Fourier components of scattered electron wave (10) $C_{i\leftrightarrow}(\mathbf{k}, \mathbf{k}_0)$ are expressed in terms of the sums of partial Fourier components $C^N_{e\leftrightarrow}$ described by recurrent relations (38)

$$C_{i\leftrightarrow} = \sum_N C^N_{i\leftrightarrow}. \quad (39)$$

The expressions for these sums (39) follow from recurrent relations (38) (see Appendix B),

$$\begin{aligned} C^{\text{res}}_{e\rightarrow} &= \frac{1}{Q}\{L_e \tilde{G}_e(d^e_{\rightarrow} + \hat{\Sigma}_e) + r_{e\leftarrow} a_n \tilde{G}_h \hat{\Sigma}_h\}, \\ C^{\text{pot}}_{e\leftarrow} &= \frac{a_e a_h r_{h\leftarrow} D_0(k_y)}{Q}, \\ C^{\text{res}}_{e\leftarrow} &= \frac{a_h}{Q}\{r_{h\leftarrow} a_e L_e \tilde{G}_e(d^e_{\rightarrow} + \hat{S}_e) + L_h \tilde{G}_h \hat{\Sigma}_h\}, \end{aligned} \quad (40)$$

where

$$\begin{aligned} Q &= 1 - r_{e\leftarrow} r_{h\leftarrow} a_e a_h, \\ \hat{\Sigma}_e &= \frac{c_1 a_2 + (1-c_2) a_1}{(1-b_1)(1-c_2) - c_1 b_2}, \\ \hat{\Sigma}_h &= \frac{b_2 a_1 + a_2(1-b_1)}{(1-b_1)(1-c_2) - c_1 b_2}, \end{aligned} \quad (41)$$

$$\begin{aligned} a_1 &= \frac{1}{2\pi}\int dk_y \frac{d_{e\leftarrow} r_{h\leftarrow} a_e a_h C^0_{e\rightarrow}}{Q}, \\ a_2 &= \frac{1}{2\pi}\int dk_y \frac{d_{h\leftarrow} a_e C^0_{e\rightarrow}}{Q}, \\ b_1 &= \frac{1}{2\pi}\int dk_y \frac{d_{e\leftarrow} r_{h\leftarrow} a_e a_h L_e \tilde{C}_e}{Q}, \\ b_2 &= \frac{1}{2\pi}\int dk_y \frac{d_{h\leftarrow} a_e L_e \tilde{C}_e}{Q}, \\ c_1 &= \frac{1}{2\pi}\int dk_y \frac{d_{e\leftarrow} L_h a_h \tilde{G}_h}{Q}, \\ c_2 &= \frac{1}{2\pi}\int dk_y \frac{d_{h\leftarrow} r_{e\leftarrow} a_e a_h L_h \tilde{G}_h}{Q}. \end{aligned} \quad (42)$$

The structure of expressions (40) has a clear physical meaning. The numerators in formulas (40) define the power of the sources of electrons moving in the directions towards and away from the superconductor. In the potential term, these sources are the amplitude of the electron passage at potential (2) $D_0(k_y)$ and the quantity $a_e a_h r_{h\leftarrow} D_0(k_y)$, respectively. The coefficient $a_e a_h r_{h\leftarrow}$ indicates that the first electron wave propagating toward the normal metal arises as a result of three successive reflections, namely, the Andreev reflection from the superconductor that converts the electron, which passed with the amplitude $D_0(k_y)$, to a hole with the probability a_e; the reflection of the latter hole from potential (2) with the probability $r_{h\leftarrow}$; and the subsequent transformation of this hole to an electron moving toward normal metal, as a result of Andreev reflection of the hole from the superconductor with the probability a_h.

Two sources of electrons are available in a resonance channel in each direction. This is due to the fact that, as a result of multiple Andreev and normal reflections, electron and hole states with self-consistent amplitudes $\hat{\Sigma}_e$ and $\hat{\Sigma}_h$ are formed on a defect. There-

fore, in a resonance channel, the sources of electrons moving towards the superconductor are equal to the sum of products of the amplitudes of existence on a defect of the electron ($d^e_{\rightarrow} + \hat{\Sigma}_e$) and hole ($\hat{\Sigma}_h$) states into the probabilities of tunneling from a defect, which are $L_e\tilde{G}_e$ and $r_{e\leftarrow}a_nL_h\tilde{G}_h$, respectively. In so doing, the coefficient $r_{e\leftarrow}a_n$ allows for the fact that a forward-moving electron generated by the amplitude $\hat{\Sigma}_h$ of the hole state is obtained as a result of Andreev reflection of the hole from the superconductor and subsequent potential reflection of the electron from potential (2). The backward-moving electrons appear as a result of Andreev reflection of holes ($a_hL_h\tilde{G}_h\hat{\Sigma}_h$) and three successive reflections $r_{h\leftarrow}a_ea_h$ which transform the electron wave $L_e\tilde{G}_e(d^e_{\rightarrow} + \hat{\Sigma}_e)$ (propagating toward the superconductor) into an electron wave propagating in the opposite direction.

The denominator $Q = 1 - r_{e\leftarrow}r_{h\leftarrow}a_ea_h$ in all expressions (40) is a result of summation of repeated Andreev and potential reflections of waves initiated by the sources described above.

The contributions made to the electron current by holes from normal metal ($C_{h\leftrightarrow}(\mathbf{k}, \mathbf{k}_0)$) are calculated analogously,

$$C^{\text{pot}}_{h\rightarrow} = \frac{D^*_0(k_y)r_{e\leftarrow}a_n}{Q},$$

$$C^{\text{res}}_{h\rightarrow} = \frac{1}{Q}\{L_hr_{e\leftarrow}a_h\tilde{G}_h(\hat{\Sigma}^h_e + d^*_{e\rightarrow}) + L_e\tilde{G}_e\hat{\Sigma}^h_h\},$$

$$C^{\text{pot}}_{h\leftarrow} = \frac{D^*_0(k_ya_n)}{Q}, \tag{43}$$

$$C^{\text{res}}_{h\leftarrow} = \frac{a_h}{Q}\{L_h\tilde{G}_h(\hat{\Sigma}^h_e + d^*_{e\rightarrow}) + L_e\tilde{G}_er_{h\leftarrow}a_e\hat{\Sigma}^h_h\},$$

as well as the contributions made by electron-like ($C_{el\leftrightarrow}(\mathbf{k}, \mathbf{k}_0)$) and holelike ($C_{hl\leftrightarrow}(\mathbf{k}, \mathbf{k}_0)$) excitations of superconductor,

$$C^{\text{pot}}_{el\rightarrow} = \frac{r_{e\leftarrow}\sqrt{1-|a_h|^2}}{Q},$$

$$C^{\text{res}}_{el\rightarrow} = \frac{1}{Q}\{L_e\tilde{G}_e\hat{\Sigma}^{el}_e + r_{e\leftarrow}a_nL_h\tilde{G}_h\hat{\Sigma}^{el}_h\},$$

$$C^{\text{pot}}_{el\leftarrow} = \frac{\sqrt{1-|a_h|^2}}{Q}, \tag{44}$$

$$C^{\text{res}}_{el\leftarrow} = \frac{a_h}{Q}\{r_{h\leftarrow}a_eL_e\tilde{G}_e\hat{\Sigma}^{el}_e + L_h\tilde{G}_h\hat{\Sigma}^{el}_h\},$$

$$C^{\text{pot}}_{hl\rightarrow} = \frac{r_{e\leftarrow}r_{h\leftarrow}a_h\sqrt{1-|a_e|^2}}{Q},$$

$$C^{\text{res}}_{hl\rightarrow} = \frac{1}{Q}\{L_e\tilde{G}_e\hat{\Sigma}^{hl}_e + r_{e\leftarrow}a_nL_h\tilde{G}_h\hat{\Sigma}^{hl}_h\},$$

$$C^{\text{pot}}_{hl\leftarrow} = \frac{r_{h\leftarrow}a_h\sqrt{1-|a_e|^2}}{Q}, \tag{45}$$

$$C^{\text{res}}_{hl\rightarrow} = \frac{a_h}{Q}\{r_{h\leftarrow}a_eL_e\tilde{G}_e\hat{\Sigma}^{hl}_e + L_h\tilde{G}_h\hat{\Sigma}^{hl}_h\}.$$

The quantities $\hat{\Sigma}^i_{e,h}$ in formulas (43)–(45) are defined by relations (41). The difference consists in the expressions for the coefficients of (42). In formulas (43), the coefficients a_i, b_i, and c_i are described by formulas (42) with the $e \leftrightarrow h$ replacement. In formulas (44) and (45), for the same coefficients b_i and c_i, one must use different coefficients a_i,

$$a^{el}_1 = \frac{\sqrt{1-|a_h|^2}d_{e\leftarrow}}{Q},$$

$$a^{el}_2 = \frac{\sqrt{1-|a_h|^2}d_{h\leftarrow}a_er_{e\leftarrow}}{Q}, \tag{46}$$

$$a^{hl}_1 = \frac{\sqrt{1-|a_e|^2}d_{e\leftarrow}a_hr_{h\leftarrow}}{Q},$$

$$a^{hl}_2 = \frac{\sqrt{1-|a_e|^2}d_{h\leftarrow}}{Q}. \tag{47}$$

As in formula (40), the numerators in expressions (43)–(45) define the sources of electron waves propagating towards and away from the superconductor, while the denominators allow for repeated Andreev and potential reflections of these waves.

In numerical calculations, the Andreev coefficients $a_{e,h}(k_x, k_y)$ in the foregoing expressions were calculated self-consistently in accordance with the procedure described in [10, 11]. The calculations were performed for arbitrary angles α between the normal to the interface and the crystallographic direction *a* of the *d*-type superconductor. Note that it is only for the orientation angle $\alpha = 0$ that the results coincide with the values obtained in a non-self-consistent manner,

$$a_e(k_x, k_y) = \begin{cases} \dfrac{\varepsilon - \text{sgn}(\varepsilon)\sqrt{\varepsilon^2 - |\Delta(\mathbf{k})|^2}}{\Delta(\mathbf{k})}, & \varepsilon \geq |\Delta|, \\ \dfrac{\varepsilon - i\sqrt{|\Delta(\mathbf{k})|^2 - \varepsilon^2}}{\Delta(\mathbf{k})}, & \varepsilon < |\Delta|, \end{cases} \tag{48}$$

$$\Delta(\mathbf{k}) = \Delta_0\cos[2(\theta - \alpha)],$$

and used previously [7, 8] in analyzing resonant transport in weak couplings on the basis of *d*-type superconductors.

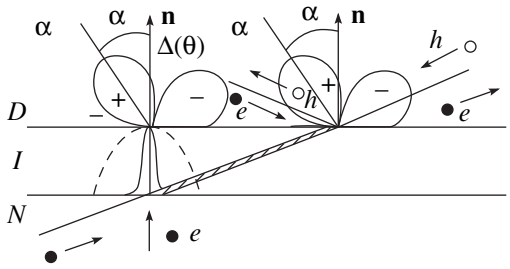


Fig. 1. A diagram illustrating the process of tunneling in a two-dimensional *N–I–D* junction without defects. The broken line in region I indicates the angular dependence of penetrability of a δ-functional barrier, and the solid line indicates the angular dependence of penetrability of a thick square barrier.

4. ZBA UNDER CONDITIONS OF RESONANCE TUNNELING

In the low-voltage region, the amplitudes of the Andreev reflection coefficients $a_{h,e}$ tend to $\pm i$ (see Eq. (48)) irrespective of the behavior of spatial variation of the order parameter in the neighborhood of the boundary. If, in so doing, as a result of an event of being scattered by this boundary, a quasiparticle enters the region with the same sign of the order parameter, then $(a_e a_h)_{e=0} = -1$. This is observed, for example, in the case of superconductor with pairing of the *s*-type (*N–I–S* junction) or in the particular case of $\alpha = 0$ in *N–I–D* structures. However, if $\alpha \neq 0$, this change of sign may occur and lead to the relation

$$(a_e a_h)_{\varepsilon=0} = 1. \tag{49}$$

In this case, Eqs. (40) and (43)–(45) immediately produce anomalously high values of the coefficients $C_{i\leftrightarrow}$ and, consequently, anomalies in the conductance as well.

On substituting the expressions for Fourier components $C_{i\leftrightarrow}(\mathbf{k}, \mathbf{k}_0)$ (40)–(43) into those for current (5) and (7) and using Eq. (41), one can derive the following expression for the potential component of conductance at $eU \longrightarrow 0$:

$$G_d^{\text{pot}}(0) = \frac{2e^2}{\pi\hbar}\frac{k_0}{2\pi}L_y \int\limits_{ZBA} d\theta \cos\theta. \tag{50}$$

By integration over ZBA is meant the integration with respect to the region of angles in the direction of which the condition of electron-hole resonance (49) is valid, which leads to the formation of a bound state with the energy $E_M = 0$ in the superconductor.

Relation (50) is a well-known expression for ZBA obtained previously [7] (though with some inaccuracies: the cosine under the integral was not included) for potential scattering in the model of the δ-potential barrier. It is interesting to note that in the model being treated of a "thick" barrier with penetrability, described by formula (15), the value of ZBA is the same as in the case of the model of the δ-potential barrier, in spite of the fact that the penetrability modulus of a thick square barrier has a much sharper maximum in the region of small angles of incidence $\theta = 0$ (see Fig. 1) than the δ-potential barrier. This result comes naturally: ZBAs arise as a result of resonance between the Fermi energies of a normal metal and of a bound electron–hole state with $E_M = 0$ on the *I–D* interface. The width of the electron–hole state with $E_M = 0$ is defined by the probability of electron (hole) escape from the *I–S* region as a result of tunneling back to the normal metal. Therefore, even though the first transmission of the electron wave function in the direction satisfying condition (49) may be very unlikely for a thick square barrier, this low probability is compensated by the formation of a narrower electron–hole resonance. The width of the ZBA peak with respect to voltage is of the order of $|D(\theta_{ZBA})|^2\Delta_0$ (see Figs. 2 and 3).

Figures 2a and 3a give the conductance of the potential channel of an *N–I–D* junction as a function of voltage for orientation angles $\alpha = \pi/4$ and $\alpha = \pi/8$, respectively, with the parameters $\kappa_0 d = 2$ and $\kappa_0/k_0 = 0.1$, calcu-

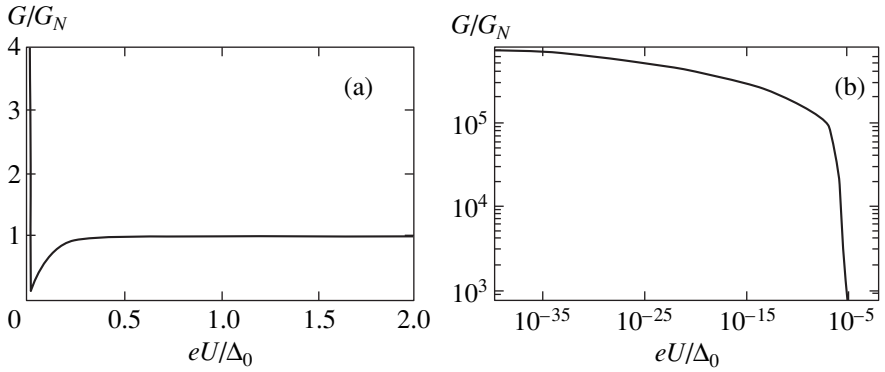


Fig. 2. (a) The conductance of an *N–I–D* junction with a thick square barrier without defects with the orientation angle $\alpha = \pi/4$ and parameters $\kappa_0 d = 2$ and $\kappa_0/k_0 = 0.1$, calculated using non-self-consistent Andreev coefficients (48). The conductance is normalized to that of an analogous *N–I–N* junction. (b) The low-voltage conductance on a logarithmic scale.

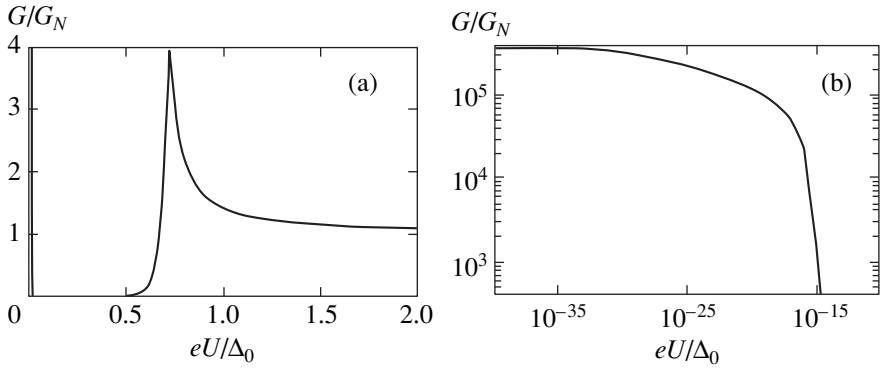


Fig. 3. (a) The conductance of an *N–I–D* junction with a thick square barrier without defects with the orientation angle $\alpha = \pi/8$ and parameters $\kappa_0 d = 2$ and $\kappa_0/k_0 = 0.1$, calculated using non-self-consistent Andreev coefficients (48). The conductance is normalized to that of an analogous *N–I–N* junction. (b) The low-voltage conductance on a logarithmic scale.

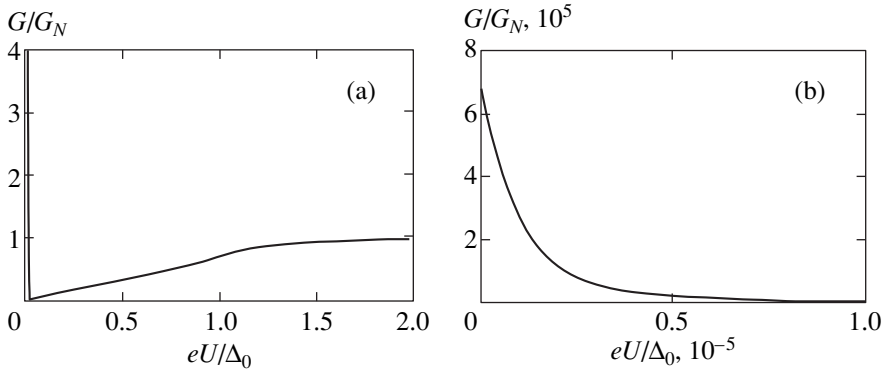


Fig. 4. (a) The conductance of an *N–I–D* junction with a δ-functional barrier with normal conductance equal to that of a thick square barrier with the parameters $\kappa_0 d = 2$ and $\kappa_0/k_0 = 0.1$. The orientation angle $\alpha = \pi/4$. The calculations were performed using non-self-consistent Andreev coefficients (48). The conductance is normalized to that of an analogous *N–I–N* junction. (b) The low-voltage conductance.

lated using non-self-consistent Andreev coefficients (48). These dependences are normalized to the conductance of an analogous *N–I–N* junction. Shown separately in Figs. 2b and 3b are low-voltage regions. One can see that, in both cases, the normalized conductance exhibits a singularity at $eU \longrightarrow 0$. In so doing, the width of this singularity at $\alpha = \pi/8$ is approximately 10 orders of magnitude less than at $\alpha = \pi/4$. This is attributed to the extremely low penetrability of a thick square barrier for directions of $\theta_{ZBA} \neq 0$. For the orientation angle $\alpha = \pi/8$, this region is such that $\theta_{ZBA} \in [-5\pi/32, -\pi/8] \cup [\pi/8, 5\pi/32]$, while for the orientation angle $\alpha = \pi/8$ condition (49) is valid for any values of θ, including $\theta = 0$.

For comparison, Figs. 4a and 5a give the normalized conductance of an *N–I–D* junction as a function of voltage, calculated in the model of a δ-functional scatterer using non-self-consistent Andreev coefficients (48) (see [7]). The coefficient at the δ function corresponds to the conductance of the junction as in the case of the square barrier in Figs. 2 and 3. Figures 4b and 5b give the same dependences in the low-voltage region. One can see that the width of the peak of conductance for the orientation angle $\alpha = \pi/8$ is less than for the orientation angle $\alpha = \pi/4$; however, both quantities are of the same order of magnitude. This is attributed to the fact that the angular dependence of penetrability in the model of δ-functional scatterer is less sharp than in the case of a long square barrier (see Fig. 1).

Figure 6 gives the normalized conductance of an *N–I–D* junction with a square barrier as a function of voltage, calculated using self-consistent Andreev coefficients [10, 11] for the case of mirror interface. A comparison of Fig. 6 with Figs. 2 and 3 reveals a qualitative agreement in the behavior of the dependences.

In the case of isotropic scattering of quasiparticles by localized states, directions will always be available, the tunneling along which will cause a quasiparticle to get to the region of existence of electron-hole resonance (49) (Fig. 7). The interaction with a defect in the process of tunneling causes an effective "scanning" of the interface by a scattered wave, which inevitably leads to the formation of resonance trajectories. It is qualitatively clear that the fact

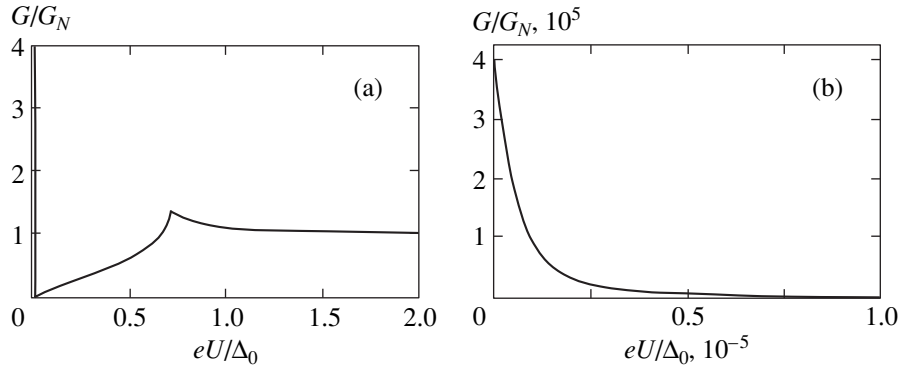


Fig. 5. (a) The conductance of an *N–I–D* junction with a δ-functional barrier with normal conductance equal to that of a thick square barrier with the parameters $\kappa_0 d = 2$ and $\kappa_0/k_0 = 0.1$. The orientation angle $\alpha = \pi/8$. The calculations were performed using non-self-consistent Andreev coefficients (48). The conductance is normalized to that of an analogous *N–I–N* junction. (b) The low-voltage conductance.

of formation of such a channel is associated neither with the position of the energy level of localized state nor with its spatial arrangement. The fact of isotropic rescattering by any defect in the process of scattering will suffice.

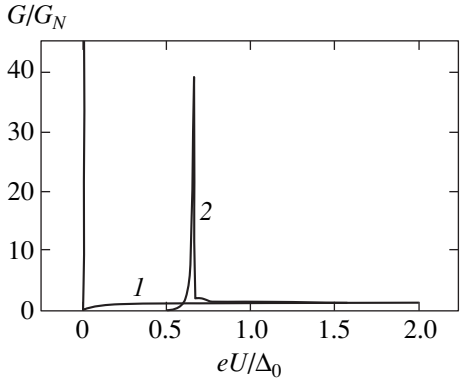


Fig. 6. The conductance of an *N–I–D* junction with a thick square barrier without defects with the parameters $\kappa_0 d = 2$ and $\kappa_0/k_0 = 0.1$, calculated using self-consistent Andreev coefficients. The conductance is normalized to that of an analogous *N–I–N* junction. Curve *1* corresponds to the orientation angle $\alpha = \pi/4$, and curve *2* corresponds to the orientation angle $\alpha = \pi/8$.

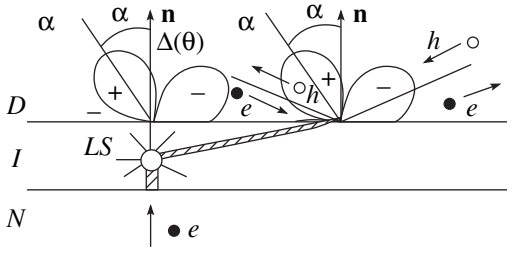


Fig. 7. A diagram illustrating the process of tunneling in a two-dimensional *N–I–D* junction with a defect in the *I*-layer.

Given the validity of the easy-to-satisfy condition for the coordinate of localized state x_0 and ZBA angles,

$$\int_{\text{ZBA}} d\theta \frac{\kappa_x^2 + k_x^2}{4\kappa_x^2} \exp[2\kappa_x(d + x_0)] \gg 1, \quad \kappa_0(d + x_0) \gg 1, \tag{51}$$

the amplitudes of the electron $(d^e_{\rightarrow} + \hat{\Sigma}_e)$ and hole $(\hat{\Sigma}_h)$ states on a localized state are defined mainly by waves rescattered backward, away from the superconductor. When condition (51) is valid and after the substitution of expressions for Fourier components $C_{i\leftrightarrow}(\mathbf{k}, \mathbf{k}_0)$ (40)–(43) into formulas for current (5), (8), and (9), one can derive the following relations for the interference and resonance components of conductance at $eU \longrightarrow 0$, which describe the electron wave scattering from a defect:

$$G_d^{\text{int}}(0) \approx -4\frac{e^2}{\pi\hbar} = -8R_K^{-1}, \quad R_K = \frac{2\pi\hbar}{e^2}, \tag{52}$$

$$G_d^{\text{res}}(0) \approx 0. \tag{53}$$

In analyzing the processes of resonance tunneling, the interference component (52) of current (8) was not included, as a rule, because of the sharper (compared with the resonance channel) dependence on the barrier thickness. In *N–I–D* contacts, this contribution to current proves to be significant and the most nontrivial. It reflects the process of destruction of resonance in a potential channel due to the interaction with the defect. Indeed, in addition to the validity of condition (49), the conservation of the transverse component of quasiparticle momentum in the process of reflection of the quasiparticle from a spatially homogeneous barrier is required for the formation of a bound electron-hole state with $E_M = 0$ on the *I–D* interface. However, the presence of a scattering center on the trajectory inevitably leads to the quasiparticle rescattering in other directions, which results in the breaking of electron-hole resonance. This effect must not depend considerably either on the posi-

tion of the energy level of localized state or on its spatial arrangement on a preferred trajectory. What this effect leads to is the reduction of the conductance by eight quanta R_K^{-1} per each defect. Note, however, that within the given model this cannot bring about the change of sign of the overall differential conductance with increasing concentration of localized states, because it will be preceded by the condition of independence of tunneling via separate scattering centers being violated.

Note that, in the absence of Andreev reflection (in an *N–LS–N* structure), the interference contribution to current depends considerably on the shape of the resonance curve and goes to zero both at $eU = \varepsilon_R$ and upon averaging with respect to ε_R.

The paradoxical, at first glance, result for resonance channel (53) is associated with the fact that, in accordance with Eqs. (40), the power of the sources of electrons moving toward one another is defined by the amplitudes of probabilities of the existence on a defect of both electron $(d^e_{\rightarrow} + \hat{\Sigma}_e)$ and hole $(\hat{\Sigma}_h)$ states. In so doing, it turns out that the "number" of "forward"-propagating (i.e., towards the superconductor) electrons generated by the electron states exceeds, by a factor of $(r_{h\leftarrow}a_e a_h)^{-1}$, the number of "backward"-moving electrons. For electrons generated by the hole state, the situation on the localized state is quite the contrary: the number of electrons moving forward proves to be r_e times less than the number of electrons moving backward. In view of Eq. (49), as well as of the fact that the resonance contribution to current is defined, according to formula (9), by the difference of squares of modules $\left|C^{\text{res}}_{i\rightarrow}(\mathbf{k}, \mathbf{k}_0)\right|^2$ and $\left|C^{\text{res}}_{i\leftarrow}(\mathbf{k}, \mathbf{k}_0)\right|^2$, one can readily find that the contributions to resonance current (9) by the electron and hole states on a localized state at $eU \longrightarrow 0$ fully compensate each other and lead to the results given by Eq. (53).

Therefore, the result given by Eqs. (52) and (53) proves to be fairly general and independent of the shape of the resonance curve. Moreover, it is conserved also during the change of sign in defect potential (3) from negative (resonance) to positive (nonresonance scattering). Consequently, the effect of ZBA suppression occurs in the case of both resonance and direct scattering by a defect; what is important is only the nonconservation of the transverse component of scattered wave momentum upon interaction with defect. In addition, it is not only a part of the localized states with the energy in the vicinity of the Fermi energy (differing from the latter by a value of several half-widths of the resonance line Γ or less) and with the coordinate in the vicinity of the middle of the barrier (spaced from the middle at a distance of several radii of localized state κ_0^{-1} or less), which defined the resonance current in the case of structures with normal or superconducting electrodes with pairing of the *s*-type [20–23], that make the contribution to the suppression of the conductance anomaly (50) in the potential channel; this contribution is also due to almost all of the defects inside the barrier with coordinates satisfying condition (51).

5. RESONANCE TUNNELING WITH THE ORIENTATION ANGLE $\alpha = 0$

Because, in the case of orientation angle $\alpha = 0$, the effects associated with the emergence of ZBA are absent, a *d*-type superconductor may be treated as an ordinary anisotropic *s*-type superconductor. Therefore, it is sufficient to treat the scattering of electron and hole wave functions, defined by relations (40)–(47), only in the narrow cone of angles,

$$\theta \approx \frac{\kappa_0}{k_0}\frac{1}{\sqrt{\kappa_0 d}} \ll 1. \tag{54}$$

Therefore, for thick and relatively low barriers which satisfy condition (54), resonance scattering becomes one-dimensional. Integrals (42) are taken in an explicit form and, after the substitution of the obtained Fourier components $C_{i\leftrightarrow}(\mathbf{k}, \mathbf{k}_0)$ into the expressions for current (5)–(9), we arrive at formula (3) in [23] for the conductance of a structure at low voltages $eU \ll \Delta_0$ with the reflection and transmission coefficients

$$r(\varepsilon) = \frac{(\varepsilon - \varepsilon_R) + i\left(\dfrac{\Gamma_l^{2D} - \Gamma_r^{2D}}{2}\right)}{(\varepsilon - \varepsilon_R) + i\Gamma^{2D}}, \tag{55}$$

$$|t(\varepsilon)|^2 = 1 - |r(\varepsilon)|^2,$$

which allow for the 2D behavior of scattering in the interlayer in expressions for $\Gamma^{2D}_{l,r}$ (17).

6. CONCLUSION

The results of sequential theoretical analysis of resonance tunneling in two-dimensional *N–I–D* structures with defects in the interlayer have demonstrated that, in the presence of *d*-pairing in electrodes, one cannot restrict oneself to a one-dimensional approximation. Qualitatively, this is associated with the fact that, in the case of *N–I–D* structures, direct tunneling in the ZBA region is also resonance tunneling to the bound electron-hole state, which leads to the formation of ZBA. In so doing, the interference of two processes occurring in the low-voltage region brings about both a partial suppression of ZBA and a sharp enhancement of the process of tunneling via localized states, which proves to involve almost all localized states, irrespective of their position in the interlayer. This result is fairly general and independent of the shape of the resonance curve and describes both resonance (at $\beta > 0$) and direct scattering by the potential of defect (3). It is only with the orientation angle $\alpha = 0$ that the problem reduces to a one-dimensional one, and we derive the relation for conductance, which is in formal agreement with the analogous expression derived for the 1D case.

ACKNOWLEDGMENTS

This study received partial support from State Contract 107-6(00)-P.

APPENDIX A

Two-Dimensional One-Electron Resonance Scattering

We will treat the electron wave scattering by 2D-potential (1). For this purpose, we will write the solution of Schrödinger's equation for potential (1) in the form

$$\psi(\mathbf{r}) = \psi_0(\mathbf{r}) + \int (d^2r')G_0(\varepsilon, \mathbf{r}, \mathbf{r}')V_{imp}(\mathbf{r}' - \mathbf{r}_0)\psi(\mathbf{r}'), \quad (A.1)$$

where $\psi_0(\mathbf{r})$ and $G_0(\varepsilon, \mathbf{r}, \mathbf{r}')$ denote the electron wave function and Green's function of two-dimensional square potential (2), respectively, which satisfy the equations

$$[\varepsilon - H_0(\mathbf{r})]\psi_0(\mathbf{r}) = 0,$$
$$[\varepsilon - H_0(\mathbf{r})]G_0(\varepsilon, \mathbf{r}, \mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}'), \quad (A.2)$$
$$H_0 = -\frac{\hbar^2\nabla^2}{2m} + V_{rect}(x).$$

The defect radius ρ has atomic dimensions, i.e., $\rho \ll k_0^{-1}$, while the wave function $\psi(\mathbf{r})$ is, on an atomic scale, a slow-varying function of the variable $\mathbf{r}$. This enables one to remove the wave function in Eq. (A.1) from the integrand and represent this expression as

$$\psi(\mathbf{r}) = \psi_0(\mathbf{r}) + \psi(\mathbf{r}_0)\int d^2r' G_0(\varepsilon, \mathbf{r}, \mathbf{r}')V_{imp}(\mathbf{r}' - \mathbf{r}_0). \quad (A.3)$$

We assume in Eq. (A.3) that $\mathbf{r} = \mathbf{r}_0$ to readily derive, for the local value of $\psi(\mathbf{r}_0)$,

$$\psi(\mathbf{r}_0) = \frac{\psi_0(\mathbf{r}_0)}{1 - \int d^2r' G_0(\varepsilon, \mathbf{r}_0, \mathbf{r}')V_{imp}(\mathbf{r}' - \mathbf{r}_0)}. \quad (A.4)$$

We will further substitute expression (A.4) into (A.3) in view of the fact that Green's function $G_0(\varepsilon, \mathbf{r}, \mathbf{r}_0)$ appearing in Eq. (A.3) on a scale ρ, on which variations of the defect potential $V_{imp}(\mathbf{r} - \mathbf{r}_0)$ occur, is also a slow-varying function of the variable $\mathbf{r}$ and derive the final expression for an electron wave scattered by a localized state (see our formulas (13)–(17)),

$$\psi(\mathbf{r}) = \psi_0(\mathbf{r}) + L_e\psi_0(\mathbf{r}_0)G_0(\varepsilon, \mathbf{r}, \mathbf{r}_0), \quad (A.5)$$

$$L_e = \frac{\int d^2r' V_{imp}(\mathbf{r}' - \mathbf{r}_0)}{1 - \int d^2r' G_0(\varepsilon, \mathbf{r}, \mathbf{r}')V_{imp}(\mathbf{r}' - \mathbf{r}_0)}. \quad (A.6)$$

The first term in formula (A.5) describes potential scattering by potential (1), and the second term describes resonance scattering by a localized state.

In order to calculate the electron amplitude of resonance scattering L_e, one must know the expression for Green's electron function in the barrier,

$$G_e(\mathbf{r}, \mathbf{r}') = \frac{1}{2\pi}\int dk_y \exp[ik_y(y - y')]G_e(k_y, x, x'),$$
$$G_e(k_y, x, x_0)_{|x, x'| < d} = G_V + G_1,$$
$$G_V = -\frac{m}{\kappa_x^e\hbar^2}\exp(-\kappa_x^e|x - x'|), \quad (A.7)$$
$$G_1 = \frac{2m}{\kappa_x^e\hbar^2}\frac{\kappa_x^e + ik_x^e}{ik_x^e - \kappa_x^e}\exp(-2\kappa_x^e d)\cosh[\kappa_x^2(x + x')],$$

where G_V is Green's electron function of rectangular barrier (2) of infinite thickness $d \longrightarrow \infty$, and G_1 denotes the corrections to this function due to finiteness of the barrier,

$$k_x^e = \sqrt{k_0^{e2} - k_y^2}, \quad k_0^{e2} = 2m(E_f + \varepsilon)/\hbar^2,$$
$$\kappa_x^e = \sqrt{\kappa_0^{e2} + k_y^2}, \quad \kappa_0^{e2} = 2m(V_0 - E_f - \varepsilon)/\hbar^2.$$

We substitute Green's functions (A.7) into (A.6) in view of the form of local potential $V_{imp}(\mathbf{r} - \mathbf{r}_0)$ (3) to derive the following expression for the electron amplitude of resonance scattering L_e:

$$L_e = \frac{\pi}{m}\frac{\hbar^2}{\ln(\kappa_{res}^2/\kappa_0^e) + J' + iJ''},$$
$$\kappa_{res}^e \equiv \frac{\sqrt{2m(V_0 - E_f - \varepsilon_0)}}{\hbar} = \frac{\eta}{\rho}\exp\left(-\frac{\hbar^2}{m\beta\rho^2}\right), \quad (A.8)$$

where $\hbar\kappa_{res}^e$ is the resonance value of momentum defining the value of resonance energy ε_0, $\eta = 2\exp(0.5)/\gamma$, $\gamma \approx 1.78$ is Euler's constant, and J' and J'' have the form

$$J' = \frac{\hbar^2\sqrt{\pi}}{2mV_0}(\kappa_0^{e2} - k_0^{e2})\phi(x_0),$$
$$J'' = \frac{\hbar^2\sqrt{\pi}}{mV_0}(\kappa_0^e k_0^e)\phi(x_0), \quad (A.9)$$
$$\phi(x_0) = \frac{1}{2}\left(\frac{\exp[-2\kappa_0^e(d - x_0)]}{\sqrt{\kappa_0^e(d - x_0)}} + \frac{\exp[-2\kappa_0^e(d + x_0)]}{\sqrt{\kappa_0^e(d + x_0)}}\right).$$

We expand the logarithm in formula (A.8) in the vicinity of the resonance value of its argument ($\kappa_{res}^e = \kappa_0^e$) to derive formulas (16) and (17) in which

$$\varepsilon_R = \varepsilon_0 + \delta_\varepsilon, \quad \delta_\varepsilon = -2(V_0 - E_f)J', \quad (A.10)$$

where $\delta\varepsilon$ is the shift of the resonance energy, which is caused by the finite width of the barrier.

The amplitude of resonance scattering for holes L_h is calculated analogously. In so doing, one must change the sign of energy ($\varepsilon \longrightarrow -\varepsilon$) and take into account that Green's hole functions are related to electron functions by the relation $G_h(k_y, x, x') = G_e^*(k_y, x, x')_{\varepsilon \to -\varepsilon}$.

APPENDIX B

Solution of Recurrent Equations

In order to calculate the electric current through a structure, one must know the Fourier components of scattered electron wave $C_{i\leftrightarrow}(\mathbf{k}, \mathbf{k}_0)$, which are sums of the respective partial Fourier components $C_{e\leftrightarrow}^N$ satisfying integral recurrent relations (38) with initial conditions (13) and (14). In order to find them in an explicit form, we will use the formulas

$$C_{e\leftarrow}^{2N+1} = a_e(\mathbf{k})a_h(\mathbf{k})r_{h\leftarrow}(k_y)C_{e\rightarrow}^{2N} + a_h(\mathbf{k})L_h\tilde{G}_h(k_y, x_0)d_{h\leftarrow}^{2N}(\mathbf{r}_0), \quad \text{(B.1)}$$

$$C_{e\rightarrow}^{2N+2} = r_{e\leftarrow}(k_y)r_{h\leftarrow}(k_y)a_e(\mathbf{k})a_h(\mathbf{k})C_{e\leftarrow}^{2N} + L_e\tilde{G}_e(k_y, x_0)d_{e\leftarrow}^{2N+1}(\mathbf{r}_0) + r_{e\leftarrow}(k_y)a_h(\mathbf{k})L_h\tilde{G}_h(k_y, x_0)d_{h\leftarrow}^{2N}(\mathbf{r}_0), \quad \text{(B.2)}$$

which follow from recurrent relations (38). We sum recurrent formula (B.2) with respect to the parameter $q = r_{e\leftarrow}(k_y)r_{h\leftarrow}(k_y)a_e(\mathbf{k})a_h(\mathbf{k})$, $|q| < 1$ to derive the following expression for $C_{e\rightarrow}$:

$$C_{e\rightarrow} = \frac{1}{Q}\{C_{e\rightarrow}^0 + L_e\tilde{G}_e(k_y, x_0)\hat{\Sigma}_e + r_{e\leftarrow}(k_y)a_h(\mathbf{k})L_h\tilde{G}_h(k_y, x_0)\hat{\Sigma}_h\}, \quad \text{(B.3)}$$

in which

$$\hat{\Sigma}_e = \sum d_{e\leftarrow}^{2N+1}(\mathbf{r}_0), \quad \hat{\Sigma}_h = \sum d_{h\leftarrow}^{2N}(\mathbf{r}_0). \quad \text{(B.4)}$$

Analogously, formulas (B.1) and (38) yield the expressions for $C_{e\leftarrow}$, $C_{h\leftarrow} = \sum C_{h\leftarrow}^{2N+1}$,

$$C_{e\leftarrow} = \frac{1}{Q}\{a_e(\mathbf{k})a_h(\mathbf{k})r_{h\leftarrow}(k_y)C_{e\rightarrow}^0 + a_e(\mathbf{k})a_h(\mathbf{k})r_{h\leftarrow}(k_y)L_e\tilde{G}_e(k_y, x_0)\hat{\Sigma}_e + a_h(\mathbf{k})L_h\tilde{G}_h(k_y, x_0)\hat{\Sigma}_h\}, \quad \text{(B.5)}$$

$$C_{h\leftarrow}(\mathbf{k}) = a_e(\mathbf{k})C_{e\leftarrow}(\mathbf{k}). \quad \text{(B.6)}$$

It follows from formulas (28), (36), and (B.4) that $\hat{\Sigma}_e$ and $\hat{\Sigma}_h$ are the sums of respective electron and hole partial Fourier components, integrated with respect to k_y with the weight factors $d_{e,h\leftarrow}(k_y)$; therefore, in order to find these sums, it is sufficient to multiply formula (B.5) by $d_{e\leftarrow}(k_y)$ and (B.6) by $d_{h\leftarrow}(k_y)$, integrate the resultant expressions with respect to k_y, and solve the resultant set of two linear equations relative to $\hat{\Sigma}_e$ and $\hat{\Sigma}_h$. As a result, we derive formula (40) for the Fourier components of scattered electron wave with coefficients defined by relations (41) and (42).

REFERENCES

1. C. C. Tsuei, J. R. Kirtley, C. C. Chi, *et al.*, Phys. Rev. Lett. **73**, 593 (1994).
2. D. A. Wollman, D. J. van Harlingen, D. J. Lee, *et al.*, Phys. Rev. Lett. **71**, 2134 (1993).
3. D. A. Brawner and H. R. Ott, Phys. Rev. B **50**, 6530 (1994).
4. D. A. Wollman, D. J. van Harlingen, J. Giapintzakis, *et al.*, Phys. Rev. Lett. **74**, 797 (1995).
5. Yu. S. Barash, A. V. Galaktionov, and A. D. Zaikin, Phys. Rev. B **52**, 665 (1995).
6. C.-R. Hu, Phys. Rev. Lett. **72**, 1526 (1994).
7. Y. Tanaka and S. Kashiwaya, Phys. Rev. Lett. **74**, 3451 (1995).
8. Y. Tanaka and S. Kashiwaya, Phys. Rev. B **53**, R11957 (1996).
9. S. Kashiwaya, Y. Tanaka, M. Koyanagi, and K. Kajimura, Phys. Rev. B **53**, 2667 (1996).
10. Yu. S. Barash, A. A. Svidzinsky, and H. Burkhardt, Phys. Rev. B **55**, 15282 (1997).
11. A. A. Golubov and M. Yu. Kupriyanov, Pis'ma Zh. Éksp. Teor. Fiz. **69**, 242 (1999) [JETP Lett. **69**, 262 (1999)].
12. L. Alff, R. Gross, A. Marx, *et al.*, Phys. Rev. B **58**, 11197 (1998).
13. A. A. Golubov, M. A. J. Verthoeven, I. A. Devyatov, *et al.*, Physica C (Amsterdam) **235**, 1361 (1994).
14. T. Satoh, M. Hidaka, M. Yu. Kupriyanov, *et al.*, IEEE Trans. Appl. Supercond. **5**, 2612 (1995).
15. I. I. Vengrus, M. Yu. Kupriyanov, O. V. Snigirev, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **60**, 372 (1994) [JETP Lett. **60**, 381 (1994)].
16. R. Dommel, C. Horstmann, M. Siegel, *et al.*, Appl. Phys. Lett. **67**, 1775 (1995).
17. J. Yoshida, T. Nagano, and T. Hashimoto, Phys. Rev. B **53**, 8623 (1996).
18. Y. Savada, H. Terai, A. Fujimaki, *et al.*, IEEE Trans. Appl. Supercond. **5**, 2099 (1995).

19. J. Yoshida, IEICE Trans. Electron. **E83-C**, 49 (2000).
20. A. I. Larkin and K. A. Matveev, Zh. Éksp. Teor. Fiz. **93**, 1030 (1987) [Sov. Phys. JETP **66**, 580 (1987)].
21. H. Knauer, J. Richer, and P. Siedel, Phys. Status Solidi A **44**, 303 (1977).
22. I. M. Lifshits and V. Ya. Kirpichenkov, Zh. Éksp. Teor. Fiz. **77**, 989 (1979) [Sov. Phys. JETP **50**, 499 (1979)].
23. I. L. Aleiner, H. Clarke, and L. I. Glazman, Phys. Rev. B **53**, R7630 (1996).
24. G. Johansson, E. N. Bratus, V. S. Shumeiko, and G. Wendin, cond-mat/9807240 (1998).
25. I. A. Devyatov and M. Yu. Kupriyanov, Zh. Éksp. Teor. Fiz. **114**, 687 (1998) [JETP **87**, 375 (1998)].
26. G. E. Blonder, M. Tinkham, and T. M. Klahwijk, Phys. Rev. B **25**, 4515 (1982).
27. M. Belogolovskiĭ, M. Graiger, P. Kus, *et al.*, Phys. Rev. B **59**, 9617 (1999).
28. A. F. Andreev, Zh. Éksp. Teor. Fiz. **46**, 1823 (1964) [Sov. Phys. JETP **19**, 1228 (1964)].
29. C. Bruder, Phys. Rev. B **41**, 4017 (1990).

Translated by H. Bronstein

Journal of Experimental and Theoretical Physics, Vol. 92, No. 4, 2001, pp. 665–676.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 4, 2001, pp. 763–776.

SOLIDS
Electronic Properties

Boson–Fermion Mixed Site States

A. I. Agafonov* and É. A. Manykin

Institute of Superconductivity and Solid State Physics, Kurchatov Institute, Russian State Scientific Centre, Moscow, pl. Kurchatova 1, 123182 Russia

**e-mail: aai@isssph.kiae.ru*

Received August 11, 2000

Abstract—It is shown that an essential aspect in the study of polarons, which belong to multiparticle systems, is a correct selection of the Fock space from a manifold of unitarily inequivalent representations. Using the quantum field theory, we have studied a single-site polaron model with an allowance for the electron correlations at a lattice site. This model takes into account the formation of a charged site boson state in addition to the single-particle fermion states. It is demonstrated that, provided the initial site sates are half-occupied, the formation of boson–polaron mixed site states takes place. The thermodynamics of these states is considered.

1. INTRODUCTION

A fundamental problem of solid state physics is to determine conditions under which the electron–phonon interaction may qualitatively change the eigenstates of charge carriers in a dielectric [1–4]. At present, a considerable attention of researchers is drawn to the study of small polarons, which is partly related to the rapid development of computational methods for the study of polarons at large values of the electron–phonon interaction parameters [5–8]. On the other hand, the interest in these investigations is due to the possibility of constructing a high-temperature superconductivity (HTSC) theory and providing a description for the unusual properties of manganites [9–14].

Since polarons belong to multiparticle systems, there exist a large number of representations that can be unitarily inequivalent to each other. This situation poses a problem of correctly selecting a representation for the Fock space [15]. In the quantum field theory (QFT), the Fock space is constructed by means of a cyclic action of the production operators of physical particles (i.e., of the particles observed in a physical experiment) upon their vacuum states. Here, an important observable quantity is the quantum system energy equal to the sum of the energies of all physical particles in a given state. The operator commutation relationships by themselves cannot unambiguously determine the correct representation because it is necessary to specify the vacuum state as well.

The key role in the study of polaron and bipolaron bands in multisite polaron models [10, 11, 16–21] belongs to the single-site Hamiltonian. For the Fröhlich polaron, this Hamiltonian has the following form:

$$H_{\text{site}} = \sum_{\sigma}\varepsilon_0 n_\sigma + \sum_{\sigma} n_\sigma \sum_{\mathbf{q}} \hbar\Omega\xi_{\mathbf{q}}(b_{\mathbf{q}} + b^{+}_{-\mathbf{q}}) + \sum_{\mathbf{q}} \hbar\Omega\left(b^{+}_{\mathbf{q}} b_{\mathbf{q}} + \frac{1}{2}\right), \quad (1)$$

where ε_0 is the seeding electron ground state energy in the site; n_σ is the occupancy operator of the state with a spin projection $\sigma = \pm(1/2)(\uparrow, \downarrow)$; $b_{\mathbf{q}}$ is the destruction operator for the phonon with the wavevector $\mathbf{q}$ and the energy Ω (for simplicity, we consider dispersionless phonons); and $\xi_{\mathbf{q}}$ is the matrix element of the electron–phonon coupling. In solving Eq. (1), we must take into account the initial site state occupancy:

$$\left\langle \Psi_i \left| \sum_{\sigma} n_\sigma \right| \Psi_i \right\rangle = 2A,$$

where $n_\sigma = a^{+}_{\sigma} a_\sigma$ and $|\Psi_i\rangle$ is the initial state for an unperturbed electron Hamiltonian determined by the first term in the right-hand part of Eq. (1).

Note that, neglecting the hopping between sites in a multipolaron system, an electron–phonon Hamiltonian used in constructing the HTSC theory [10, 11] also reduces to Eq. (1). Using, in addition, the approximation of the spin-nondegenerate single electron state in the site, the polaron model [16] reduces to Eq. (1) as well.

An approach that is widely used in studying multisite polaron models is based on the canonical Lang–Firsov transformation for the electron–phonon Fröhlich

Hamiltonian [16]. This transformation ensures diagonalization of the single-site Hamiltonian (1). Then, the transformed Hamiltonian for a multipolaron model is studied within the framework of the perturbation theory with a parameter (assumed to be small) $\lambda^{-1} = zt/E_p$, where t is the modulus of the hopping integral between nearest sites, z is the number of these sites, and E_p is the polaron shift of the occupied ground electron state in the site according to the Lang–Firsov theory. This is a basic approach used in the superconductivity theory using the concept of interstitial bipolarons.

In this section, we will demonstrate that there exist unitarily inequivalent representations for the Fröhlich site polaron. Using these representations leads to far-reaching consequences that provide for the possibility of making a correct choice of the Fock space in the problem under consideration.

A canonical Lang–Firsov transformation for Hamiltonian (1) is [16]

$$\tilde{H}_{\text{site}} = \exp(S) H_{\text{site}} \exp(-S), \tag{2}$$

where

$$S = \sum_{\sigma} a_{\sigma}^{+} a_{\sigma} \sum_{\mathbf{q}} \xi_{\mathbf{q}} (b_{\mathbf{q}} - b_{-\mathbf{q}}^{+}). \tag{3}$$

Using Eqs. (2)∧(3) and the well-known commutation relationships for the operators in (1), we readily arrive at the electron operator transformed to

$$\tilde{a}_{\sigma} = a_{\sigma} \exp\left(-\sum_{\mathbf{q}} \xi_{\mathbf{q}} (b_{\mathbf{q}} - b_{-\mathbf{q}}^{+}) \right) \tag{4}$$

and at the phonon operator transformed to

$$\tilde{b}_{\mathbf{q}} = b_{\mathbf{q}} + \sum_{\sigma} a_{\sigma}^{+} a_{\sigma} \xi_{\mathbf{q}}. \tag{5}$$

The transformed Hamiltonian acquires the form

$$\tilde{H}_{\text{site}} = \sum_{\sigma} (\varepsilon_0 - E_p) a_{\sigma}^{+} a_{\sigma} - 2E_p n_{\uparrow} n_{\downarrow} + \sum_{\mathbf{q}} \hbar\Omega \left(b_{\mathbf{q}}^{+} b_{\mathbf{q}} + \frac{1}{2} \right), \tag{6}$$

where $n_{\sigma} = 0$ or 1 and the polaron shift is

$$E_p = \sum_{\mathbf{q}} \xi_{\mathbf{q}}^2 \hbar\Omega. \tag{7}$$

The polaron shift E_p is a very important parameter in the theory of bipolaron superconductivity [11], where this value is considered as a measure of the electron–phonon coupling in high-T_c materials and as a reference quantity in the analysis of all other interaction parameters (for example, the on-site electron correlation energy).

The eigenstates of $\tilde{H}_{\text{site}}$ (6) are

$$|\tilde{\Psi}\rangle = |n_{\uparrow}, n_{\downarrow}\rangle |\{N_{\mathbf{q}}\}\rangle.$$

Accordingly, using this expression and Eq. (3), we obtain for the eigenstates of H_{site} (1)

$$|\Psi\rangle = |n_{\uparrow}, n_{\downarrow}\rangle \exp(-(n_{\uparrow} + n_{\downarrow})) \sum_{\mathbf{q}} \xi_{\mathbf{q}} (b_{\mathbf{q}} - b_{-\mathbf{q}}^{+}) |\{N_{\mathbf{q}}\}\rangle,$$

which implies that the electron vacuum state in the Lang–Firsov theory,

$$|0\rangle = |0_{\uparrow}, 0_{\downarrow}\rangle, \tag{8}$$

contains only two single-particle electron states corresponding to the electron vacuum state of the unperturbed electron Hamiltonian $H_{\text{site}}^{0} = \sum_{\sigma} \varepsilon_0 a_{\sigma}^{+} a_{\sigma}$. In fact, this choice of the vacuum state with two single-particle electron states for the electron subsystem is inherent in the transformation (2)∧(3), where the operators a_{σ} are considered as the operators of the destruction of the physical particles. In other words, the vacuum state (8) of the Fock state for Hamiltonian (1) in the Lang–Firsov theory is determined only by the electron vacuum state of the unperturbed electron Hamiltonian H_{site}^{0}, while the eigenvectors of this space are constructed by applying operators a_{σ}^{+} upon the vacuum state (8).

A completely rigorous approach to the single-site polaron problem under consideration is to apply methods of the quantum field theory (QFT). In this case, the Fröhlich polaron corresponds to a solution of the many-particle problem in the space of single-particle fermion states. If the Fock space is expanded to two-particle site states, then a solution can emerge that corresponds to the formation of a boson–fermion mixed state on the site.

It should be noted that using the two aforementioned approaches in the model (1) leads to significantly different results. An analysis of these results provides for the possibility of making a correct choice of the Fock space in the polaron problem under consideration. We will carry out such an analysis for model (1) in two cases of a fully occupied site state. Lang and Firsov [16] considered the sites with a single spin-nondegenerate electron level. In this case, transformation (2)∧(3) leads to the following value of the electron energy for the fully occupied site state (i.e., for the initial condition $\langle \Psi_i | a^{+} a | \Psi_i \rangle = 1$):

$$E_{\text{el}} = \varepsilon_0 - E_p. \tag{9}$$

For the spin-degenerate ground state, the electron energy is [11]

$$E_{\text{el}} = (\varepsilon_0 - E_p)(n_{\uparrow} + n_{\downarrow}) - 2E_p n_{\uparrow} n_{\downarrow}. \tag{10}$$

For the initial condition $A = 1$ (fully occupied level), the polaron energy shift (decrease) is $2E_p$ per spin.

Now we will analyze the polaronic level shift in the same cases using the perturbation theory. The system occurs initially in the fully occupied ground state and features no excited electron states. According to Eq. (1), the first diagonal correction to the electron energy is evidently zero. In the second order of the perturbation theory, the energy variation is also zero because no vacant excited states are available. As can be readily seen, the energy corrections in all orders of the perturbation theory for the fully occupied ground level are zero. As will be shown below using the QFT approach, the average electron energy in the fully occupied site state corresponds to ε_0 per spin without any polaron shift.

Polarons belong to the class of Fermi particles (fermions). A polaron state can be presented as a superposition of single-particle fermion states. These spin-degenerate states form a kind of the energy ladder related to multiphonon processes. It is intuitively clear that we may select from the two spin subspaces a manifold of pairs of the single-particle states, such that a change in the electron energy upon scattering (during which the pairs exchange with a single phonon) in each of the two spin pairs will be small as compared to the phonon energy. Then the effective electron interaction will be attractive and a boson state can appear in the site. In this case, the site state can be represented in the following form:

$$|\Psi\rangle = (1-\beta^2)^{1/2}|\Psi_1\rangle + \beta\exp(i\phi)|\Psi_2\rangle, \qquad (11)$$

where β is the spectral weight of the boson state. For $\beta \neq 0$, vacuum state (9) should be extended to take into account the vacuum state of bosons.

When $\beta = 0$, the wavefunction (11) is invariant relative to the gauge transformation $\hat{O} = \exp(i\hat{N}\phi')$ (N is the particle number operator and ϕ' is an arbitrary phase) and the site state represents a normal polaron state. For $\beta \neq 0$, wavefunction (11) is not invariant relative to the above gauge transformation and the site state represents a superposition of the states with different $N = 1, 2$. However, the state (11) corresponding to this spontaneously violated symmetry is a solution to Eq. (1), which is invariant with respect to the gauge transformation. Therefore, the initial invariance must be retained in a certain manner in solution (11). This conservation is reflected by taking into account the initial site state occupancy, which allows the chemical potential in the state (11) to be determined.

Note that the formation of a boson–fermion mixed state (11) can only take place provided that the initial site occupancy is incomplete ($A < 1$). This is evident, since otherwise ($A = 1$) no free single-particle states would be available for the scattering of quasiparticles.

However, the formation of such a boson–fermion mixed site state is hindered by the Coulomb correlations in the site. For small polarons, the effective Bohr radius a_{eff} is on the order of one interatomic distance. Then, assuming the usual value of the high-frequency dielectric constant $\epsilon_{\text{opt}} = 3$ and taking $a_{\text{eff}} = 5$ Å, we may estimate the typical energy of the on-site interaction between localized electrons:

$$U \approx \frac{e^2}{\epsilon_{\text{opt}} a_{\text{eff}}} \approx 0.45 \text{ eV}.$$

A question arises as to what the relationship is between the on-site correlation energy and the electron–phonon coupling energy at which a boson–fermion mixed site state may form. The next section of this paper is devoted to the study of conditions favoring the formation of such boson–fermion mixed site states and to the thermodynamics of these states. An analysis based on the method of Green's functions will be carried out for a paramagnetic state.

2. MODEL FORMULATION

Supplementing Eq. (1) by a term corresponding to the electron correlations at the site, we may represent the Hamiltonian of the single-site polaron model in the following form:

$$H_{\text{site}} = \varepsilon_0(\hat{n}_\uparrow + \hat{n}_\downarrow) + U\hat{n}_\uparrow\hat{n}_\downarrow + \sum \hbar\Omega\xi_q(\hat{n}_\uparrow + \hat{n}_\downarrow)(b_{\mathbf{q}} + b^+_{-\mathbf{q}}) + H_{\text{ph}}, \qquad (12)$$

where H_{ph} is the phonon Hamiltonian that may include an anharmonic part as well.

Evidently, a substance may contain a certain ensemble of such like site (or interstitial) states. Neglecting the hopping between sites, these site states can be considered as independent of each other and the analysis can be reduced to a single-site Hamiltonian (12).

We have determined the initial site state occupancy as

$$2A = \left\langle \Psi_i \left| \sum_\sigma n_\sigma \right| \Psi_i \right\rangle.$$

The wavefunction Ψ_i of the unperturbed electron Hamiltonian for a particle possessing a spin of 1/2 has two components with the spin projections onto the z axis $\sigma = \pm 1/2$ [22]. For $A = 1/2$, the particle may occur in the site state with a certain spin value σ and the wavefunction $\Psi_i(\sigma)$, while the other site state with spin projection $-\sigma$ will be vacant. Since the spin operator commutates with the Hamiltonian (1), this occupancy distribution in the two spin subspaces also holds for the solution to Eq. (1). As will be seen from the results presented below (see Section 3.1), this solution for the site state is characterized by the absence of a polaron shift. For $A = 1/2$, the initial wavefunction of the particle may be also prepared in the state with a normalized wavefunction $\Psi_i = (\Psi_i(\sigma) + \Psi_i(-\sigma))/\sqrt{2}$ [22]. This state is of considerable importance in multisite models with hopping between sites intended to describe the state of

paramagnetic high-T_c materials. For this reason, below we will study a solution to Eq. (12) for an initial wavefunction in a half-occupied site state ($A = 1/2$).

Let us introduce the Matsubara Green's functions (GFs) [23]:

$$g_{\sigma\sigma_1}(\tau) = -\langle T_\tau \tilde{a}_\sigma(\tau)\tilde{\bar{a}}_{\sigma_1}(0)\rangle, \tag{13}$$

$$f^+_{\sigma\sigma_1}(\tau) = \langle T_\tau \tilde{\bar{a}}_\sigma(\tau)\tilde{\bar{a}}_{\sigma_1}(0)\rangle, \tag{14}$$

$$f_{\sigma\sigma_1}(\tau) = \langle T_\tau \tilde{a}_\sigma(\tau)\tilde{a}_{\sigma_1}(0)\rangle, \tag{15}$$

where $\langle\ldots\rangle$ denotes averaging over the Gibbs ensemble.

Using Eqs. (12)–(15), taking into account only the diagrams with noncrossing phonon GFs for self-energy parts (see Section 3.1), and passing to the spectral representation in discrete $\omega_n = (2n + 1)\pi T$ values, we obtain the following set of equations (without losing generality, we take $\varepsilon_0 = 0$):

$$\begin{aligned} &g_{ll}^{-1} g_{\uparrow\uparrow}(\omega_n) + [U g_{\uparrow\downarrow}(0^+) - \Sigma^s_{\uparrow\downarrow}] g_{\downarrow\uparrow}(\omega_n) \\ &- [U f_{\uparrow\downarrow}(0^+) - \Sigma^a_{\uparrow\downarrow}] f^+_{\downarrow\uparrow}(\omega_n) = 1, \end{aligned} \tag{16}$$

$$[U g_{\downarrow\uparrow}(0^+) - \Sigma^s_{\downarrow\uparrow}(\omega_n)] g_{\uparrow\uparrow}(\omega_n) + g_{ll}^{-1} g_{\downarrow\uparrow}(\omega_n) = 0, \tag{17}$$

$$[U f^+_{\downarrow\uparrow}(0^+) - \Sigma^{+a}_{\downarrow\uparrow}(\omega_n)] g_{\uparrow\uparrow}(\omega_n) - f_{ll}^{-1} f^+_{\downarrow\uparrow}(\omega_n) = 0, \tag{18}$$

$$g_{ll} = (i\omega_n - UA - \Sigma^s_{\uparrow\uparrow}(\omega_n) + \mu)^{-1}, \tag{19}$$

$$f_{ll} = (i\omega_n + UA + \bar{\Sigma}^s_{\uparrow\uparrow}(\omega_n) - \mu)^{-1}, \tag{20}$$

where μ is the chemical potential. Then the self-energy parts have the form

$$\begin{aligned} \Sigma^s_{\uparrow\uparrow}(\omega_n) &= -T(\hbar\Omega)^2 \\ &\times \sum_{n_1,\mathbf{q}} \xi^2_{\mathbf{q}} d_{\mathbf{q}}(i\omega_n - i\omega_{n_1}) g_{\uparrow\uparrow}(\omega_{n_1}), \end{aligned} \tag{21}$$

$$\begin{aligned} \bar{\Sigma}^s_{\uparrow\uparrow}(\omega_n) &= -T(\hbar\Omega)^2 \\ &\times \sum_{n_1,\mathbf{q}} \xi^2_{\mathbf{q}} d_{\mathbf{q}}(i\omega_n + i\omega_{n_1}) g_{\uparrow\uparrow}(\omega_{n_1}), \end{aligned} \tag{22}$$

$$\begin{aligned} \Sigma^s_{\downarrow\uparrow}(\omega_n) &= -T(\hbar\Omega)^2 \\ &\times \sum_{n_1,\mathbf{q}} \xi^2_{\mathbf{q}} d_{\mathbf{q}}(i\omega_n - i\omega_{n_1}) g_{\downarrow\uparrow}(\omega_{n_1}), \end{aligned} \tag{23}$$

$$\begin{aligned} \Sigma^{+a}_{\downarrow\uparrow}(\omega_n) &= -T(\hbar\Omega)^2 \\ &\times \sum_{n_1,\mathbf{q}} \xi^2_{\mathbf{q}} d_{\mathbf{q}}(i\omega_n - i\omega_{n_1}) f^+_{\downarrow\uparrow}(\omega_{n_1}), \end{aligned} \tag{24}$$

$$\begin{aligned} \Sigma^a_{\downarrow\uparrow}(\omega_n) &= -T(\hbar\Omega)^2 \\ &\times \sum_{n_1,\mathbf{q}} \xi^2_{\mathbf{q}} d_{\mathbf{q}}(i\omega_n - i\omega_{n_1}) g_{\uparrow\downarrow}(\omega_{n_1}), \end{aligned} \tag{25}$$

where $d_{\mathbf{q}}(i\omega_n - i\omega_{n_1})$ is the phonon propagator.

The function $g_{\uparrow\uparrow}(\omega_n)$ can be presented in the following form [23]:

$$g_{\uparrow\uparrow}(\omega_n) = \int_{-\infty}^{\infty} \frac{\rho_{\uparrow\uparrow}(x)dx}{x - i\omega_n}, \tag{26}$$

where $\rho_{\uparrow\uparrow}(x)$ is the imaginary part of the retarded double-time GF $G^{(R)}_{\uparrow\uparrow}(\omega)$ at a finite temperature

$$\rho_{\uparrow\uparrow}(\omega) = \frac{1}{\pi}\mathrm{Im}G^{(R)}_{\uparrow\uparrow}(\omega) \le 0. \tag{27}$$

Note that in the case of a nonmagnetic state,

$$\rho_{\uparrow\uparrow}(\omega) = \rho_{\downarrow\downarrow}(\omega) = -\rho(\omega),$$

where $\rho(\omega)$ is a real function determining the spectral density per spin for a fermion quasiparticle at the site, which is normalized as

$$\int_{-\infty}^{\infty} \rho(\omega)d\omega = 1.$$

Using expression (26), applying the sum rule to ω_n [23], and taking into account the definition of the site state occupancy A per spin, we obtain

$$A = \lim_{\tau\to 0^+} g_{\sigma\sigma}(-\tau) = \int_{-\infty}^{\infty} dx \rho(x) n(x - \mu), \tag{28}$$

where $n(t) = 1/\{\exp(t/T) + 1\}$. Since the A value is determined by the initial site state occupancy, relationship (28) is essentially an equation for determining the chemical potential μ.

The analytical continuation of the self-energy parts (21) and (22) from a discrete set of points to the entire upper half-plane ω has the following form:

$$\begin{aligned} \Sigma^s_{\uparrow\uparrow}(\omega) &= (\hbar\Omega)^2\xi^2 \int_{-\infty}^{\infty} \rho(x)dx \\ &\times \left[\frac{n(\mu - x) + N}{\omega - x - \Omega + i\gamma} + \frac{n(x - \mu) + N}{\omega - x + \Omega + i\gamma}\right], \end{aligned} \tag{29}$$

$$\begin{aligned} \bar{\Sigma}^s_{\uparrow\uparrow} &= -(\hbar\Omega)^2\xi^2 \int_{-\infty}^{\infty} \rho(x)dx \\ &\times \left[\frac{n(\mu - x) + N}{\omega + x + \Omega - 2\mu + i\gamma} + \frac{n(x - \mu) + N}{\omega + x - \Omega - 2\mu + i\gamma}\right]. \end{aligned} \tag{30}$$

Here $N = 1/\{\exp(\Omega/T) - 1\}$ and ξ is a dimensionless parameter of the electron–phonon coupling given by the formula

$$\xi^2 = \int \frac{\xi_{\mathbf{q}}^2 d\mathbf{q}}{(2\pi)^3},$$

and γ is a positive constant ($\gamma \ll \Omega$) describing the phonon damping.

For a paramagnetic state $g_{\downarrow\uparrow} = g_{\uparrow\downarrow}$ and $\Sigma^s_{\downarrow\uparrow}(\omega_n) = \Sigma^s_{\uparrow\downarrow}(\omega_n)$. In the representation of

$$g_{\downarrow\uparrow}(\omega_n) = \int_{-\infty}^{\infty} \frac{\rho_{sf}(x)dx}{x - i\omega_n} \tag{31}$$

the real spectral density function $\rho_{sf}(x)$ is determined as the imaginary part of the retarded double-time GF $G^{(R)}_{\downarrow\uparrow}(\omega)$ at a finite temperature

$$\rho_{sf}(\omega) = \frac{1}{\pi}\mathrm{Im}G^{(R)}_{\downarrow\uparrow}(\omega), \tag{32}$$

which obeys the condition

$$\int_{-\infty}^{\infty} \rho_{sf}(\omega)d\omega = 0.$$

By the same token, in the representation of

$$f^+_{\downarrow\uparrow}(\omega) = \int_{-\infty}^{\infty} \frac{\rho^+(x)dx}{x - i\omega_n}, \tag{33}$$

the spectral function $\rho^+(x)$ is determined as the imaginary part of the retarded double-time GF $F^{+(R)}_{\downarrow\uparrow}(\omega)$ at a finite temperature

$$\rho^+(\omega) = \frac{1}{\pi}\mathrm{Im}F^{+(R)}_{\downarrow\uparrow}(\omega), \tag{34}$$

which obeys the condition

$$\int_{-\infty}^{\infty} \rho^+(\omega)d\omega = 0.$$

The analytical continuation of the self-energy parts $\Sigma^s_{\downarrow\uparrow}(\omega_n)$ (23) and $\Sigma^{+a}_{\downarrow\uparrow}(\omega_n)$ (24) to the entire upper half-plane ω has the following form:

$$\Sigma^s_{\downarrow\uparrow}(\omega) = -(\hbar\Omega)^2\xi^2 \int_{-\infty}^{\infty} \rho_{sf}(x)dx \times\left[\frac{n(\mu - x) + N}{\omega - x - \Omega + i\gamma} + \frac{n(x - \mu) + N}{\omega - x + \Omega + i\gamma}\right], \tag{35}$$

$$\Sigma^{+a}_{\downarrow\uparrow}(\omega) = -(\hbar\Omega)^2\xi^2 \int_{-\infty}^{\infty} \rho^+(x)dx \times\left[\frac{n(\mu - x) + N}{\omega - x - \Omega + i\gamma} + \frac{n(x - \mu) + N}{\omega - x + \Omega + i\gamma}\right]. \tag{36}$$

The system of equations (16)–(18) involves two spin fluctuation parameters for the paramagnetic state. The first is $\lambda = g_{\downarrow\uparrow}(0^+) = g_{\uparrow\downarrow}(0^+)$. Using the representation (31) and applying the sum rule to ω_n [23], we obtain

$$\lambda = 2T\sum_{n>0} \mathrm{Re}g_{\downarrow\uparrow}(\omega_n),$$

and eventually

$$\lambda = \int_{-\infty}^{\infty} dx\rho_{sf}(x)n(\mu - x) = -\int_{-\infty}^{\infty} dx\rho_{sf}(x)n(x - \mu). \tag{37}$$

The second parameter is

$$\beta = \lim_{\tau \to +0} f^+_{\downarrow\uparrow}(\tau) = 2T\sum_{n>0} \mathrm{Re}\, f^+_{\downarrow\uparrow}(\omega_n).$$

Taking into account representation (33), this expression can be transformed to

$$\beta = \int_{-\infty}^{\infty} dx\rho^+(x)n(\mu - x) = -\int_{-\infty}^{\infty} dx\rho^+(x)n(x - \mu). \tag{38}$$

The β value is the weight of the boson state in a mixed site state.

Now we will write a solution to system (16)–(18) analytically continued to the entire upper half-plane ω:

$$G^{(R)}_{\uparrow\uparrow}(\omega) = \frac{G^{-1}_{ll}}{G^{-2}_{ll} - [\lambda U - \Sigma^s_{\downarrow\uparrow}(\omega)]^2 - F_{ll}G^{-1}_{ll}[\beta U - \Sigma^{+a}_{\downarrow\uparrow}(\omega)]^2}, \tag{39}$$

$$G^{(R)}_{\downarrow\uparrow}(\omega) = \frac{\Sigma^s_{\downarrow\uparrow}(\omega) - \lambda U}{G^{-2}_{ll} - [\lambda U - \Sigma^s_{\downarrow\uparrow}(\omega)]^2 - F_{ll}G^{-1}_{ll}[\beta U - \Sigma^{+a}_{\downarrow\uparrow}(\omega)]^2}, \tag{40}$$

$$F^{+(R)}_{\downarrow\uparrow}(\omega) = \frac{F_{ll}G^{-1}_{ll}[\beta U - \Sigma^{+a}_{\downarrow\uparrow}(\omega)]}{G^{-2}_{ll} - [\lambda U - \Sigma^s_{\downarrow\uparrow}(\omega)]^2 - F_{ll}G^{-1}_{ll}[\beta U - \Sigma^{+a}_{\downarrow\uparrow}(\omega)]^2}, \tag{41}$$

where $G^{-1}_{ll} = \omega - AU - \Sigma^s_{\uparrow\uparrow}(\omega)$ and $F^{-1}_{ll} = \omega + AU + \bar{\Sigma}^s_{\uparrow\uparrow}(\omega) - 2\mu$.

In the case when the initial site states are half-occupied ($A = 1/2$), a simple analysis of the above solution shows that the chemical potential is constant ($\mu = U/2$),

the spectral density $\rho(x)$ is an even function with respect to μ,

$$\rho(x-\mu) = \rho(\mu-x),$$

while the spectral functions $\rho_{sf}(x)$ and $\rho^+(x)$ are odd,

$$\rho^+(x-\mu) = -\rho^+(\mu-x),$$
$$\rho_{sf}(x-\mu) = -\rho_{sf}(\mu-x).$$

Thus, we obtained a solution to Eq. (12) describing the boson–fermion mixed state (39)–(41). This solution determines the spectral densities ρ (27), ρ_{sf} (32), and ρ^+ (34). These densities set the self-energy parts (29), (30), (35), and (36), which, in turn, determine the solution (39)–(41).

This closed system was solved by means of an iterative procedure. Up to 100 iterations were necessary, after which a maximum relative variation of the spectral density peaks did not exceed 10^{-3} and a change in the λ and β values did not exceed 3×10^{-5} for the two last iterations.

3. RESULTS AND DISCUSSION

The solution (39)–(41) will be analyzed in four particular cases. The on-site electron correlations can be considered within the framework of the Hartree–Fock approximation with neglect of the spin fluctuations ($\lambda = 0$), and the solution can be studied both in the space of single-particle electron state for $\beta = 0$ (normal polaron state) and in the space of boson–polaron mixed states for $\beta \neq 0$. Then we will consider the case with the spin fluctuations taken into account ($\lambda \neq 0$) and study this solution both in the space of the single-particle electron state for $\beta = 0$ (spin-fluctuating polaron state) and in the space of spin-fluctuating boson–polaron mixed states for $\beta \neq 0$.

In the results presented below for $A = 1/2$, the energy variable ω is counted from the $\varepsilon_0 + UA$ level. In the general case, the electron energy of a given state was calculated as

$$\tilde{E} = -\tilde{E}_p - U(A^2 - \lambda^2 + \beta^2), \quad (42)$$

where

$$\tilde{E}_p = -\int_{-\infty}^{\infty} \omega\rho(\omega)n(\omega-\mu)d\omega \quad (43)$$

is a polaron shift for the normal polaron state.

Below we will present the spectral densities for the above states with various initial site state occupancies. In the case of half-occupied initial states, we will consider a phase diagram of these states depending on the temperature.

3.1. The Polaron State at $A = 1$

When finding solution (39)–(41), we neglected the diagrams of higher orders ($n \geq 4$) with intersecting phonon GFs for irreducible self-energy parts. Therefore, the parameters in this solution have to be selected so as to provide that the diagrams would not lead to any significant variation of the spectral densities. In this section, we will calculate the irreducible self-energy parts of the second and fourth orders at $T = 0$ for the normal polaron state ($U = 0$, $\beta = 0$) and present an expression for the spectral density in this state for the completely occupied state ($A = 1$).

The corresponding self-energy parts have the following form:

$$\Sigma^{(2)}(\omega) = i(\hbar\Omega)^2\xi^2\int_{-\infty}^{\infty}\frac{d\omega_1}{2\pi}D(\omega-\omega_1)G(\omega_1), \quad (44)$$

$$\Sigma^{(4)}(\omega) = -(\hbar\Omega)^4\xi^4\int\frac{d\omega_1}{2\pi}\int\frac{d\omega_2}{2\pi}D(\omega-\omega_1) \times G(\omega_1)G(\omega_2)D(\omega_1-\omega_2)G(\omega-\omega_1+\omega_2), \quad (45)$$

where

$$D(\omega) = \frac{1}{\omega-\Omega+i\gamma} - \frac{1}{\omega+\Omega-i\gamma}. \quad (46)$$

The total electron GF for $A = 1$ has the following spectral representation ($\delta \longrightarrow +0$):

$$G(\omega) = \int_{-\infty}^{\infty}\frac{\rho(x)dx}{\omega-x-i\delta}. \quad (47)$$

Substituting expressions (46) and (47) into (44) and (45), we obtain

$$\Sigma^{(2)}(\omega) = (\hbar\Omega)^2\xi^2\int_{-\infty}^{\infty}\frac{\rho(x)dx}{\omega-x+\Omega-i\gamma}, \quad (48)$$

$$\Sigma^{(4)}(\omega) = (\Sigma^{(2)}(\omega))^2\int_{-\infty}^{\infty}\frac{\rho(x)dx}{\omega-x+2\Omega-2i\gamma}. \quad (49)$$

The spectral density of the polaron state for $A = 1$ with the self-energy part approximated as $\Sigma(\omega) = \Sigma^{(2)}$ (48) is presented by a solid curve in Fig. 1. The chemical potential is everywhere above the spectral density of the polaron state. Clearly distinguished in Fig. 1 are seven peaks in the $\rho(\omega)$ structure, which correspond to multiphonon processes involved in polaron-state formation. Note that these peaks are not equidistant, their shifts being determined by the real part of (48). For the low-energy peaks, the distance between the neighboring peaks gradually decreases, approaching the phonon energy $\Omega = 30$ meV.

For the spectral density depicted by the solid curve in Fig. 1, expression (43) yields $\tilde{E}_p = 0$. Thus, the

Fröhlich polaron with $A = 1$ exhibits no polaron energy shift, which is a quite natural result from the standpoint of the quantum-mechanical approach (see section 1). It should be emphasized that $\tilde{E}_p = 0$ at $A = 1$ for any values of the model parameters.

The dashed curve in Fig. 1 presents the spectral density of the polaron state calculated for the self-energy part in the form $\Sigma(\omega) = \Sigma^{(2)} + \Sigma^{(4)}$. A comparison of the dashed and solid curves in Fig. 1 shows that the two spectral densities exhibit a qualitatively similar structure of peaks. The peaks are somewhat shifted in energy due to the real part of $\Sigma^{(4)}$ (49). At the same time, an allowance for the contribution due to the lowest-order diagram with intersecting phonon GFs leads (for the model parameters employed) to the appearance of a negative spectral density within a narrow energy region (Fig. 1). As the electron–phonon coupling parameter ξ increases, this region of the nonphysical density of states grows and new negative-density regions appear between the peaks where $\mathrm{Im}\Sigma^{(2)}$ is small. This was accompanied by a significantly deteriorated convergence of the iterative procedure. Nevertheless, for the spectral density presented by the dashed curve in Fig. 1, Eq. (43) also yields $\tilde{E}_p = 0$.

In the HTSC model [11] based on the concept of interstitial bipolaron formation, a very important model parameter is the polaron shift. We have determined for $A = 1$ that the average energy of the Fröhlich polaron state (which is precisely the polaron shift for the energy scale adopted) is $\tilde{E}_p = 0$ to within the fourth-order term ξ^4 with respect to the electron–phonon coupling parameter. At the same time, an expression of the second order $E_p = \xi^2\hbar\Omega$ (see also Eq. (11) in [11]) for the polaron shift yields $E_p = 1/3$ eV in the case of the Fröhlich polaron in the site with $A = 1$. We believe that the latter polaron energy value is incorrect, which is related to an incorrect representation of the Fock space for polarons employed in the Lang–Firsov theory.

In the case of an unoccupied site state ($A = 0$), we can use representation (47) upon substitution $-i\delta \longrightarrow +i\delta$, which leads to the substitution $\Omega - i\gamma \longrightarrow -\Omega + i\gamma$ in expressions (48) and (49). As a result, $\rho(\omega)$ for the polaron state at $A = 0$ corresponds to the density of states for the $A = 1$ mirror reflected relative to the axis $\omega = 0$.

If the initial electron wavefunction at the site is such that the electron occurs in a state with definite spin projection (e.g., $\sigma = 1/2$) and the second site state ($-\sigma$) is empty, the occupancy distribution in the two spin subspaces will be retained for a solution to Eq. (1) as well (because the spin operator commutates with Hamiltonian (1)). Therefore, $\rho_{\sigma\sigma}$ represents the spectral density depicted in Fig. 1 and $\rho_{-\sigma,-\sigma}(\omega - \mu) = \rho_{\sigma\sigma}(\mu - \omega)$. In this case, the polaron shift is obviously equal to zero.

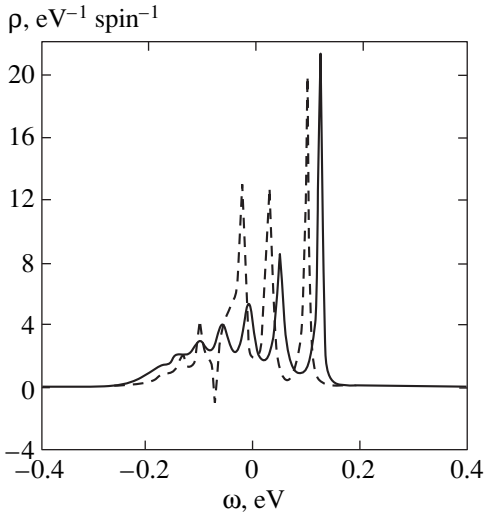


Fig. 1. The spectral density of polaron states in the case of the complete initial site occupancy ($A = 1$) calculated for $T = 0$ using the self-energy part approximated as $\Sigma(\omega) = \Sigma^{(2)}$ (solid curve) and $\Sigma(\omega) = \Sigma^{(2)} + \Sigma^{(4)}$ (dashed curve). The model parameters: $\xi = 10/3$; $\Omega = 30$ meV; $\gamma = 0.1\Omega$.

3.2. Half-Filled Polaron States at Finite Temperatures

Figure 2 shows the spectral densities of polaron states $\rho(\omega)$ for $A = 1/2$ and $T = 20$ K calculated for two different sets of model parameters. As seen, both curves exhibit a central peak at $\omega = 0$ with the spectral density peaks corresponding to multiphonon processes symmetrically arranged on the left and right wings. These peaks, the number of which increases with the electron–phonon coupling parameter (Fig. 2), are not equidistant, their shift being determined by the real part of the self-energy component (29). The distance between the neighboring peaks decreases with increasing peak number, eventually approaching the phonon energy.

In the case of half-filling ($A = 1/2$), the chemical potential occurs at the central peak ($\mu = 0$) and the decrease in the electron energy related to the polaron effect (43) is $\tilde{E}_p = 38.84$ meV/spin for the solid curve and $\tilde{E}_p = 58.15$ meV/spin for the dashed curve (Fig. 2). At the same time, a decrease in the electron energy calculated by the formula $E_p = \hbar\Omega\xi^2 A$ is 1/6 eV/spin for the parameters corresponding to the solid curve in Fig. 2 and $E_p = 273.8$ meV/spin for the parameters corresponding to the solid curve in Fig. 2. These values are significantly overstated as compared to the $\tilde{E}_p$ values indicated above.

3.3. Boson–Polaron Mixed Site State

In the Hartree–Fock approximation ($\lambda = 0$), the polaron sate at $A = 1/2$ (Fig. 2) corresponds to a solution of Eq. (12) in the space of single-particle electron

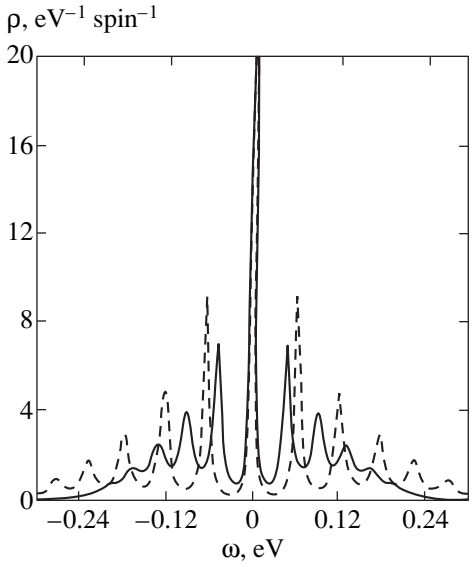


Fig. 2. The spectral density of polaron states in the case of the half-filled initial site states ($A = 1/2$) for $T = 20$ K. The solid curve was calculated using the same model parameters as in Fig. 1; the dashed curve was calculated for $\xi = 3.7$; $\Omega = 40$ meV; $\gamma = 0.05\Omega$. The maximum density of states at $\omega = 0$ is 35.54 eV^{-1} spin^{-1} (solid curve) and 45.47 eV^{-1} spin^{-1} (dashed curve).

states, that is, in the case when $\beta = 0$. Actually, this situation takes place for the completely occupied site states. In the case of half-filling ($A = 1/2$) and a temperature below a certain critical level ($T < T_\beta$), a boson–polaron mixed state possesses a lower energy as compared to that of the polaron state. The boson–polaron mixed state is defined by the single-particle density $\rho(\omega)$ and the spectral density of the intrasite boson state $\rho^+(\omega)$. Figure 3 shows the spectral densities of the boson–polaron mixed states at two temperatures. It was found that the chemical potential of this system for $A = 1/2$ is independent of the temperature ($\mu = 0$).

The solid curves in Figs. 3a and 3b correspond to the mixed state at $T = 20$ K. A decrease in the electron energy is $-\tilde{E} = 73.2$ meV. The boson–polaron state formation leads to splitting of the central peak (see Fig. 2) into two components spaced by $\Delta = 92.2$ meV (Fig. 3a). The chemical potential μ occurs in the middle of the gap, that is, in the region where the spectral density $\rho(\omega)$ is almost zero. There are four symmetrically arranged (but not equidistant) peaks on both left and right sides of μ.

The spectral density of the on-site boson state $\rho^+(\omega)$ is an odd function of ω (Fig. 3b). For the half-filled site state, $\mu = 0$ and, consequently, the β value determined by formula (38) is finite ($\beta = 0.1606$). The dashed curves in Figs 3a and 3b correspond to this state at $T = 120$ K (i.e., near T_β). Here, the decrease in the electron energy is $-\tilde{E} = 65.0$ meV and $\beta = 0.0967$.

Note that, in the case of completely occupied site states ($A = 1$), the chemical potential falls on the right of the energy region of the spectral density $\rho(\omega)$ distribution. Since

$$\int_{-\infty}^{\infty} \rho^+(\omega)d\omega = 0,$$

this always corresponds to $\beta = 0$ and, hence, the system occurs in the normal polaron state (Fig. 1).

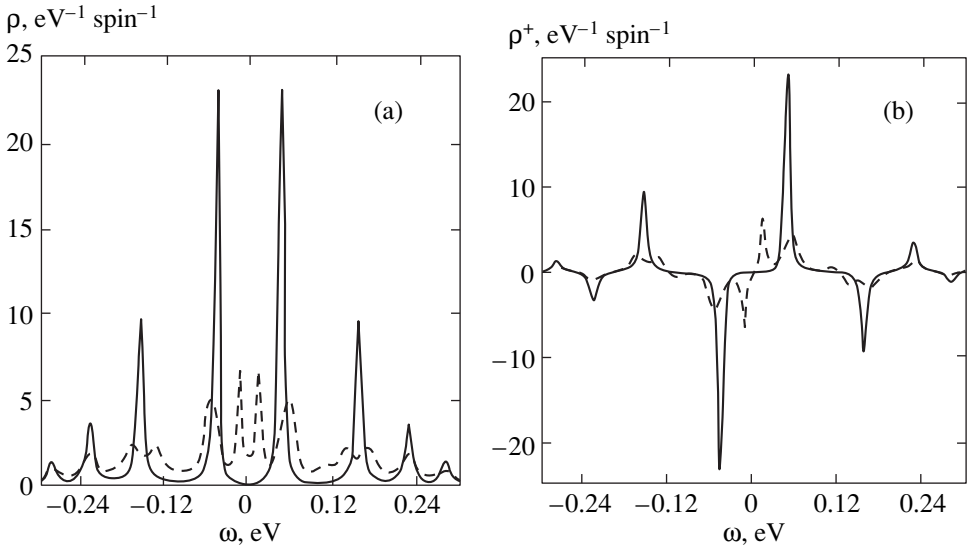


Fig. 3. The spectral density of boson–polaron mixed site states calculated for $T = 20$ K (solid cures) and 120 K (dashed curves) in the case of half-filling ($A = 1/2$): (a) single-particle states; (b) boson states. The model parameters: $\xi = 3.7$; $\Omega = 40$ meV; $\gamma = 0.05\Omega$.

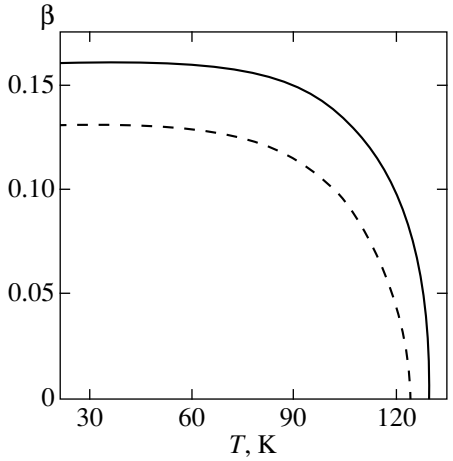


Fig. 4. Temperature dependence $\beta(T)$ of the spectral weight of a two-particle site state for $A = 1/2$. The solid curve was calculated using the same model parameters as in Fig. 3; the dashed curve was calculated for $\xi = 5.774$; $\Omega = 30$ meV; $\gamma = 0.033\Omega$.

Figure 4 illustrates the temperature-induced transition from a boson–polaron mixed state to the polaron state for two sets of parameters. The character of this transition is determined by a temperature dependence of the spectral weight of the boson site state $\beta(T)$. As is seen in Fig. 4, there is a certain critical temperature T_β at which $\beta = 0$, which corresponds to the transition from a boson–polaron to a polaron state. As $T \longrightarrow T_\beta$, the energy gap vanishes ($\Delta \longrightarrow 0$), the two spectral density peaks in the vicinity of the chemical potential μ merge together, and the amplitudes of $\rho^+(\omega)$ peaks tend to zero (see the dashed cures in Figs. 3a and 3b). For $T \geq T_\beta$, the solution corresponds to only a normal polaron state (Fig. 2).

It was established that the temperature dependence of β in the vicinity of T_β is described by the function $\beta = \beta_0\sqrt{1 - T/T_\beta}$. For a set of parameters corresponding to the solid curve in Fig. 4, $\beta_0 = 0.4336$ and $T_\beta = 129.85$ K. For the dashed curve in Fig. 4, $\beta_0 = 0.2354$ and $T_\beta = 124.63$ K. Note that the electron–phonon coupling parameter for the solid curve is smaller, and the phonon energy is greater, than the analogous values for the dashed curve, while the critical temperatures in the two cases are close.

3.4. Spin-Fluctuating Polaron State

Now let us consider the polaron states with an allowance for the spin fluctuations in the site with half-filling ($A = 1/2$). Assuming $\beta = 0$, the spin-fluctuating polaron state appears at temperatures below a certain critical level ($T < T_\lambda$) and is characterized by the spin fluctuation parameter $\lambda \leq 1/2$. The spin-fluctuating polaron state is determined by the single-particle density of states $\rho(\omega)$ and the spectral density of spin fluctuations $\rho_{sf}(\omega)$. Figure 5 shows the spectral density of the spin-fluctuating polaron state at two temperatures. It was found that the chemical potential of this state is independent of the temperature ($\mu = 0$).

Using Eqs. (39)–(41) with $\beta = 0$, it is possible to show that the presence of spin fluctuations leads to doubling of the spectral features in $\rho(\omega)$. As a result, the central peak of the polaron sate (see Fig. 2) splits into two peaks separated by the energy gap Δ. The solid curves in Figs 5a and 5b represent the spin-fluctuating polaron state at $T = 20$ K, which corresponds to the spin

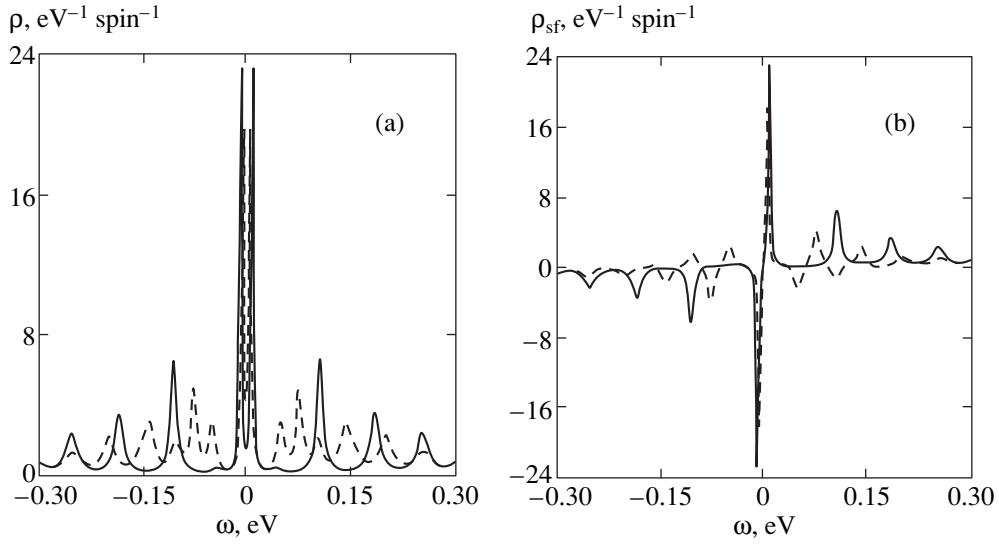


Fig. 5. The spectral density of spin-fluctuating polaron states calculated at $T = 20$ K (solid curves) and 87 K (dashed curves) in the case of the half-occupied initial site states ($A = 1/2$): (a) single-particle states; (b) spin fluctuations. The model parameters are the same as in Fig. 3, and $U = 0.4$ eV.

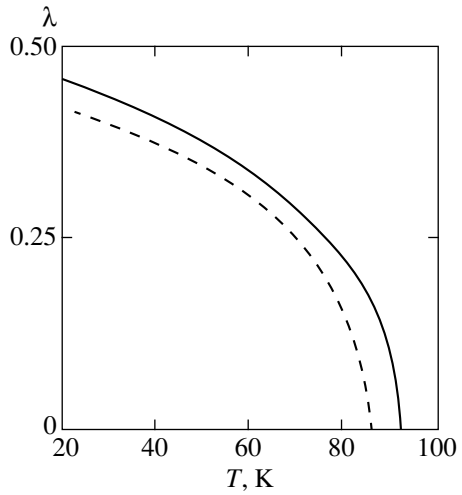


Fig. 6. Temperature dependence $\lambda(T)$ of the spectral weight of a spin-fluctuating polaron state for $A = 1/2$. The solid curve was calculated using the same model parameters as in Fig. 5; the dashed curve was calculated for $\xi = 5.774$; $\Omega = 30$ meV; $\gamma = 0.033\Omega$; $U = 0.5$ eV.

fluctuation parameter $\lambda = 0.4552$ and the gap width $\Delta = 16.32$ meV, with the chemical potential "occurring" at the middle of the gap. The dashed curves in Fig. 5a and 5b illustrate the case of $T = 87$ K, with $\lambda = 0.1585$ and $\Delta = 10.88$ meV.

Note that, with the initial condition $A = 1/2$, the function $\rho(\omega)$ is even (Fig. 5a), while $\rho_{sf}(\omega)$ is odd (Fig. 5b) relative to $\mu = 0$, which results in that the λ is finite for $T < T_\lambda$. For the initial condition $A = 1$, the chemical potential falls on the right of the energy region of the spectral density $\rho(\omega)$ distribution. In this case, we obtain only a trivial solution of Eq. (37), $\lambda = 0$, whereby the system with $A = 1$ is always in the normal polaron state (Fig. 1).

As the temperature increases, the spin fluctuations decay (Fig. 5) and the λ value decreases. Figure 6 shows the temperature dependence of the spin fluctuation parameter $\lambda(T)$ calculated for a spin-fluctuating polaron state with $A = 1/2$ using two sets of parameters. This dependence determines the transition from the spin-fluctuating to the normal polaron state observed with increasing temperature. As seen, there is a certain critical temperature T_λ at which λ turns zero, which corresponds to the transition. As $T \longrightarrow T_\lambda$, the energy gap vanishes ($\Delta \longrightarrow 0$), the two spectral density peaks $\rho(\omega)$ in the vicinity of the chemical potential μ merge together, and the amplitudes of $\rho_{sf}(\omega)$ peaks tend to zero (see the dashed cures in Figs. 5a and 5b). For $T \geq T_\beta$, there is a single solution corresponding to a normal polaron state (Fig. 2).

3.5. Spin-Fluctuating Boson–Polaron Mixed State

The appearance of the energy gap in the single-particle spectral density $\rho(\omega)$ is related both to the formation of an intrasite boson state (Fig. 3) and to the spin fluctuations leading to the aforementioned doubling of the spectral density features (Fig. 5). It was established that there may exist a temperature interval in which the parameter λ for the spin-fluctuating polaron state is finite and the spectral weight β of the boson–polaron mixed state is nonzero. It was also found that, in the temperature range where the spin fluctuation parameter λ is nonzero, the energy of a spin-fluctuating boson–polaron mixed state is lower than that of the simple spin-fluctuating polaron state. This is related to a strong mutual influence of the two factors, λ and β, in the case of the former state formation. The spin-fluctuating boson–polaron mixed state is characterized by the density of single-particle states $\rho(\omega)$, the spectral density of spin fluctuations $\rho_{sf}(\omega)$, and the spectral density of the two-particle site state $\rho^+(\omega)$.

Figure 7 shows the spectral densities for the spin-fluctuating boson–polaron mixed states at $A = 1/2$ calculated for $T = 20$ and 70 K. At $T = 20$ K, the electronic energy of the spin-fluctuating polaron state is $\tilde{E} = -110.52$ meV, while the spin-fluctuating boson–polaron mixed state is characterized by $\tilde{E} = -141.83$ meV. As the temperature increases, the energy of the former state grows (to $\tilde{E} = -140.4$ meV at $T = 70$ K).

As is seen in Fig. 7a, the spectrum of single-particle states exhibits an energy gap at $\mu = 0$. This energy gap decreases from $\Delta = 45.34$ meV at $T = 20$ K to 18.13 eV at $T = 70$ K. The spectral density of spin fluctuations depicted in Fig. 7b shows that the spin fluctuation parameter at $T = 20$ K is significantly lower for the spin-fluctuating boson–polaron mixed state ($\lambda = 0.2904$) than for the spin-fluctuating polaron state ($\lambda = 0.4552$). As the temperature grows, the spin fluctuations decrease. For $T = 70$ K, the spin fluctuation parameter for the spin-fluctuating boson–polaron mixed state ($\lambda = 0.2741$) is still somewhat lower than that ($\lambda = 0.2881$) for the spin-fluctuating polaron state (Fig. 6). Figure 7c shows the spectral density of the two-particle site state. At $T = 20$ K, the spectral weight of the boson state ($\beta = 0.0526$) is lower than the value ($\beta = 0.1606$) for the boson–polaron mixed state (Fig. 3). The β value of the spin-fluctuating boson– polaron mixed state monotonically decreases when the temperature increases up to a certain level ($\beta = 0.01527$ at $T = 70$ K); the boson–polaron state at this temperature has $\beta = 0.1582$ (Fig. 4).

For the model parameters employed, the critical temperature T_β for the transition from boson–polaron to polaron state (Fig. 4) is greater than the T_λ value corresponding to the transition from a spin-fluctuating to normal polaron state (Fig. 6). This interplay leads to the unusual phase diagram depicted in Fig. 8. In the spin-fluctuating boson–polaron mixed state, the mutual influence of the spin fluctuations and the channel of the boson site state formation leads to a decrease both in λ (as compared to the value in the spin-fluctuating polaron state) and in β (as compared to the value in boson–polaron mixed state). At a certain temperature,

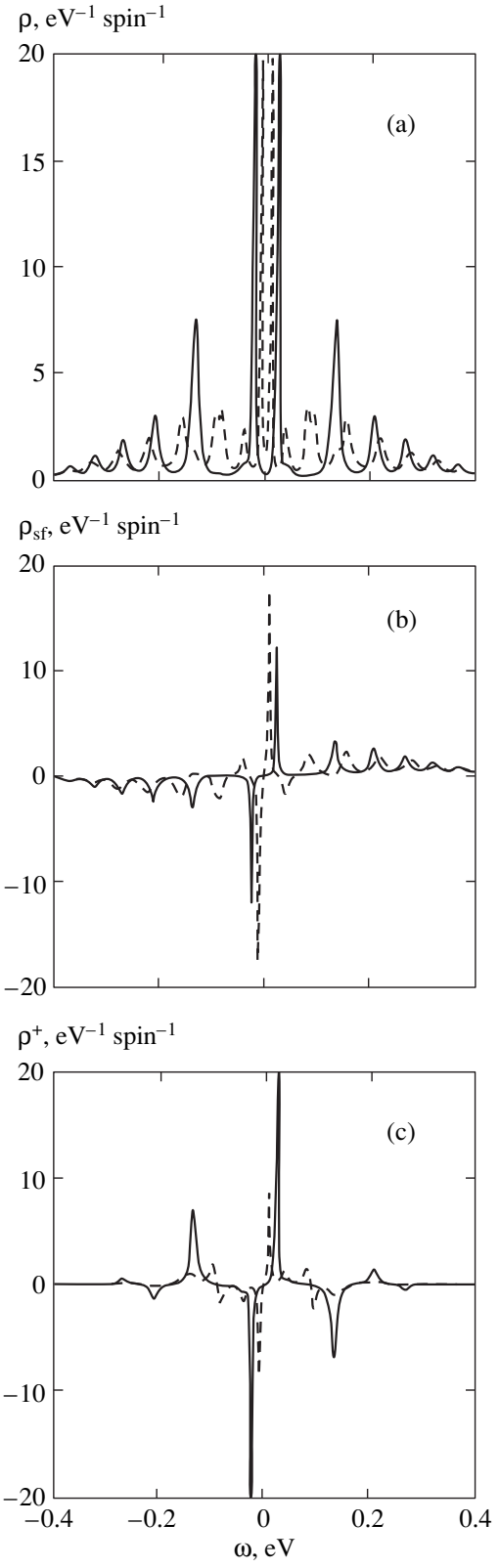


Fig. 7. The spectral density of the spin-fluctuating boson–polaron mixed state at $T = 20$ K (solid curves) and $T = 70$ K (dashed curves): (a) single-particle states; (b) spin fluctuations; (c) two-particle site state. The model parameters are the same as in Fig. 5.

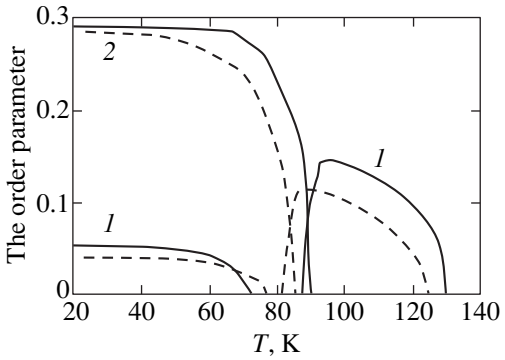


Fig. 8. A phase diagram of the order parameter versus temperature involving (*1*) $\beta(T)$ and (*2*) $\lambda(T)$ curves. The model parameters for the solid and dashed curves are the same as in Figs. 5 and 6, respectively.

the electron correlations completely suppress the boson formation channel ($T \approx 73$ and 77 K for the solid and dashed curves *1* in Fig. 8, respectively). At this point, $\beta = 0$ and the system features a transition from the spin-fluctuating boson–polaron mixed state to the spin- fluctuating polaron state. As the temperature grows further, the spin fluctuations and, accordingly, the λ value decrease (Fig. 8). Even before λ turns zero, the spin fluctuations become too weak to suppress the boson formation channel and β becomes nonzero. Thus, there is a certain temperature ($T \approx 87$ and 83 K for the solid and dashed curves *1* in Fig. 8, respectively) at which the system features a transition from the spin-fluctuating polaron state to the spin-fluctuating boson–polaron state. The critical temperature at which $\lambda \longrightarrow 0$ (see the solid and dashed curves *2* in Fig. 8) for the spin-fluctuating boson–polaron state is somewhat lower than the value of T_λ for the transition from a spin-fluctuating polaron to normal polaron state (Fig. 6). At this temperature, we observe a transition from the spin-fluctuating boson–polaron state to boson–polaron state (Fig. 8). Above this temperature level, $\beta(T)$ keeps increasing to approach the value for the boson–polaron state (Fig. 4). Finally, at $T \longrightarrow T_\beta$, the system is characterized by $\beta \longrightarrow 0$ and $\Delta \longrightarrow 0$ and we observe the transition from boson–polaron to polaron state.

In concluding, it should be noted that a superconducting state in multisite polaron models used to describe the phase diagrams of high-T_c materials appears provided that extended charged bosons are formed in the system. The formation of intrasite boson–polaron mixed states might well be a precursor of such a superconducting state. Taking into consideration an ensemble to such site states with an allowance for the particle hopping between the sites, the single-particle transitions over an ensemble of the sites would lead to the appearance of extended fermion states (equivalent to the insulator–metal transition upon reaching a certain threshold doping level), while the two-particle transitions of initially localized charged bosons between the site

states may result in the formation of extended charged boson states [24].

In this approach to the HTSC description, it is possible to take into account a hybridization between the initial localized single-particle site states and the band states in the parent insulator, which leads already in the second order of the hybridization matrix element to the particle hopping over the ensemble of sites [25]. This behavior is related to the following facts. First, in contrast to the tunneling (or jumping) mechanism, there is no exponential smallness related to overlap of the wavefunctions of small polarons in the neighboring sites. Second, the data of angle-resolved X-ray photoelectron spectroscopy indicate that the symmetry of the wavefunctions of extended single-particle states near the Fermi surface in doped high-T_c materials is close to the symmetry of the wavefunctions of the initial cuprates near the dielectric gap [26]. The hybridization between the initial localized site states and the band states in the parent insulator provides for a simple explanation of the wavefunction symmetry of the extended charged states in narrow polaron bands.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research (project no. 99-02-16292), by the Soros Higher Education Program (grant HBA no. 802), and by the MMTsFPI (grant no. 9-1998).

REFERENCES

1. H. Frölich, S. Machlup, and T. K. Mitra, Phys. Condens. Mater. **1**, 359 (1963).
2. *Polarons*, Ed. by Yu. A. Firsov (Nauka, Moscow, 1975).
3. D. Emin and T. Holstein, Phys. Rev. Lett. **36**, 323 (1976).
4. B. K. Chakraverty, M. J. Sienko, and J. Bonnerot, Phys. Rev. B **17**, 3781 (1978).
5. H. de Raedt and Ad. Lagendijk, Phys. Rev. B **27**, 6097 (1983).
6. H. Fehske, H. Roder, G. Wellein, and A. Mistriotis, Phys. Rev. B **51**, 16582 (1995).
7. A. S. Mishchenko, N. V. Prokof'ev, A. Sakamoto, and B. V. Svistunov, Phys. Rev. B **62**, 6317 (2000).
8. P. E. Kornilovitch and E. R. Pike, Phys. Rev. B **55**, R8634 (1997).
9. K. A. Muller, J. Supercond. **12**, 3 (1999).
10. A. S. Alexandrov and N. F. Mott, Rep. Prog. Phys. **57**, 1197 (1994).
11. A. S. Alexandrov, Phys. Rev. B **61**, 12315 (2000).
12. E. V. L. de Mello and J. Ranninger, Phys. Rev. B **55**, 14872 (1997).
13. B. K. Chakraverty, J. Ranninger, and D. Feinberg, Phys. Rev. Lett. **82**, 2621 (1999).
14. Z.-X. Shen and J. R. Schrieffer, Phys. Rev. Lett. **78**, 1771 (1997).
15. H. Umezawa, H. Matsumoto, and M. Tachiki, *Thermo-Field Dynamics and Condensed States* (North-Holland, Amsterdam, 1982; Mir, Moscow, 1985).
16. I. G. Lang and Yu. A. Firsov, Zh. Éksp. Teor. Fiz. **43**, 1843 (1962) [Sov. Phys. JETP **16**, 1301 (1963)].
17. Yu. A. Firsov and E. K. Kudinov, Fiz. Tverd. Tela (St. Petersburg) **39**, 2159 (1997) [Phys. Solid State **39**, 1930 (1997)].
18. M. Capone, W. Stephan, and M. Grilli, Phys. Rev. B **56**, 4484 (1997).
19. E. V. L. de Mello and J. Ranninger, Phys. Rev. B **58**, 9098 (1998).
20. M. Zoli, Phys. Rev. B **61**, 14523 (2000).
21. S. Ciuchi, F. de Pasquale, S. Fratini, and D. Feinberg, Phys. Rev. B **56**, 4494 (1997).
22. L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-Relativistic Theory* (Nauka, Moscow, 1989, 4th ed.; Pergamon, Oxford, 1977, 3rd ed.).
23. A. A. Abrikosov, L. P. Gor'kov, and I. E. Dzyaloshinskiĭ, *Methods of Quantum Field Theory in Statistical Physics* (Fizmatgiz, Moscow, 1962; Prentice-Hall, Englewood Cliffs, 1963).
24. A. I. Agafonov and É. A. Manykin, Zh. Éksp. Teor. Fiz. **117**, 182 (2000) [JETP **90**, 160 (2000)].
25. A. I. Agafonov and E. A. Manykin, Phys. Rev. B **52**, 14571 (1995).
26. Z.-X. Shen and D. S. Dessau, Phys. Rep. **253**, 1 (1995).

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 92, No. 4, 2001, pp. 677–695.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 4, 2001, pp. 777–798.

SOLIDS
Electronic Properties

Evolution of the Fermi Surface of Cuprates on the Basis of the Spin–Polaron Approach

A. F. Barabanov[a, *], A. A. Kovalev[a], O. V. Urazaev[a], A. M. Belemuk[a], and R. Hayn[b]

[a]*Institute of High-Pressure Physics, Russian Academy of Sciences, Troitsk, Moscow oblast, 142190 Russia*
[b]*Institute for Solid State and Materials Research (IFW), P.O. Box 270016, D-01171 Dresden, Germany*
**e-mail: abarab@ns.hppi.troitsk.ru*

Received November 3, 2000

Abstract—A spin-polaron approach is proposed for describing the evolution of the Fermi surface (FS) and the electronic structure of cuprates with various doping levels on the basis of the spin-fermion model. The complexity of the internal structure of a spin polaron is taken into account by introducing the superposition of spin-polaron states with various radii, while doping is effectively described by frustration in the spin Hamiltonian. The calculations of the polaron spectrum, the spectral weight of bare charge carriers, and the Fermi surface demonstrate radical changes in the electronic structure upon an increase of the doping level. The results obtained make it possible to use a unified approach for describing the experimental data on photoemission such as the isotropic bottom of the band and the residual Fermi surface of undoped compounds, the large Fermi surface and the extended saddle-type singularity for optimally doped compounds, as well as the pseudogap in the case of intermediate doping. *© 2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The problems associated with the topology and evolution of the Fermi surface (FS) in *p*-doped high-temperature superconductor are being intensely studied at present. The results of recent experiments on photoemission with angular resolution (ARPES) indicate the difference between the spectra of optimally doped cuprates and the spectra of insulator cuprates. The experimental data obtained for undoped compounds [1, 2] indicate the isotropic bottom of the band near point $(\pi/2, \pi/2)$ in the momentum space, the large difference in the energies corresponding to points $N = (\pi/2, \pi/2)$ and $X = (\pi, 0)$, and the so-called remanent FS (the surface at which an abrupt decrease in the one-particle spectral weight takes place) [3]. In optimally doped cuprates [4–10], a plane zone region, a large FS with the center at point $M = (\pi, \pi)$ and the so-called shadow FS emerging as a result of antiferromagnetic spin correlations are observed [11]. The plane zone region has the shape of a saddle point extended in the $(\pi/2, 0)$–$(\pi, 0)$ direction. The shadow FS resembles the main FS, but is displaced by the antiferromagnetic vector $\mathbf{Q} = (\pi, \pi)$. Besides, a high-energy pseudogap is observed for intermediate doping near point X with an energy of the order of 0.1 to 0.2 eV [2, 12–14]. In the rigid zone model, the isotropic minimum in the spectrum of undoped compounds must be transformed into small hole pockets near point N upon an increase in the doping level. However, the experimental data indicate the absence of such pockets [14], and the FS apparently has the shape of an arc for the optimal and intermediate doping levels [15]. Such an obvious contradiction can be resolved only by analyzing the evolution of the spectral density under doping.

Many publications are devoted to an analysis of the hole spectrum of a two-dimensional (2D) doped antiferromagnet on the basis of various theoretical models such as the generalized *t–J* model, the effective three-band model, the Kondo-lattice model, and the Hubbard model. Most of these works are based on exact diagonalization of small clusters [16, 17], the application of the quantum-mechanical Monte Carlo method [18] and the self-consistent Born approximation (SCBA) [19–22], the "string" ansatz for the hole wave function [23], or the diagrammatic methods using the semiphenomenological spin susceptibility [24]. The analysis is usually carried out on the basis of either the two-sublattice (Néel-type) state of the spin subsystem [19, 20, 23, 25] or proceeding from the spherically symmetric state of the spin liquid [26, 27]. This leads to qualitatively similar results: a hole moves mainly over one of the sublattices; i.e., effective jumps to the second nearest neighbors play the dominating role in the spectrum formation, and the spectrum has a minimum at point N for a low doping level.

On the other hand, only a few theoretical works are devoted to the FS evolution depending on the doping level and the frustration of the spin subsystem, i.e., the antiferromagnetic exchange interaction between nearest neighbors (J_1) and the next nearest neighbors (J_2). An analysis of the frustrated generalized *t–J* model [26] and the frustrated effective three-band model [27] reveals a strong frustration dependence of the spin–polaron spectrum. It should be noted that these investi-

gations [26, 27] were carried out only for a local polaron (the disadvantages of this approximation will be considered below), and the calculations made do not lead to the spectral weight of a bare hole and describe only certain hypothetical FSs. In a recent work by Sibata *et al.* [28], the frustrated t–t'–t''–J_1–J_2 model was investigated in the self-consistent Born approximation, in which holes were treated as zero-spin fermions and spins as normal bosons [20]. The SCBA formalism in [28] corresponds to the two-sublattice approximation, which imposes the requirement of the spectrum symmetry relative to the reduced magnetic Brillouin zone (which means that points Γ and M are equivalent). However, such a symmetry of the hole spectrum is not observed in the ARPES data.

As regards the calculation of quasi-particle bands on the basis of the exact diagonalization of clusters containing from 16 to 20 sites [28–31], it should be noted that in the case of doping, it is difficult to reconstruct the FS and to study the effect of small variations in the doping level in view of a limited number of **k** points. For this reason, the lowest energy of the quasiparticle energy in some of such calculations corresponds to point X (and not to point M) [31] or the insulator under investigation does not display the presence of a remanent FS [29]. The limited size of clusters apparently does not permit the description of spin-polaron states with a radius exceeding two lattice constants.

In the present paper, we will prove that many ARPES results obtained in a wide range of doping can be explained using the spin-polaron approach taking into account the complex structure of the spin polaron (the superposition of spin-polaron states with various radii [32]) and the frustration in the spin subsystem. We will consider this problem using the effective three-band (spin-fermion) model [33, 34], which explicitly takes into account direct oxygen–oxygen jumps in the CuO_2 plane of cuprates.

A distinguishing feature of our analysis is that we consider a spin polaron as a complex quasiparticle: a coherent superposition of a local polaron (bare hole forming a Zang–Reiss singlet around a Cu^{2+} ion) and a large-radius antiferromagnetic polaron. The latter is a local polaron coupled with a spin wave having a quasi-momentum **q** close to **Q**. It turns out that the introduction of such a basis of spin polaron operators describes the splitting of the lower band of a local polaron [32] (among other things, this makes it possible to reproduce the abrupt decrease in the intensity of ARPES peaks during the motion in **k** from point N to point M).

We will use the projection method for Green's functions of basis operators, which is a sort of the mean-field approximation. The correctness of this method can be verified by comparing the results obtained for $T = 0$ [35] with the spectral function $A_h(\mathbf{k}, \omega)$ of a bare hole, which was obtained using the SCBA for a local spin polaron [36]. It can be seen that the quasiparticle peak obtained in the SCBA and its intensity are correctly reproduced by the lower band of the complex polaron in the projection method approximation. As regards the upper bands corresponding to excited states, they effectively describe the incoherent component of $A_h(\mathbf{k}, \omega)$.

Let us consider the approach adopted by us for describing the spin subsystem. The typical behavior of the magnetic 2D subsystem corresponds to undoped CuO_2 planes in La_2CuO_4. Such a system is correctly described by the Heisenberg antiferromagnetic 2D model ($S = 1/2$) with the interaction between nearest spins in the square lattice. The antiferromagnetic exchange interaction between the first nearest neighbors (the spins of Cu^{2+} ions in a certain CuO_2 plane) is very strong (of the order of 0.13 eV $\approx$ 1500 K for La_2CuO_4 [37]) and is much larger than the interplanar exchange. The interplanar exchange is mainly responsible for the long-range order observed in the insulator phase of CuO_2 planes (the characteristic Néel temperature for La_2CuO_4 is $T_N \approx 300$ K). However, even in the case of a comparatively low level of the system doping with holes, the long-range antiferromagnetic order vanishes in the entire temperature range. It is usually assumed that doping leads to the antiferromagnetic interaction between the second nearest neighbors in the Cu^{2+} subsystem in an individual plane, i.e., to frustration [38]. The cluster calculations indicate that the frustration parameter has a large value ($J_2/J_1 \sim 0.1$) even for undoped La_2CuO_4 [39].

In the subsequent analysis, we will assume that it is frustration that modifies the spectrum of a spin polaron; speaking about the evolution of the spectrum, we will presume a certain qualitative equivalence between doping and frustration in the spin subsystem.

It should be noted that doping is not completely equivalent to frustration. For example, the doped t–J model and the frustrated J_1–J_2 model lead to different results for the dynamic spin–spin structural factor for the Raman scattering spectrum [40]. Nevertheless, it is well known that both doping and frustration lead to a decrease in the magnetic correlation length. Moreover, the numerical calculations on a finite net indicate the equivalence of these models as regards the static spin–spin correlation functions [41]. It will be shown below that the elementary excitation spectrum in our approach is determined just by static spin–spin correlation functions.

Our interpretation of the spin subsystem in the spherically symmetric theory is essential. The motion of a local polaron is often studied using the two-sublattice approximation for the spin substrate, which may lead to the spectrum periodicity relative to the magnetic Brillouin zone [42–44]. However, ARPES experiments indicate the absence of such a periodicity: the spectrum turns out to be periodic relative to the entire Brillouin zone of the CuO_2 plane. In the spherically symmetric approximation, it is the latter symmetry which is pre-

served in the Green's functions of spin excitations as well as in the spectrum of charge carriers.

2. HAMILTONIAN AND COMPUTATIONAL METHOD

The main properties Of the motion of a hole in a CuO_2 plane are described by the following model [33, 34]:

$$\hat{H} = \hat{\tau} + \hat{J} + \hat{h}, \tag{1}$$

$$\hat{\tau} = 4\tau\sum_{\mathbf{R}} p_{\mathbf{R}}^{+}\left(\frac{1}{2} + \tilde{S}_{\mathbf{R}}\right)p_{\mathbf{R}}, \tag{2}$$

$$\hat{J} = \frac{J_1}{2}\sum_{\mathbf{R},\,\mathbf{g}} \mathbf{S}_{\mathbf{R}}\cdot\mathbf{S}_{\mathbf{R}+\mathbf{g}} + \frac{J_2}{2}\sum_{\mathbf{R},\,\mathbf{d}} \mathbf{S}_{\mathbf{R}}\cdot\mathbf{S}_{\mathbf{R}+\mathbf{d}}, \tag{3}$$

where

$$p_{\mathbf{R}} = \frac{1}{2}\sum_{\mathbf{a}} c_{\mathbf{R}+\mathbf{a}}, \quad \tilde{S}_{\mathbf{R}} \equiv S_{\mathbf{R}}^{\alpha}\hat{\sigma}^{\alpha},$$
$$\{p_{\mathbf{R}}, p_{\mathbf{R}'}^{+}\} = \delta_{\mathbf{R},\,\mathbf{R}'} + \frac{1}{4}\sum_{\mathbf{g}} \delta_{\mathbf{R},\,\mathbf{R}'+\mathbf{g}}, \tag{4}$$

J_1, $J_2 > 0$ are the antiferromagnetic interactions between the first nearest ($\mathbf{g} = \pm\mathbf{g}_x, \pm\mathbf{g}_y$) and second nearest neighbors ($\mathbf{d} = \pm\mathbf{g}_x, \pm\mathbf{g}_y$) on the square lattice. Direct oxygen–oxygen jumps are described by the term

$$\hat{h} = -h\sum_{\mathbf{R}} [c_{\mathbf{R}+\mathbf{a}_x}^{+}(c_{\mathbf{R}+\mathbf{a}_y} + c_{\mathbf{R}-\mathbf{a}_y} + c_{\mathbf{R}+\mathbf{g}_x+\mathbf{a}_y} + c_{\mathbf{R}+\mathbf{g}_x-\mathbf{a}_y}) + \text{H.c.}], \tag{5}$$

where $\mathbf{a}_{x,y} = \mathbf{g}_{x,y}/2$. Here and below, recurrent indices indicate summation; {…, …} and […, …] indicate the anticommutator and commutator, respectively; $\mathbf{g}_{x,y}$ are the basis vectors of the square lattice of copper ($|\mathbf{g}| \equiv 1$); $\mathbf{R} + \mathbf{a}$ are four vectors corresponding to O lattice sites near a Cu ion at site $\mathbf{R}$; $c_{\mathbf{R}+\mathbf{a}}^{+}$ is the operator of creation of a hole at an oxygen site (spin indices are omitted to simplify the notation); $\hat{\sigma}^{\alpha}$ are the Pauli matrices; and $\mathbf{S}$ is the operator of a spin localized at a copper site. We do not introduce the relative phases of the p and d orbitals since they can be reconstructed by redefining the creation and annihilation operators through the multiplication by the phase factor $\exp(i\mathbf{Q}\cdot\mathbf{R})$. At the end of computations, we will reconstruct these phases through the substitution $\mathbf{k} \longrightarrow \mathbf{k}' = \mathbf{k} - \mathbf{Q}$. The amplitude τ of oxygen hole jumps couples the motion of a hole with the spin subsystem of copper. Parameters J_1 and J_2 can be expressed in terms of the frustration parameter p:

$$J_1 = (1-p)J, \quad J_2 = pJ, \quad 0 \le p \le 1, \quad J > 0. \tag{6}$$

The frustration parameter p can be treated as an analogue of the number of holes x per copper atom. The estimate based on the one-band Hubbard model for $U/t \sim 5$ gives the value of $p \sim 0.1$ for $x = 0.1$. It should be noted that the spin subsystem of the CuO_2 plane for $La_{2-x}Sr_xCuO_4$ loses the long-range order for $x > 0.02$.

The effective spin-fermion Hamiltonian (1) adopted by us can be obtained as a result of the canonical transformation from the Emery model [45] for $U_d \gg t \gg \varepsilon$, $\varepsilon = \varepsilon_p - \varepsilon_d > 0$, where U_d is the Coulomb one-site repulsion at copper sites. As regards the choice of the set of the energy parameters ε, t, and U_d in the Emery model, they can be obtained either from band calculations or from cluster calculations. We give below two characteristic sets: $\varepsilon = \varepsilon_p - \varepsilon_d = 3.6$ eV, $t = 1.3$ eV, $U_d = 10.5$ eV [46] and $\varepsilon = 2.0$ eV, $t = 1.0$ eV, $U_d = 8$ eV [47, 48]. Considering that $\tau \sim t^2/\varepsilon$ and $J \sim 4\tau(t/\varepsilon)^2$, we choose the values of the parameters $\tau \approx 0.4$ eV, $J = 0.4\tau$, and $h = 0.4\tau$. In the subsequent analysis, we put τ equal to unity. The copper subsystem is considered in the state of a spin liquid, which possesses the spherical symmetry in the spin space. In particular, the spin–spin correlation functions satisfy the relation

$$C_{\mathbf{r}} = \langle S_{\mathbf{R}}^{\alpha}S_{\mathbf{R}+\mathbf{r}}^{\alpha}\rangle = 3\langle S_{\mathbf{R}}^{x(y,z)}S_{\mathbf{R}+\mathbf{r}}^{x(y,z)}\rangle, \quad \langle S_{\mathbf{R}}^{\alpha}\rangle = 0.$$

Analyzing a spin polaron by the projection method, we introduce a finite set of basis operators for each cell $\mathbf{R}$: $A_{\mathbf{R},i}^{+}$ (with spin σ), where i is the number of the operator, $i \le n$. The set includes the creation operator for a hole (with spin σ) as well as other operators describing a hole against the background of spin excitations. The specific form of the operators in the set is dictated by the physical meaning of the problem and will be explained below. In order to calculate the spectral function for a bare hole, we take into account the sum of the contributions from two operators, $c_{\mathbf{R}+\mathbf{a}_x}^{+}$ and $c_{\mathbf{R}+\mathbf{a}_y}^{+}$, describing the two possible positions of the oxygen hole in a unit cell.

As usual, we introduce the retarded two-time Green's functions $G_{ij}(t, \mathbf{k})$ for the Fourier components $A_{\mathbf{k},i}$ of operators $A_{\mathbf{R},i}$:

$$G_{ij}(t,\mathbf{k}) \equiv \langle A_{\mathbf{k},i}(t)|A_{\mathbf{k},j}^{+}(0)\rangle = -i\theta(t)\langle\{A_{\mathbf{k},i}(t), A_{\mathbf{k},j}^{+}(0)\}\rangle, \tag{7}$$

$$A_{\mathbf{k},j} = \frac{1}{\sqrt{N}}\sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}}A_{\mathbf{R},j}, \quad i, j = 1, \ldots, n.$$

The equation of motion for the Fourier components of Green's functions has the form

$$\omega\langle A_{\mathbf{k},i}|A_{\mathbf{k},j}^{+}\rangle_{\omega} = K_{i,j} + \langle B_{\mathbf{k},i}|A_{\mathbf{k},j}^{+}\rangle_{\omega},$$
$$K_{i,j}(\mathbf{k}) = \langle\{A_{\mathbf{k},i}, A_{\mathbf{k},j}^{+}\}\rangle, \quad B_{\mathbf{k},i} = [A_{\mathbf{k},i}, H]. \tag{8}$$

In the projection method, the new operators $B_{\mathbf{k},i}$ are approximated by the projections onto the space $\{A_{\mathbf{k},i}\}$ of the basis operators:

$$B_{\mathbf{k},i} \approx \sum_l L_{i,l}(\mathbf{k}) A_{\mathbf{k},l}, \quad L(\mathbf{k}) = D(\mathbf{k})K^{-1}, \tag{9}$$
$$D_{ij}(\mathbf{k}) = \langle\{B_{\mathbf{k},i}, A^+_{\mathbf{k}j}\}\rangle.$$

After substituting the approximate expressions (9) for operators $B_{\mathbf{k},i}$ into the equation of motion (8), the system of equations for Green's functions $\langle A_{\mathbf{k},i} | A^+_{\mathbf{k},j}\rangle_\omega$ becomes closed and can be presented in matrix form:

$$(\omega E - DK^{-1})G = K, \tag{10}$$

where E is the unit matrix.

The quasiparticle spectrum $\varepsilon(\mathbf{k})$ is determined by the poles of the Green's function G and can be found from the equation

$$\det|K\varepsilon(\mathbf{k}) - D| = 0.$$

3. BASIS OPERATORS AND APPROXIMATIONS

For the first three basis operators, we chose the operators describing the states of a local spin polaron:

$$A_{\mathbf{R},1} = \frac{1}{2}(c_{\mathbf{R}+\mathbf{a}_x} + c_{\mathbf{R}-\mathbf{a}_x}),$$
$$A_{\mathbf{R},2} = \frac{1}{2}(c_{\mathbf{R}+\mathbf{a}_y} + c_{\mathbf{R}-\mathbf{a}_y}), \tag{11}$$
$$A_{\mathbf{R},3} = \tilde{S}_{\mathbf{R}} p_{\mathbf{R}}.$$

Their superposition defines the Zang–Reiss polaron and the state of a bare hole. In particular, we have

$$c_{\mathbf{k},x} = \frac{2}{1+\exp(ik_x)} A_{\mathbf{k},1},$$
$$c_{\mathbf{k},y} = \frac{2}{1+\exp(ik_y)} A_{\mathbf{k},2}, \tag{12}$$
$$c_{\mathbf{k},x(y)} = \frac{1}{\sqrt{N}} \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} c_{\mathbf{R}+\mathbf{a}_{x(y)}}.$$

The elementary excitation spectrum in the local polaron approximation was investigated by us earlier [27]. It was found, in particular, that frustration in the spin subsystem and oxygen–oxygen jumps can explain the emergence of a protracted saddle point in the spectrum. However, a drawback of the local polaron approximation is that the elementary excitation spectrum depends on the spin–spin correlation functions which are valid for the short-range order only (even at zero temperature $T = 0$). Thus, the local polaron approximation does not take into account the effect of the long-range spin order on $\varepsilon(\mathbf{k})$.

Schrieffer [42] pointed out the importance of the coherence factor associated with the existence of a long-range order in an analysis of a 2D periodic Kondo lattice in the framework of the two- sublattice spin subsystem. The simplest Hamiltonian in this model has the form

$$H_K = \sum_{\mathbf{R},\mathbf{g}} t_{\mathbf{g}} c^+_{\mathbf{R}+\mathbf{g}} c_{\mathbf{R}} + I \sum_{\mathbf{R}} c^+_{\mathbf{R}} \tilde{S}_{\mathbf{R}} c_{\mathbf{R}} + \frac{1}{2} J \sum_{\mathbf{R},\mathbf{g}} S^\alpha_{\mathbf{R}+\mathbf{g}} S^\alpha_{\mathbf{R}}, \tag{13}$$

where the intrasite interaction described by constant I is an analogue of Hamiltonian $\hat{\tau}$ (2) in our model.

Schrieffer [42] used the mean-field approximation for the Néel state at $T = 0$ and interpreted spins as classical vectors:

$$S^\alpha_{\mathbf{R}} = \delta_{\alpha z} S_0 e^{i\mathbf{Q}\cdot\mathbf{R}}, \quad S_0 = \text{const}. \tag{14}$$

In this approximation, the Hamiltonian of the Kondo interaction (13) is a potential energy with a doubled period. As a result, such an interaction leads to the hybridization of the states of a bare hole with momenta $\mathbf{k}$ and $\mathbf{k} + \mathbf{Q}$. In order to take such a hybridization into account, we must carry out the standard u–v transformation from the very outset. In the Néel state, the amplitude $S_{\mathbf{Q}}$ of a spin wave with $\mathbf{q} = \mathbf{Q}$ ($\mathbf{Q}$ wave) has a macroscopically large value and possesses the properties similar to those of the amplitude of a Bose particle with zero momentum in a superfluid Bose gas. For this reason, this amplitude can be regarded as a c number. In this case, hybridization corresponds to the pairing of a $\mathbf{Q}$ wave with a local electronic state. However, this does not lead to a new state and only indicates the mixing of states with momenta $\mathbf{k}$ and $\mathbf{k} + \mathbf{Q}$.

A distinguishing feature of this research is the analysis of the one-particle motion over the spin substrate which is in a spherically symmetric state. On this substrate, the mean value $\langle\mathbf{S}_{\mathbf{Q}}\rangle = 0$, and the above approximation is inapplicable. At $T = 0$ and for zero frustration, the only quantity which can be regarded as macroscopic is $\langle\mathbf{S}_{\mathbf{Q}} \cdot \mathbf{S}_{\mathbf{Q}}\rangle$. In this case, the pairing of a local state with $\mathbf{S}_{\mathbf{Q}}$ corresponds to a new delocalized spin-polaron state, viz., the local polaron (11) "dressed" in the antiferromagnetic spin wave $\mathbf{S}_{\mathbf{Q}}$. In the model under investigation, at $T = 0$, such states were introduced in [35] and have the form

$$\tilde{Q}_{\mathbf{R}} A_{\mathbf{R},i}, \quad i = 1, 2, 3, \tag{15}$$

$$\tilde{Q}_{\mathbf{R}} = \frac{1}{\sqrt{N}} e^{-i\mathbf{Q}\cdot\mathbf{R}} \tilde{S}_{\mathbf{Q}}, \quad \tilde{S}_{\mathbf{Q}} = S^\alpha_{\mathbf{Q}} \hat{\sigma}^\alpha,$$
$$\mathbf{S}_{\mathbf{Q}} = \frac{1}{\sqrt{N}} \sum_{\mathbf{R}_1} e^{i\mathbf{Q}\cdot\mathbf{R}_1} \mathbf{S}_{\mathbf{R}_1}. \tag{16}$$

It was shown in [35] that it is important to take into account the quantum nature of the spin $\mathbf{Q}$ wave since the transitions between the states of local and delocalized polarons lead to the splitting of lower bands (in the

local polaron approximation) and considerably modify their properties. It was noted in the Introduction that it is just the results obtained with the help of the complex spin polaron approximation (in which the set of basis vectors contains, along with the operators of a local spin polaron, the polarons dressed in the **Q** wave also) that reproduce the SCBA quasiparticle band at $T = 0$ [36].

In the present work, we consider this model for T, $p \neq 0$. In this case, the spin subsystem loses the long-range order, and the average value $N^{-1}\langle \mathbf{S_Q} \cdot \mathbf{S_Q} \rangle$ is equal to zero. For this reason, we must introduce new operators in order to preserve the results in the complex polaron approximation at $T = 0$ and to take into account the finite spin correlation length ξ. A natural generalization of the set of operators (15) for T, $p \neq 0$ is the addition of the local polaron states $A_{\mathbf{r},1}$, $A_{\mathbf{r},2}$, and $A_{\mathbf{r},3}$ dressed in the set of spin waves with momenta **q** close to the antiferromagnetic vector **Q** (the spin–spin structural factor for such waves has a sharp peak even for T, $p \neq 0$). Such operators can be written in the form

$$A_{\mathbf{R},4} = \tilde{Q}_{\mathbf{R}}^{(\Omega)} A_{\mathbf{R},1}, \quad A_{\mathbf{R},5} = \tilde{Q}_{\mathbf{R}}^{(\Omega)} A_{\mathbf{R},2}, \quad A_{\mathbf{R},6} = \tilde{Q}_{\mathbf{R}}^{(\Omega)} A_{\mathbf{R},3}, \tag{17}$$

$$\tilde{Q}_{\mathbf{R}}^{(\Omega)} = N^{-1} \sum_{\boldsymbol{\rho},\, \mathbf{q} \in \boldsymbol{\Omega}} e^{i\mathbf{q}\cdot\boldsymbol{\rho}} \tilde{S}_{\mathbf{R}+\boldsymbol{\rho}}, \quad \boldsymbol{\Omega} = \{\mathbf{q}, |\pm\pi - q_{x,y}| < \kappa_0\}. \tag{18}$$

Here, **Ω** is the square surrounding point **Q** and equivalent points (i.e., four squares $\Omega = \kappa_0 \times \kappa_0$ at the corners of the first Brillouin zone; see Fig. 1a below). The choice of the parameter κ_0 is dictated by the spin correlation length ξ, which is primarily determined by the frustration parameter p.

Let us substantiate the choice of $\kappa_0(p)$ which will be adopted by us below. For this purpose, we present each operator in (17), say, $A_{\mathbf{R},4}$, in the form

$$A_{\mathbf{R},4} = \sum_{\mathbf{R}'} \alpha(\mathbf{R} - \mathbf{R}') \times \exp[i\mathbf{Q}\cdot(\mathbf{R} - \mathbf{R}')]\tilde{S}_{\mathbf{R}'} A_{\mathbf{R},1}, \tag{19}$$

where

$$\alpha(\mathbf{l}) = \int_{-\kappa_0}^{\kappa_0}\int_{-\kappa_0}^{\kappa_0} \frac{d^2k}{(2\pi)^2} e^{-i\mathbf{k}\cdot\mathbf{l}}. \tag{20}$$

The absolute value of quantity $\alpha(\mathbf{R} - \mathbf{R}')$ depends only on the modulus of difference $\mathbf{l} = \mathbf{R} - \mathbf{R}'$, decreases upon an increase in **l**, and describes the degree of pairing of a local polaron in cell **R** with the spin in cell $\mathbf{R} + \mathbf{l}$. The dependence $\tilde{\alpha}(\mathbf{l}) = |\alpha(\mathbf{l})/\alpha(0)|$ has the form

$$\tilde{\alpha}(\mathbf{l}) = \left|\frac{\sin(l_x\kappa_0)\sin(l_y\kappa_0)}{l_x l_y \Omega}\right|. \tag{21}$$

The value of l_0 satisfying the condition

$$l_0\kappa_0 = \pi \tag{22}$$

can be treated qualitatively as the radius of pairing of a local polaron with spin correlations since $\tilde{\alpha}(\mathbf{l}) \sim 1$ for $l < l_0$ and $\tilde{\alpha}(\mathbf{l}) \ll 1$ for $l \gg l_0$. On the other hand, the radius of paring of a spin polaron must be of the order of the correlation length, $l_0 \sim \xi$. This qualitative estimate leads to the choice of the range of $\Omega = \pi^2/\xi^2$, $\kappa_0 = \pi/\xi$. Such a choice of Ω gives a correct description of the system in two limiting cases. Indeed, if ξ tends to infinity (the spin system possesses a long-range order, $T = 0$), the polaron pairing radius also tends to infinity, κ_0 tends to zero, and $\tilde{Q}_{\mathbf{R}}^{(\Omega)}$ (18) is transformed into $\tilde{Q}_{\mathbf{R}}$ (16). As a result, we obtain a continuous description of the model from finite temperatures to zero. Clearly, in the opposite limit of high temperatures, when the spin correlations between neighboring sites are small, the system must be described by a local spin polaron. It can be seen from (22) that if ξ ($\sim l_0$) tends to unity, the value of κ_0 is close to π. In this case, $\alpha(\mathbf{R} - \mathbf{R}') = \delta_{\mathbf{R},\mathbf{R}'}$, and operators $A_{\mathbf{R},i}$, $i = 4, 5, 6$, are local spin polarons close in their properties to $A_{\mathbf{R},i}\}$, $i = 1, 2, 3$ (for example, $A_{\mathbf{R},6}$ ($\kappa_0 = \pi$) is a linear combination of $A_{\mathbf{R},1}$, $A_{\mathbf{R},2}$, and $A_{\mathbf{R},3}$).

The spin correlation length will be determined below from the spectrum of spin waves, which was calculated for the frustrated Heisenberg model in the spherically symmetric approximation [49]. If the spin system is close to the Néel phase, quantity ξ can be determined from the expansion of the spectrum in the vicinity of the antiferromagnetic vector **Q** [54] (it will be denoted by $\xi_{\mathbf{Q}}$). Then the criterion $\kappa_0 = \pi/\xi_{\mathbf{Q}}$ is used. The explicit form of the spectrum and $\xi_{\mathbf{Q}}$ are given in Appendix A.

The choice of parameter κ_0 for small values of $\xi_{\mathbf{Q}}$ must be different. The value of $\xi_{\mathbf{Q}}$ strongly depends on the frustration parameter p. If $p \geq 0.15$, $\xi_{\mathbf{Q}} \leq 2$–3 ($\kappa_0 = \pi/\xi_{\mathbf{Q}} \leq 0.35\pi$), and the actual value of ξ for such p may differ significantly from $\xi_{\mathbf{Q}}$. Indeed, in the presence of frustration, the spin–spin correlation function $C_{\mathbf{R}} = \langle \mathbf{S}_{\mathbf{R}_0} \cdot \mathbf{S}_{\mathbf{R}_0+\mathbf{R}} \rangle$ has the following dependence on $\mathbf{R} = n_x\mathbf{g}_x + n_y\mathbf{g}_y$:

$$C_{\mathbf{R}} = m_1(\mathbf{R})(-1)^{n_x+n_y} + m_2(\mathbf{R})[(-1)^{n_x} + (-1)^{n_y}], \tag{23}$$

where $m_1(\mathbf{R}) \gg m_2(\mathbf{R})$ for $p \ll 1$ (Néel-type phase) and $m_1(\mathbf{R}) \ll m_2(\mathbf{R})$ for p close to unity (the stripe phase with a gapless spectrum at points $\mathbf{Q}_1 = (\pm\pi, 0), (0, \pm\pi)$).

Frustration parameter p, doping level x, spin correlation length $\xi_{\mathbf{Q}}$, and truncation parameter κ_0. The values of $\varepsilon_1(\pi, 0)$, $\varepsilon_2(\pi, 0)$, and ε_F are given in the τ units; $n_{h,\sigma}^{(1,2)}(\mathbf{k})$ is the spectral weight of the first and second bands

p	κ_0	$\xi_{\mathbf{Q}}$	x	$n_{h,\sigma}^{(1)}(\pi, 0)$	$n_{h,\sigma}^{(2)}(\pi, 0)$	$n_{h,\sigma}^{(1)}(\pi/2, \pi/2)$	$\varepsilon_1(\pi, 0)$	$\varepsilon_2(\pi, 0)$	ε_F
0.05	0.08π	12	0.00	0.23	0.02	0.23	−4.06	–	–
0.10	0.17π	6	0.02	0.25	0.00	0.26	−4.09	–	−4.38
0.13	0.30π	3.4	0.06	0.13	0.11	0.27	−4.21	−4.04	−4.34
0.15	0.35π	<3	0.11	0.06	0.18	0.23	−4.38	−4.03	−4.40
0.2	0.35π	<2	0.14	0.08	0.16	0.24	−4.35	−4.08	−4.40
0.25	0.35π	<1	0.19	0.10	0.13	0.25	−4.28	−4.10	−4.36

The choice of $m_1(\mathbf{R})$ an $m_2(\mathbf{R})$ is dictated by the gap in the spin spectrum at points $\mathbf{Q}$ and $\mathbf{Q}_1$. For intermediate values of frustration $0.15 \le p \le 0.55$, these gaps in the spin spectrum are comparable, and it is meaningless to determine ξ from the expansion of the spectrum in the vicinity of points $\mathbf{Q}$ or $\mathbf{Q}_1$ in this interval of p. On the other hand, the structure of the local polaron is taken into account by operators $A_{\mathbf{R},1}$, $A_{\mathbf{R},2}$, and $A_{\mathbf{R},3}$, while operators $A_{\mathbf{R},4}$, $A_{\mathbf{R},5}$, and $A_{\mathbf{R},6}$ are required for describing polaron states with a large or intermediate radius $l_0 = 2$ to 3. This means that we must fix $\kappa_0 \approx 0.35\pi$ for $p \ge 0.15$ ($\kappa_0 = \pi/l_0$, $l_0 = 2$ to 3) and choose $\kappa_0 = \pi/\xi_{\mathbf{Q}}$ for $p < 0.15$. The value of $\xi_{\mathbf{Q}}$ calculated in this way, as well as the relations between parameter κ_0 and the value of frustration p, are given in the table.

In order to determine the excitation spectrum from operators (11) and (17), we calculate the matrix elements of matrices D and K. Such calculations are usually cumbersome since they involve determining complex commutators of operators $A_{\mathbf{k},i}$ and $B_{\mathbf{k},i}$ with the Hamiltonian. As a rule, such commutators cannot be expressed using only binodal Green's functions and, hence, certain approximations are required. The expressions for the matrix elements are simplified considerably in the one-hole approximation. We will use this approximation since the problem is considered in the limit of low doping ($x < 0.2$). In this case, the matrix elements can be expressed in terms of the binodal and multinodal correlators of spin operators (to be more precise, in terms of binodal, trinodal, tetranodal and pentanodal correlation functions). Taking the spherical symmetry into consideration, we can reduce trinodal to binodal and pentanodal to tetranodal correlators. Tetranodal correlators can be reduced to the form

$$V_{\mathbf{R}_1\mathbf{R}_2\mathbf{R}_3\mathbf{R}_4} = \langle (\mathbf{S}_{\mathbf{R}_1} \cdot \mathbf{S}_{\mathbf{R}_2})(\mathbf{S}_{\mathbf{R}_3} \cdot \mathbf{S}_{\mathbf{R}_4}) \rangle,$$

where we assume that all nodes are different. Such correlators are calculated using the Takahashi approximation [50]

$$V_{\mathbf{R}_1\mathbf{R}_2\mathbf{R}_3\mathbf{R}_4} = C_{R_{12}} C_{R_{34}} + \frac{1}{3} C_{R_{13}} C_{R_{24}} + \frac{1}{3} C_{R_{14}} C_{R_{23}}.$$

As a result, the matrix elements are expressed in terms of the static spin–spin structural factor $C_{\mathbf{q}}$ (the Fourier component $C_{\mathbf{R}}$). The explicit form of matrices K and D is given in Appendix B. The typical components of these matrices have the form

$$u_g = N^{-1} \sum_{\mathbf{q} \in \mathbf{\Omega}} e^{i\mathbf{q} \cdot \mathbf{g}} C_{\mathbf{q}},$$
$$W_{g1}^{(J)} = N^{-2} \sum_{\mathbf{q}_1, \mathbf{q}_2 \in \mathbf{\Omega}} e^{i\mathbf{q}_2 \cdot \mathbf{g}} C_{\mathbf{q}_1 - \mathbf{q}_2} C_{\mathbf{q}_2}. \quad (24)$$

Expressions similar to u_g and $W_{g1}^{(J)}$ contain one or two sums over $\mathbf{q} \in \mathbf{\Omega}$. Each sum over $\mathbf{q}$ is proportional to the ratio Ω/π^2, which is a small parameter in our approximation: as mentioned above, $\Omega/\pi^2 = \kappa_0^2/\pi^2 \le 0.1$ (see the table). This smallness justifies the disregard of some terms proportional to Ω^2.

4. RESULTS AND DISCUSSION

After solving system (10) for the set of operators $A_{\mathbf{R},i}$ (11) and (17), we can write the resultant Green's functions in the form

$$G_{ij}(\omega, \mathbf{k}) = \sum_{l=1}^{6} \frac{z_{(i,j)}^{l}(\mathbf{k})}{\omega - \varepsilon_l(\mathbf{k})}, \quad i, j = 1, \ldots, 6. \quad (25)$$

Relation (12) shows that residues $z_{(1,1)}^{(l)}(\mathbf{k})$ and $z_{(2,2)}^{(l)}(\mathbf{k})$ define the value of $n_{h,\sigma}^{(l)}(\mathbf{k})$, i.e., the number (spectral weight) of bare oxygen holes with momentum $\mathbf{k}$ and spin σ in the quasiparticle state $|\mathbf{k}, \sigma, l\rangle$ of the quasiparticle band $\varepsilon_l(\mathbf{k})$:

$$n_{h,\sigma}^{(l)}(\mathbf{k}) = \frac{2}{1 + \cos k_x} z_{(1,1)}^{(l)}(\mathbf{k}) + \frac{2}{1 + \cos k_y} z_{(2,2)}^{(l)}(\mathbf{k}). \quad (26)$$

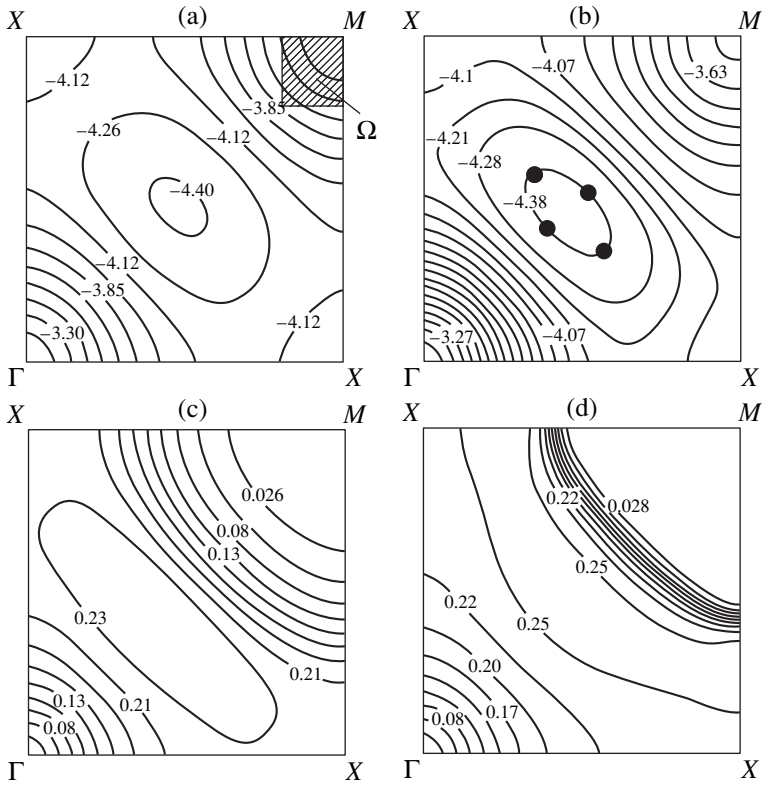


Fig. 1. (a, b) The hole spectrum $\varepsilon_1(\mathbf{k})$ and (c, d) the spectral weight $n^{(1)}_{h,\sigma}(\mathbf{k})$ of a bare hole for the lower band presented by the curves $\varepsilon_1(\mathbf{k})$ = const (in the units of $\tau = 1$) and $n^{(1)}_{h,\sigma}(\mathbf{k})$ = const for the values of frustration parameter $p = 0.05$ (a, c) and $p = 0.10$ (b, d). Dark circles in Fig. b are situated on the constant-energy contours corresponding to the FS calculated for the hole doping level x from the table. The diameters of the circles are proportional to the spectral weight $n^{(1)}_{h,\sigma}(\mathbf{k})$ on the Fermi surface. The results are presented for the first quadrant of the Brillouin zone. The symmetry points are denoted by $\Gamma = (0, 0)$, $X = \{(\pi, 0), (0, \pi)\}$, and $M = (\pi, \pi)$.

It should be recalled that the weight of a bare hole satisfies the sum rule $\sum_l n^{(l)}_{h,\sigma}(\mathbf{k}) = 2$, and the maximum number of holes per unit cell is equal to four in spite of the presence of six bands. This means that the Luttinger theorem is violated in this model.

The spin–polaron spectrum is calculated for $T = 0.2J$. It should be noted that the spectrum displays a weak temperature dependence up to $T \approx 0.4J$. Since we are interested in a relatively weak doping mode ($x \leq 0.2$), the results will be presented for two lowest bands with $l = 1, 2$. In the case of strong doping, our approximation may prove to be insufficient since it is based on the approximation of a single hole and disregards the correlation interaction between polarons.

Figures 1–3 show the $\varepsilon_1(\mathbf{k})$ spectrum and the spectral weight $n^{(1)}_{h,\sigma}(\mathbf{k})$ of a bare hole in the lower band with the help of the level lines $\varepsilon_1(\mathbf{k})$ = const and $n^{(1)}_{h,\sigma}(\mathbf{k})$ = const for the frustration parameter values $p = 0.05$, 0.1, 0.13, 0.15, 0.2, and 0.25.

It was mentioned above that in the analysis of the evolution of the spectrum for various doping levels, we presume a qualitative agreement between the doping level x and the frustration parameter p. This dependence is presented in the table. Naturally, this relation is purely phenomenological and may require certain scaling. Nevertheless, the main conclusions will remain in force if we make the realistic assumption that the low and optimal doping levels correspond to $p \approx 0$ to 0.1 and 0.1 to 0.25, respectively. It should be recalled that frustration takes place even for undoped 2D antiferromagnets if we consider the actual structure of interactions in the CuO_2 plane [39].

In Fig. 1b, dark circles are situated on the constant-energy contours corresponding to the FS calculated for the hole doping x from the table. The diameters of the circles are proportional to the spectral weight $n^{(1)}_{h,\sigma}(\mathbf{k})$ on the FS.

Figures 4–6 show the spectra of two lower hole bands in the electron representation along the symmetry lines for the chosen values of p and for $\tau = 0.4$ eV.

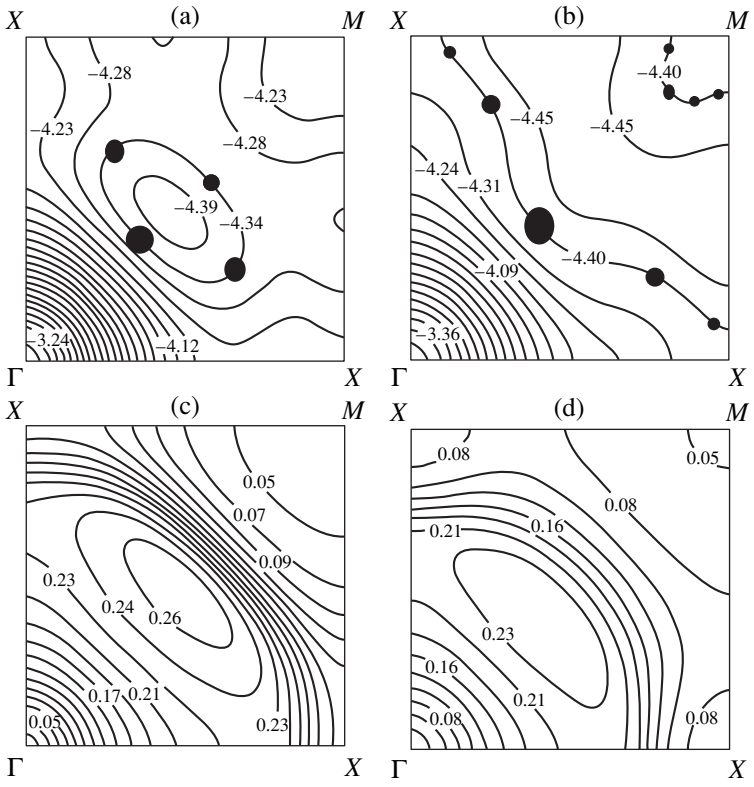


Fig. 2. The same as in Fig. 1 for $p = 0.13$ (a, c) and $p = 0.15$ (b, d).

Zero energy corresponds to the Fermi level ε_F. The regions in the bands with a large spectral weight $n_{h,\sigma}^{(1,2)}(\mathbf{k})$ of a bare particle are shown by solid curves, while the regions with a small value of $n_{h,\sigma}^{(1,2)}(\mathbf{k}) < 0.05$ are presented by dashed curves.

Let us first consider the case of a dielectric or the mode with a very low doping level presented in Figs. 1 and 4. It can be seen from Figs. 1a and 1b that the minimum of $\varepsilon_1(\mathbf{k})$ is close to point N and the spectrum is almost isotropic at the bottom of the band. The spectrum along the Γ–M and Γ–X–M directions reproduces the ARPES results (see, for example, Fig. 4a and Fig. 3 in [2]). The width of the lower hole (upper electron) band $W_1 \approx (3.3–4.4)\tau = 0.44$ eV also corresponds to the ARPES results: $W_1 \approx 0.2$ eV for $Bi_2Sr_2CaCuO_{6+0.5}$ [2], $W_1 \approx 0.3$ eV for $Ca_2CuO_2Cl_2$ [3], and $W_1 \approx 0.35$ eV for La_2CuO_4 [1]. A certain indeterminacy in these experimental data is due to a decrease in the spectral weight in the vicinity of point Γ.

The most important result obtained for a very low doping level is the strong decrease in the spectral weight of the lower band upon the displacement in $\mathbf{k}$ from point N to point M (see Figs. 1c, 1d). It can be seen that the $\mathbf{k}$ line describing such a strong decrease in spectral weight is close to that describing the remanent FS in the ARPES experiments [1, 3]. The spin polaron spectra in Figs. 1a, 1b, and 4 display a symmetry close to the symmetry of the magnetic Brillouin zone, but the spectral weight $n_{h,\sigma}^{(1)}(\mathbf{k})$ has the symmetry of the true Brillouin zone (see Figs. 1c, 1d).

Figures 3b and 6b corresponding to the optimal doping $p = 0.25$, $x = 0.19$ display a large FS with the center at point (π, π). It has the shape satisfying the Luttinger theorem, but with filling $1 + x$ instead of x. The spectral weight distribution presented in Fig. 3d shows that such a large FS appears due to the small spectral weight of a bare hole for the k states of the spin polaron lying under the FS. The mean value of $n_{h,\sigma}^{(1)}(\mathbf{k})$ is close to $0.17 \ll 1$. It should be noted that if only the local polaron approximation is used (e.g., by confining the set of basis operators to operators (11)), only one effective band $\tilde{\varepsilon}_1(\mathbf{k})$ is formed instead of the two lower bands $\varepsilon_{1,2}(\mathbf{k})$. The spectral weight $\tilde{n}_{h,\sigma}^{(1)}(\mathbf{k})$ of bare particles in this band is close to the sum of spectral weights $n_{h,\sigma}^{(1)}(\mathbf{k}) + n_{h,\sigma}^{(2)}(\mathbf{k})$. As a result, $n_{h,\sigma}^{(1)}(\mathbf{k})$ for $\mathbf{k}$ lying under the FS is approximately 1.5 times as large as $n_{h,\sigma}^{(1)}(\mathbf{k})$. Consequently, the area under the FS in the local polaron approximation would be much smaller than in Fig. 3b, which would lead to the loss of qualitative agreement with the experimental pattern.

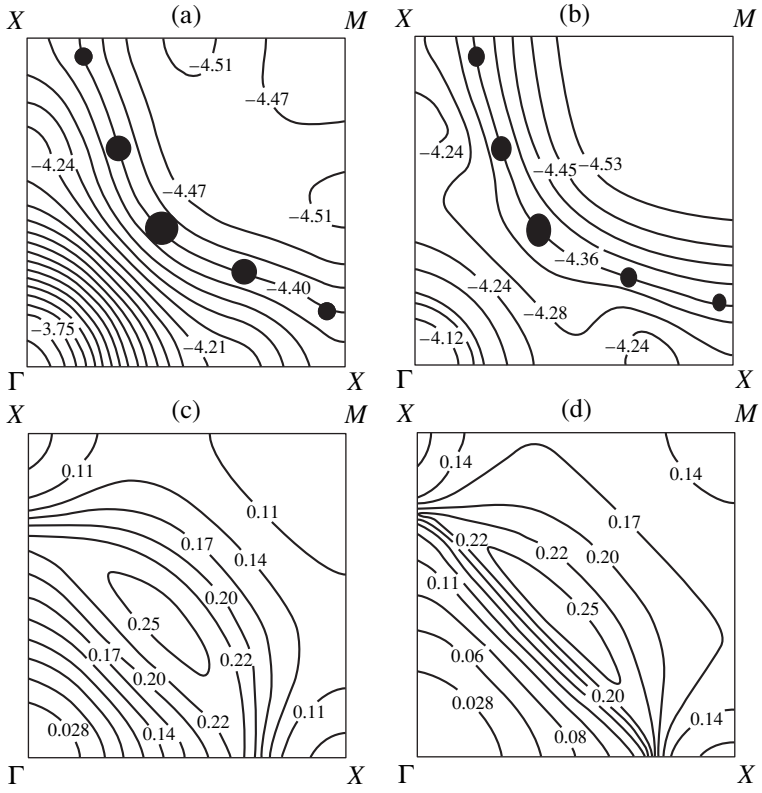


Fig. 3. The same as in Fig. 1 for $p = 0.20$ (a, c) and $p = 0.25$ (b, d).

Let us now compare the two lower bands in the case of an insulator and for the optimal doping along the N–Γ line. In the former case (see Fig. 4a), only the upper electron band is important since the second band has a small spectral weight. In the optimally doped case (Fig. 6b), the second band, whose spectral weight is mainly concentrated in the region adjoining point Γ, is also important. In this case, the matching with the ARPES results is observed (see Fig. 3 in [2]) if we disregard the spectral regions $\varepsilon_1(\mathbf{k})$ and $\varepsilon_2(\mathbf{k})$ (dashed curves in Fig. 6b), where the spectral weight is small, and present the second band in the region close to point Γ as a continuation of the first band. Such an interpretation is correct if we take into account the band broadening. In our approximation, such a broadening must be described by the additional band splitting, which can be obtained by extending the set of operators (17) for delocalized polaron states. In the case of optimal doping ($p = 0.25$), the effective band width is $W = \varepsilon_{2,\,\max} - \varepsilon_F \approx 0.55$ eV; see Fig. 6b ($W \approx 0.38$ eV for a nearly optimal doping of $Bi_2Sr_2CaCuO_{6+\delta}$ [2, 51, 52]). A comparison with the dielectric state shows a certain narrowing of the band upon a decrease in the doping level, as follows from the results of the ARPES experiments [2].

An important feature of the spectrum is the presence of the second band $\varepsilon_2(\mathbf{k})$ with a large spectral weight in the N–M region (see Fig. 6b). This branch resembles the main branch along N–Γ if we displace it by vector $\mathbf{Q}$ in the $\mathbf{k}$ space. The band $\varepsilon_2(\mathbf{k})$ is displaced relative to the principal lower branch along N–M by an energy of the order of 0.2 eV. This branch is also present in the case of intermediate doping (see Fig. 4a). A similar branch (but without an energy shift) was recently observed for Bi2212 [10] for an intermediate doping and was explained as a shadow band, which was discovered for the first time in [9]. The presence of such a band for optimally doped compounds may be important for describing plasma oscillations with a frequency of the order of $\omega_p = 1$ eV if we include the corresponding interband transitions. It should be recalled that such transitions in the conventional band theory correspond to too high energy (about 1.2 eV) and, hence, lead only to a constant contribution to the static permittivity [53].

Our results also reconstruct the presence of a protracted saddle point elongated in the X–Γ direction and close to the FS (see Fig. 3b). Such a saddle point is present in the photoemission data for optimally doped compounds. The problem of saddle point and its relation to the theory of superconductivity was considered by many authors (see, for example, [6]). A certain plane band in the vicinity of the FS was also observed in a wide region of the $\mathbf{k}$ space near point X. It should be noted that for such a plane band close to point $(2\pi/3, 0)$, our approach leads to a large spectral weight of bare charge carriers: $n^{(1)}_{h,\sigma}(\mathbf{k}) \approx 0.22$ (see Fig. 3d).

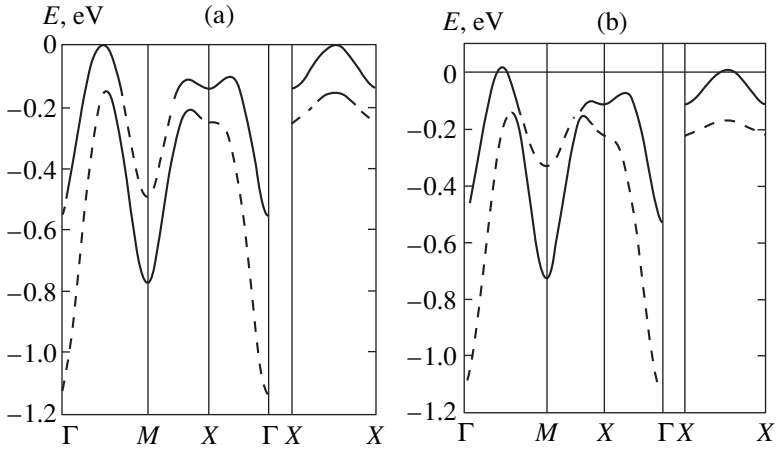


Fig. 4. The spectrum of two lower hole bands in the electron representation along the symmetry lines of the Brillouin zone for the values of frustration parameter $p = 0.05$ (a) and $p = 0.10$ (b). The zero energy corresponds to the Fermi level. The value of τ is assumed to be equal to 0.4 eV. The band regions with a large spectral weight $n^{(1,2)}_{h,\sigma}(\mathbf{k})$ are shown by solid curves, while the regions with a small weight ($n^{(1,2)}_{h,\sigma}(\mathbf{k}) < 0.05$) are presented by dashed curves.

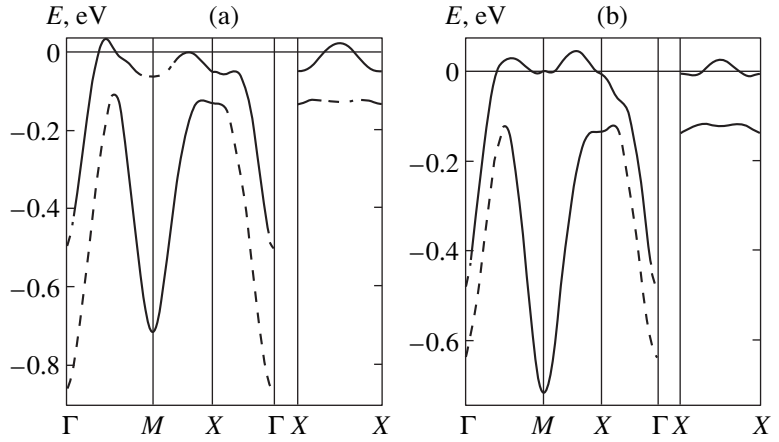


Fig. 5. The same as in Fig. 4 for $p = 0.13$ (a) and $p = 0.15$ (b).

The case of intermediate doping is presented in Figs. 2a, 2c, 5a ($p = 0.13$, $x = 0.06$) and 2b, 2d, 4a ($p = 0.15$, $x = 0.11$). As the doping level increases from $x = 0.02$ to 0.11 (see Figs. 1d, 2c, and 1c consecutively), the behavior of the spectral weight changes abruptly: the remanent FS disappears, the finite filling with $n^{(1)}_{h,\sigma}(\mathbf{k})$ appears near point M, the curvature of the lines $n^{(1)}_{h,\sigma}(\mathbf{k}) = \text{const}$ changes sign in a wide region near the X–X line, and the spectral weight along the line X–$(\pi, \pi/2)$ decreases as a result. In the region between X and $(\pi, \pi/2)$, the spectral weight is displaced into the second band and the value of $n^{(2)}_{h,\sigma}(\mathbf{k})$ increases. Since the FS intersects with the X–M line in the doping interval $x = 0.06$ to 0.11, the spectral weight $n^{(1)}_{h,\sigma}(\mathbf{k})$ (which is proportional to the diameter of circles) varies strongly along the FS.

Such a strong decrease of the spectral weight during the motion in the $\mathbf{k}$ plane along the FS from the central region near $N = (\pi/2, \pi/2)$ to points close to $X = (\pi, 0)$ can be interpreted as the opening of a high-energy pseudogap δ in the vicinity of point $(\pi, 0)$.

Pseudogap δ is often defined as $\delta = \varepsilon(\pi, 0) - \varepsilon_F$ [3]. In our case of intermediate doping ($p = 0.15$, $x = 0.11$), the value of $n^{(1)}_{h,\sigma}(\pi, 0)$ is small, and the lower hole band must be weakly manifested at point X in the ARPES experiments. For this reason, $\varepsilon_2(\pi, 0)$ (see the table) should be regarded as the value that must be taken for $\varepsilon_2(\pi, 0)$ in the expression for gap δ. In this case, the value of the pseudogap for $\tau = 0.4$ eV is equal to 0.19 eV in

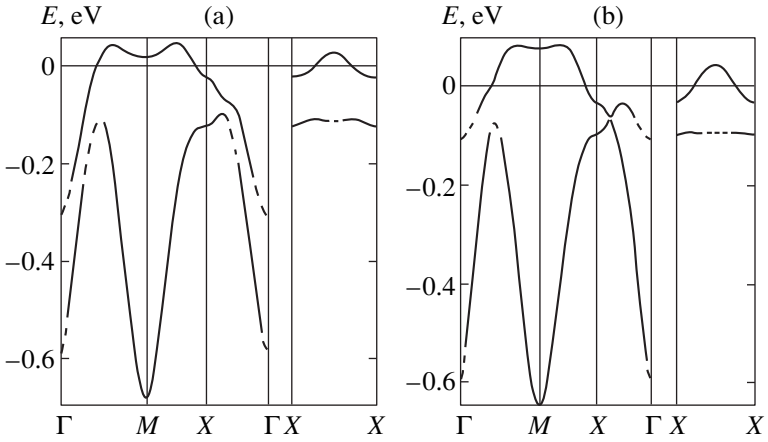


Fig. 6. The same as in Fig. 4 for $p = 0.20$ (a) and $p = 0.25$ (b).

complete accord with the ARPES results (0.1 to 0.2 eV [2, 12–14]). The dependence of the spectral weight $n^{(1)}_{h,\sigma}(\pi, 0)$ of the lower hole band on the doping level, which is presented in the table, is nonmonotonic. First, the value of $n^{(1)}_{h,\sigma}(\pi, 0)$ decreases strongly in the intermediate doping mode, and then increases significantly for the optimal doping. Such a behavior completely corresponds to the ARPES data indicating a decrease in the pseudogap width upon an increase in the doping level from 0.1 to 0.2.

An analysis of the evolution of the spectrum corresponding to the lower band $\varepsilon_1(\mathbf{k})$ under the intermediate doping conditions proves that a protracted saddle point is formed near $(2\pi/3, 0)$ precisely in this narrow doping interval. As the doping level increases above $x = 0.11$, this region of the plane band is preserved.

For any doping level, the spectral weight $n^{(1)}_{h,\sigma}(\mathbf{k})$ of a particle for values of $\mathbf{k}$ close to point N is large and has approximately the same value of $n^{(1)}_{h,\sigma}(\mathbf{k}) = 0.22$ to 0.25 (see the table). It is precisely in this region of the $\mathbf{k}$ space that the ARPES experiments demonstrate the presence of a FS with a clearly manifested quasiparticle peak. Accordingly, the second band in this region of the $\mathbf{k}$ space has low weights $n^{(2)}_{h,\sigma}(\mathbf{k}) < 0.03$ (small incoherent part of the quasiparticle peak) for any doping level.

5. CONCLUSION

We presented the spin-polaron approach to the description of the evolution of the FS and the electronic structure of doped compounds. Our approach is based on the spin-fermion model of the CuO_2 plane and takes doping into account through the introduction of frustration into the spin Hamiltonian. Strong changes in the electronic structure are associated with a simple and natural mechanism: doling leads to frustration in the spin subsystem and to a change in the spin–spin correlation functions, which in turn considerably modifies the spectrum. In other words, doping changes the state of the magnetic substrate, which determines to a considerable extent the shape of the spectrum. It should be recalled that this mechanism does not take into account the direct interaction between polarons, which is undoubtedly significant in the case of doping beyond the optimal level.

Our theory explains the isotropic bottom of the band, the large energy difference for points N and M, and the remanent FS [3] for undoped compounds. In the case of optimal doping, the theory reproduces the region of the plane band (protracted saddle point) between $(\pi/2, 0)$ and $(\pi, 0)$, the large FS [4–8], and the features of a shadow FS [9, 10]. In compounds with an intermediate doping, there exists a region in which considerable changes in the spectrum and in the spectral weight take place on the FS near X. Such a result can be interpreted as the emergence of a pseudogap in the excitation spectrum [12–14].

For a very low doping level, our results reproduce the FS in the form of small hole pockets and the remanent FS (see, for example, Fig. 1b). The spectral weight of the regions on the FS close to point M is quite small. This probably explains why such regions were not detected in the ARPES experiments. In compounds with an intermediate doping, we discovered a mode of the transition to a large FS, but with a very small spectral weight in the vicinity of point $X = (\pi, 0)$. Accordingly, we interpret the arcs of the FS [15] as the parts of the FS corresponding to a large spectral weight. It was found that the regions of the FS with a large spectral weight in the vicinity of N are very stable relative to doping in spite of considerable changes in other regions of the Brillouin zone.

Finally, let us consider the following generalization of the presented approximation, which can in principle provide a more detailed description of the incoherent component of the spectral function $A_h(\mathbf{k}, \omega)$ for a hole.

We introduce several nonoverlapping regions $\mathbf{\Omega}_1$, $\mathbf{\Omega}_2$, $\mathbf{\Omega}_3$, …, $\mathbf{\Omega}_i$, … instead of one region $\mathbf{\Omega}$ (18) (see also [32]) and construct the operators of a delocalized polaron by pairing the states of the local polaron with spin waves from various regions. Such an extension of the set of the basis operators must lead to an additional splitting of lower bands in the projection method and reproduce $A_h(\mathbf{k}, \omega)$ in the form of a set of $\mathbf{\Omega}$ functions [32]:

$$A_h(\mathbf{k}, \omega) \approx \sum_i A_i(\mathbf{k})\delta(\omega - \varepsilon_i(\mathbf{k})).$$

This will probably lead to a more satisfactory description of the shadow FS and to a stronger change in the spectral density on the FS in the case of intermediate doping; i.e., it might give a better description of the pseudogap.

ACKNOWLEDGMENT

This work was supported by INTAS (grant no. 97-11066) and by a grant from the Russian Foundation for Basic Research.

APPENDIX A

In the framework of the spherically symmetric approximation [49], we present the explicit form of Green's functions $G_s(\mathbf{k}, \omega)$ and the spin wave spectrum $\omega(\mathbf{k})$for the triply degenerate mode of spin waves:

$$G_s(\mathbf{k}, \omega) = \langle S^\alpha_{\mathbf{k}} | S^\alpha_{-\mathbf{k}} \rangle = \frac{F(\mathbf{k})}{\omega^2 - \omega^2(\mathbf{k})}, \quad \alpha = x, y, z, \tag{A.1}$$

$$F(\mathbf{k}) = -2[J_1 z_g C_g(1 - \gamma_g(\mathbf{k})) + J_2 z_d C_d(1 - \gamma_d(\mathbf{k}))], \tag{A.2}$$

$$\omega^2(\mathbf{k}) = \frac{2}{3}\Big\{(1 - \gamma_g)\times\Big[J_1 J_2 K_{gd} + I_1^2\Big(z_g(z_g - 1)C_g\alpha_1 + \frac{3}{4}z_g + K_{gg}\Big)\Big]$$

$$+(1 - \gamma_d)\Big[J_1 J_2 K_{gd} + J_2^2\Big(z_d(z_d - 1)C_d\alpha_3 + \frac{3}{4}z_d + K_{dd}\Big)\Big]$$

$$-(1 - \gamma_g^2)J_1^2 z_g^2 G_g\alpha_1 - (1 - \gamma_d^2)J_2^2 z_d^2 C_d\alpha_3 \tag{A.3}$$

$$-(1 - \gamma_g)\gamma_d J_1 J_2 Z_g Z_d C_g\alpha_1 - (1 - \gamma_d)\gamma_g J_1 J_2 Z_g Z_d C_d\alpha_3\Big\},$$

where

$$K_{gd} = \sum_{r = g + d}\alpha_r C_r, \quad K_{gg} = \sum_{\substack{r = g_1 + g_2 \\ g_2 \neq g_2}}\alpha_r C_r,$$

$$K_{dd} = \sum_{\substack{r = d_1 + d_2 \\ d_1 \neq -d_2}}\alpha_r C_r,$$

z_g and z_d are the numbers of the first and second nearest neighbors on the square lattice, α_r are vertex corrections, and

$$\alpha_g = \alpha_1, \quad \alpha_d = \alpha_3, \quad \alpha_r = \alpha_2 \text{ for } r > d,$$

$$(\alpha_1 - 1)/(\alpha_2 - 1) = R_\alpha = 0.863,$$

$$\alpha_3 = (1 - p)\alpha_2 - p\alpha_1,$$

$$J_1 = (1 - p)J, \quad J_2 = pJ,$$

$$K_{gd} = 8\alpha_1 C_g + 8\alpha_2 G_f, \quad K_{gg} = 4\alpha_2 C_{2g} + 8\alpha_3 C_d,$$

$$K_{dd} = 8\alpha_2 C_{2g} + 4\alpha_2 C_{2d}.$$

The values of a_1, C_g, C_d, C_{2g}, C_{gd}, and C_{2d} are calculated self-consistently for each p and T [49].

Equations (A.1)–(A.3) make it possible to express the spin–spin structural factor $C(\mathbf{k})$:

$$C(\mathbf{k}) = A\frac{1 + \exp(\omega_\mathbf{k}/T)}{\exp(\omega_\mathbf{k}/T) - 1}, \quad A = \frac{F(\mathbf{k})}{2\omega_\mathbf{k}}. \tag{A.4}$$

The correlation length $\xi_\mathbf{Q}$ (relative to the Néel phase) is defined with the help of the power expansion of Green's function (A.1) in $\mathbf{q} = \mathbf{Q} - \mathbf{k}$ at point $\omega = 0$ [54]:

$$G(\mathbf{q}, 0) = \frac{G(\mathbf{Q}, 0)}{1 + \xi_\mathbf{Q}^2\mathbf{q}^2}. \tag{A.5}$$

APPENDIX B

We present the explicit form of the matrix elements of matrices K and D. The following notation is adopted:

$$D_{ij}(\mathbf{k}) = \langle\{[A_{\mathbf{k}, i}, (\hat{\tau} + \hat{J} + \hat{h})], A^+_{\mathbf{k}, j}\}\rangle$$

$$= \tau\tilde{\tau}_{ij} + J_1\tilde{J}^{(1)}_{ij} + J_2\tilde{J}^{(2)}_{ij} - h\tilde{h}_{ij},$$

K_{ij}, $\tilde{\tau}_{ij}$, $\tilde{J}^{(1)}_{ij}$, $\tilde{J}^{(2)}_{ij}$, $\tilde{h}_{ij}$ are symmetric matrices,

$$\gamma_g = \frac{1}{2}(\cos k_x + \cos k_y), \quad \gamma_d = \cos k_x \cos k_y,$$

$$\gamma_{2g} = \frac{1}{2}(\cos 2k_x + \cos 2k_y).$$

The nonzero matrix elements are given below.

K Matrix

$$K_{11} = \frac{1}{2}(1+\cos k_x),$$

$$K_{16} = \frac{1}{2}u(1+\cos k_x),$$

$$K_{22} = \frac{1}{2}(1+\cos k_y),$$

$$K_{26} = \frac{1}{2}u(1+\cos k_y),$$

$$K_{33} = \frac{3}{4} + C_g\gamma_g,$$

$$K_{34} = \frac{1}{2}(u + u_g\cos k_x),$$

$$K_{35} = \frac{1}{2}(u + u_g\cos k_y),$$

$$K_{36} = -u + \gamma_g C_g(\nu_g - \nu),$$

$$K_{44} = \frac{1}{2}(u + u_g\cos k_x),$$

$$K_{46} = -u\nu + \frac{1}{2}w + \frac{1}{2}\cos k_x(-\nu_g u - \nu_g + w_g),$$

$$K_{55} = \frac{1}{2}(u + u_g\cos k_y),$$

$$K_{56} = -u\nu + \frac{1}{2}w + \frac{1}{2}\cos k_y(-\nu_g u - \nu u_g + w_g),$$

$$K_{66} = \frac{3}{4}u + 2u\nu - w + \gamma_g\Big[u_g C_g + u^2 - \frac{1}{3}u_g^2 - 4C_g\Big(\frac{1}{3}u_g\nu + u\nu_g\Big) + 2\nu\nu_g C_g + 2\nu_g^2 C_g^2 + C_g\nu^2\Big(1 + \frac{2}{3}C_g\Big)\Big] - \frac{2}{3}\gamma_g W_g^{(\tau)}.$$

$\tilde{\tau}$ *Matrix*

$$\tau_{11} = \frac{1}{2}(1+\cos k_x)^2,$$

$$\tau_{12} = \frac{1}{2} + \gamma_g + \frac{1}{2}\gamma_d,$$

$$\tau_{13} = \Big(\frac{3}{2} + 2C_g\gamma_g\Big)(1+\cos k_x),$$

$$\tau_{14} = (1+\cos k_x)(u + u_g\cos k_x),$$

$$\tau_{15} = u(1+\cos k_x) + u_g(\cos k_y + \gamma_d),$$

$$\tau_{16} = (1+\cos k_x)[u(-1+\gamma_g) + 2C_g\gamma_g(\nu_g - \nu)],$$

$$\tau_{22} = \frac{1}{2}(1+\cos k_y)^2,$$

$$\tau_{23} = (1+\cos k_y)\Big(\frac{3}{2} + 2C_g\gamma_g\Big),$$

$$\tau_{24} = u(1+\cos k_y) + u_g(\cos k_x + \gamma_d),$$

$$\tau_{25} = (1+\cos k_y)(u + u_g\cos k_y),$$

$$\tau_{26} = (1+\cos k_y)[u(-1+\gamma_g) + 2C_g\gamma_g(\nu_g - \nu)],$$

$$\tau_{33} = -\frac{9}{8} - 4C_g\gamma_g + C_g + \frac{1}{2}C_{2g}\gamma_{2g} + C_d\gamma_d,$$

$$\tau_{34} = -\frac{3}{4}u - \frac{1}{2}\nu C_g - u_g\cos k_x + \Big(\cos^2 k_x - \frac{1}{2}\Big)\Big(\nu_{2g}C_g + \frac{1}{2}u_{2g}\Big) + \gamma_g(u_g + 2\nu_g C_g - 2\nu_g C_g\cos k_x - 2\nu C_g) + \gamma_d\Big(\frac{1}{2}u_d + \nu_d C_g\Big),$$

$$\tau_{35} = -\frac{3}{4}u - \frac{1}{2}\nu C_g - u_g\cos k_y + \Big(\cos^2 k_y - \frac{1}{2}\Big)\Big(\nu_{2g}C_g + \frac{1}{2}u_{2g}\Big) + \gamma_g(u_g + 2\nu_g C_g - 2\nu_g C_g\cos k_y - 2\nu C_g) + \gamma_d\Big(\frac{1}{2}u_d + \nu_d C_g\Big),$$

$$\tau_{36} = \frac{9}{2}u - \frac{1}{4}u_g + \gamma_g(3u + 4\nu C_g + 3u_g - 4\nu_g C_g) + \gamma_{2g}\Big(C_g u + \frac{1}{2}\nu_{2g}C_{2g} - \frac{2}{3}\nu_{2g}C_g C_{2g} - \frac{1}{2}\nu C_{2g} - \frac{2}{3}\nu C_g C_{2g} - \frac{1}{3}u_g C_{2g} - 2\nu_g C_g^2 + u_{2g}C_g\Big) + \gamma_d\Big(2C_g u + \nu_d C_d - \nu C_d + \frac{4}{3}C_g C_d(\nu_d + \nu) - \frac{2}{3}u_g C_d - 4\nu_g C_g^2 + 2u_d C_g\Big),$$

$$\tau_{44} = \frac{3}{4}u - \nu_g u_g - 2u\nu + w + \frac{1}{2}w_g^{(1)} + \cos k_x(u_g - 2\nu u_g - 2\nu_g u + 2w_g) + \Big(\cos^2 k_x - \frac{1}{2}\Big)\Big(\frac{1}{2}u_{2g} - 2\nu_g u_g + w_g^{(2)}\Big),$$

$$\tau_{45} = \frac{1}{2}u + 2\gamma_g\left(-vu_g - v_g u + w_g + \frac{1}{2}u_g\right) - 2uv + w + \gamma_d\left(\frac{1}{2}u_d - 2u_g v_g + w_g^{(3)}\right),$$

$$\tau_{46} = \frac{3}{2}uv + \frac{3}{2}u - \frac{3}{2}w + \frac{1}{2}\Big(uC_g + C_g v^2 + C_g v v_g + C_g v_g^2 - \frac{8}{3}u_g v_g C_g + \frac{2}{3}uu_g - \frac{8}{3}uvC_g + \frac{8}{3}vv_g C_g^2\Big) - \frac{1}{3}W_1^{(\tau)} + \gamma_g\Big(-v_g u - vu_g + w_g + 2u_g C_g + 2v^2 C_g + \frac{4}{3}v^2 C_g^2 + 4vv_g C_g + 4v_g^2 C_g^2 + 2u^2 - \frac{2}{3}u_g^2 - \frac{8}{3}u_g vC_g - 8uv_g C_g - \frac{4}{3}W_g^{(\tau)}\Big) + \cos k_x\Big(-w_g + \frac{3}{2}u_g + uv_g + u_g v\Big) + (2\cos^2 k_x - 1) \times\Big(\frac{1}{4}w_{2g} - \frac{1}{4}uv_2 - \frac{1}{4}vu_{2g} + \frac{1}{2}C_g u_{2g} + \frac{1}{2}C_g vv_{2g} + C_g^2 v_g v_{2g} - \frac{1}{3}C_g u_{2g} v - C_g uv_{2g} - \frac{1}{6}u_g u_{2g} + \frac{1}{2}C_g vv_g + \frac{1}{2}C_g v_g^2 + \frac{1}{3}C_g^2 vv_g - \frac{4}{3}C_g u_g v_g + \frac{1}{2}uu_g - \frac{1}{3}W_2^{(\tau)}\Big) + 2\gamma_d\Big(\frac{1}{4}w_d - \frac{1}{4}uv_d - \frac{1}{4}vu_d + \frac{1}{2}C_g u_d + \frac{1}{2}C_g vv_d + C_g^2 v_g v_d - \frac{1}{3}C_g u_d v - C_g uv_d - \frac{1}{6}u_g u_d + \frac{1}{2}C_g vv_g + \frac{1}{2}C_g v_g^2 + \frac{1}{3}C_g^2 vv_g - \frac{4}{3}C_g u_g v_g + \frac{1}{2}uu_g - \frac{1}{3}W_3^{(\tau)}\Big),$$

$$\tau_{55} = \frac{3}{4}u - v_g u_g - 2uv + w + \frac{1}{2}w_g^{(1)} + \cos k_y(u_g - 2vu_{2g} - 2v_g u + 2w_g) + \Big(\cos^2 k_y - \frac{1}{2}\Big)\Big(\frac{1}{2}u_{2g} - 2v_g u_g + w_g^{(2)}\Big),$$

$$\tau_{56} = \frac{3}{2}uv + \frac{3}{2}u - \frac{3}{4}w + \frac{1}{2}\Big(uC_g + C_g v^2 + C_g vv_g + C_g v_g^2 - \frac{8}{3}u_g v_g C_g + \frac{2}{3}uu_g - \frac{8}{3}uvC_g + \frac{8}{3}vv_g C_g^2\Big) - \frac{1}{3}W_1^{(\tau)} + \gamma_g\Big(-v_g u - vu_g + w_g + 2u_g C_g + 2v^2 C_g + \frac{4}{3}v^2 C_g^2 + 4vv_g C_g + 4v_g^2 C_g^2 + 2u^2 - \frac{2}{3}u_g^2 - \frac{8}{3}u_g vC_g - 8uv_g C_g - \frac{4}{3}W_g^{(\tau)}\Big) + \cos k_y\Big(-w_g + \frac{3}{2}u_g + uv_g + u_g v\Big) + (2\cos^2 k_y - 1) \times\Big(\frac{1}{4}w_{2g} - \frac{1}{4}uv_{2g} - \frac{1}{4}vu_{2g} + \frac{1}{2}C_g u_{2g} + \frac{1}{2}C_g vv_g + C_g^2 v_g v_{2g} - \frac{1}{3}C_g u_{2g} v - C_g uv_{2g} - \frac{1}{6}u_g u_{2g} + \frac{1}{2}C_g vv_g + \frac{1}{2}C_g v_g^2 + \frac{1}{3}C_g^2 vv_g - \frac{4}{3}C_g u_g v_g + \frac{1}{2}uu_g - \frac{1}{3}W_2^{(\tau)}\Big) + 2\gamma_d\Big(\frac{1}{4}w_d - \frac{1}{4}uv_d - \frac{1}{4}vu_d + \frac{1}{2}C_g u_d + \frac{1}{2}C_g vv_d + C_g^2 v_g v_d - \frac{1}{3}C_g u_d v - C_g uv_d - \frac{1}{6}u_g u_d + \frac{1}{2}C_g vv_g + \frac{1}{2}C_g v_g^2 + \frac{1}{3}C_g^2 vv_g - \frac{4}{3}C_g u_g v_g + \frac{1}{2}uu_g - \frac{1}{3}W_3^{(\tau)}\Big),$$

$$\tau_{66} = -9vu - \frac{9}{8}u + \frac{9}{2}w - \frac{1}{4}w_g^{(1)} + u\Big(C_g + \frac{2}{3}u_g - \frac{8}{3}C_g v\Big) + v^2 C_g + v_g^2 C_g + vv_g\Big(C_g + \frac{8}{3}C_g^2\Big) + v_g\Big(\frac{1}{2}u_g - \frac{8}{3}C_g u_g\Big) - \frac{2}{3}W_1^{(\tau)} + \gamma_g\Big(uv_g\Big(-6 + \frac{16}{3}C_g\Big) + v_g u(-6 + 16C_g) - 8vv_g C_g - v^2\Big(4C_g + \frac{8}{3}C_g^2\Big) - 4u^2 + \frac{4}{3}u_g^2 - 8v_g^2 C_g^2 - 4u_g C_g + 6w_g + \frac{8}{3}W_g^{(\tau)}\Big) + \gamma_{2g}\Big[\frac{1}{2}u_{2g}C_{2g} + vv_{2g}\Big(C_{2g} + \frac{4}{3}C_g C_{2g}\Big) + v^2\Big(\frac{1}{2}C_{2g} + \frac{4}{3}C_g C_{2g} + \frac{1}{3}C_{2g}^2\Big) + v_{2g}^2\Big(C_{2g}^2 + \frac{8}{3}C_g C_{2g}\Big) + v\Big(-\frac{2}{3}u_{2g}C_{2g} + \frac{2}{3}u_g C_{2g} - 2u_{2g}C_g - 2uC_g\Big) - v_{2g}\Big(2uC_g + \frac{2}{3}u_g C_{2g} + \frac{2}{3}u_{2g}C_g + 2uC_g\Big) + \frac{1}{2}u^2$$

$$-\frac{1}{6}M_{2g}^2-\frac{14}{3}\nu_g^2C_g^2+\nu\nu_g\left(4C_g^2-\frac{4}{3}C_gC_{2g}\right)$$

$$+\frac{8}{3}\nu_g\nu_{2g}(C_g^2-C_gC_{2g})$$

$$+\nu_g\left(\frac{2}{3}u_gC_{2g}+\frac{2}{3}u_{2g}C_g+2uC_g\right)$$

$$+2w_{2g}C_g-\frac{1}{3}w_g^{(2)}C_{2g}-\frac{1}{3}W_{2g}^{(\tau)}\Big]+2\gamma_d\Big[\frac{1}{2}u_dG_d$$

$$+\nu\nu_d\left(C_d+\frac{4}{3}C_gC_d\right)+\nu^2\left(\frac{1}{2}C_d+\frac{4}{3}C_gC_d+\frac{1}{3}C_d^2\right)$$

$$+\nu_d^2\left(C_d^2+\frac{8}{3}C_gC_d\right)$$

$$+\nu\left(-\frac{2}{3}u_dC_d+\frac{2}{3}u_gC_d-2u_dC_g-2uC_g\right)$$

$$-\nu_d\left(2uC_d+\frac{2}{3}u_gC_d+\frac{2}{3}u_dC_g+2uC_g\right)+\frac{1}{2}u^2-\frac{1}{6}u_d^2$$

$$-\frac{14}{3}\nu_g^2C_g^2+\nu\nu_g\left(4C_g^2-\frac{4}{3}C_gC_d\right)+\frac{8}{3}\nu_g\nu_d(C_g^2-C_gC_d)$$

$$+\nu_g\left(\frac{2}{3}u_gC_d+\frac{2}{3}u_dC_g+2uC_g\right)$$

$$+2w_dC_g-\frac{1}{3}w_g^{(3)}C_d-\frac{1}{3}W_d^{(\tau)}\Big].$$

$\tilde{J}^{(1)}$ *Matrix*

$$J_{33}=-4C_g+C_g\gamma_g,$$

$$J_{34}=2C_g(\nu_g-\nu)+\cos k_xC_g\left(\frac{1}{2}\nu-2\nu_g+\frac{1}{2}\nu_{2g}+\nu_d\right),$$

$$J_{35}=2C_g(\nu_g-\nu)+\cos k_yC_g\left(\frac{1}{2}\nu-2\nu_g+\frac{1}{2}\nu_{2g}+\nu_d\right),$$

$$J_{36}=2C_g(\nu-\nu_g)+2u_g+\gamma_g\left(\frac{1}{2}C_g(\nu_g-\nu)\right.$$

$$-2C_g^2\nu_g-\frac{1}{2}u_g-\frac{2}{3}u_gC_g-\frac{4}{3}u_gC_d$$

$$+\frac{4}{3}\nu_dC_gC_d+\frac{2}{3}\nu_{2g}C_gC_{2g}+\frac{2}{3}u_{2g}C_g+\frac{4}{3}u_dC_d\Big),$$

$$J_{44}=2G_g(\nu_g-\nu)+C_g\cos k_x\left(\frac{1}{2}\nu-2\nu_g+\frac{1}{2}\nu_{2g}+\nu_d\right),$$

$$J_{46}=\frac{4}{3}C_g(w_g-w)+\frac{4}{3}C_gu(\nu-\nu_g)+\nu_g^2\left(C_g+\frac{4}{3}C_g^2\right)$$

$$+2\nu^2C_g+\frac{4}{3}W_{g4}^{(J)}+\nu\nu_gC_g\left(-3-\frac{4}{3}C_g\right)$$

$$+\frac{1}{2}\cos k_x\left(\frac{8}{3}u_g^2-\frac{2}{3}u^2-\frac{2}{3}uu_{2g}-\frac{4}{3}uu_d+\frac{8}{3}u\nu_gC_g\right.$$

$$-\frac{8}{3}u_g\nu_gC_g+\frac{2}{3}u\nu_gC_g+\frac{2}{3}u_{2g}\nu_gC_g+\frac{4}{3}u_d\nu_gC_g$$

$$-\frac{2}{3}u_g\nu C_g-\frac{2}{3}u_g\nu_{2g}C_g-\frac{4}{3}u_g\nu_dC_d+4\nu_g^2C_g\left(-\frac{1}{2}-\frac{2}{3}C_g\right)$$

$$+4\nu\nu_gC_g-\nu^2C_g-\nu\nu_{2g}C_g-2\nu\nu_dC_g$$

$$+\nu_g\nu\left(\frac{1}{2}C_g+\frac{2}{3}C_g^2\right)+\nu_g\nu_{2g}C_g\left(\frac{1}{2}+\frac{2}{3}C_g\right)$$

$$+2\nu_g\nu_dC_g\left(\frac{1}{2}+\frac{2}{3}C_g\right)$$

$$+\frac{2}{3}C_g(w_g^{(2)}+w_g^{(1)}+2w_g^{(3)}-4C_gw_g)+\frac{4}{3}W_{g3}^{(J)}\Big),$$

$$J_{55}=2C_g(\nu_g-\nu)+C_g\cos k_y\left(\frac{1}{2}\nu-2\nu_g+\frac{1}{2}\nu_g+\nu_d\right),$$

$$J_{56}=\frac{4}{3}C_g(w_g-w)+\frac{4}{3}C_gu(\nu-\nu_g)+\nu_g^2\left(C_g+\frac{4}{3}C_g^2\right)$$

$$+2\nu^2C_g+\frac{4}{3}W_{g4}^{(J)}+\nu\nu_gC_g\left(-3-\frac{4}{3}C_g\right)$$

$$+\frac{1}{2}\cos k_y\left[\frac{8}{3}u_g^2-\frac{2}{3}u^2-\frac{2}{3}uu_{2g}-\frac{4}{3}uu_d+\frac{8}{3}u\nu_gC_g\right.$$

$$-\frac{8}{3}u_g\nu_gC_g+\frac{2}{3}u\nu_gC_g+\frac{2}{3}u_{2g}\nu_gC_g+\frac{4}{3}u_d\nu_gC_g$$

$$-\frac{2}{3}u_g\nu C_g-\frac{2}{3}u_g\nu_{2g}C_g-\frac{4}{3}u_g\nu_dC_g$$

$$+4\nu_g^2C_g\left(-\frac{1}{2}-\frac{2}{3}C_g\right)+4\nu\nu_gC_g-\nu^2C_g-\nu\nu_{2g}C_g$$

$$-2\nu\nu_dC_g+\nu_g\nu\left(\frac{1}{2}C_g+\frac{2}{3}C_g^2\right)+\nu_g\nu_{2g}C_g\left(\frac{1}{2}+\frac{2}{3}C_g\right)$$

$$+2\nu_g\nu_dC_g\left(\frac{1}{2}+\frac{2}{3}C_g\right)$$

$$+\frac{2}{3}C_g(w_g^{(2)}+w_g^{(1)}+2w_g^{(3)}-4C_gw_g)+\frac{4}{3}W_{g3}^{(J)}\Big],$$

$$J_{66} = F_0 + \gamma_g(F_1 + F_2),$$

$$F_0 = -4uC_g - \frac{8}{3}uu_g + 8C_guv + \frac{8}{3}wC_g - \frac{8}{3}w_gC_g$$

$$+ v_gu_g\left(\frac{32}{2}C_g - 4\right) + \frac{8}{3}v_guC_g + 3v_gC_g - 3vC_g$$

$$- 8v^2C_g - 6v_g^2C_g - \frac{8}{3}v_g^2C_g^2 + 2vv_gC_g - 8C_g^2vv_g$$

$$+ 2w_g^{(1)} + \frac{8}{3}(W_1^{(\tau)} - W_{g1}^{(J)} - W_{g2}^{(J)}),$$

$$F_1 = \frac{2}{3}u^2 + u_gC_g - \frac{10}{3}uv_gC_g - 2u_gvC_g + \frac{1}{2}u_gv$$

$$+ \frac{1}{2}uv_g + v^2C_g\left(\frac{3}{4} + C_g\right) + v_g^2C_g\left(\frac{1}{4} + \frac{5}{3}C_g\right) + 2vv_gC_g$$

$$- \frac{1}{2}w_g - \frac{2}{3}w_g(C_{2g} + 2C_d) + uvC_d - uv_gC_g - \frac{2}{3}u_{2g}v_gC_g$$

$$+ \frac{2}{3}u_gvC_{2g} - \frac{2}{3}u_gvC_g + \frac{2}{3}uvC_{2g} + u_gv_gC_g$$

$$+ \frac{2}{3}uv_gC_{2g} - u_gv_{2g}C_g - \frac{2}{3}uv_{2g}C_{2g} + \frac{2}{3}v^2C_g(C_g - C_{2g})$$

$$+ v_g^2C_g\left(\frac{5}{3}C_g - \frac{7}{3}C_{2g}\right) - \frac{8}{3}vv_gC_g^2$$

$$+ vv_{2g}C_g\left(-\frac{5}{3}C_{2g} + C_g\right) + 4v_gv_{2g}C_gC_{2g}$$

$$+ 2\left(uvC_g - uv_gC_g - \frac{2}{3}u_dv_dC_g + \frac{2}{3}u_gvC_d - \frac{2}{3}u_gvC_g\right.$$

$$+ \frac{2}{3}uvC_d + u_gv_gC_g + \frac{2}{3}uv_gC_d - u_gv_dC_g - \frac{2}{3}uv_dC_d$$

$$+ \frac{2}{3}v^2C_g(C_g - C_d) + v_g^2C_g\left(\frac{5}{3}C_g - \frac{7}{3}C_d\right) - \frac{8}{3}vv_gC_g^2$$

$$+ vv_dC_g\left(-\frac{5}{3}C_d + C_g\right) + 4v_gv_dC_gC_d$$

$$\left. + \frac{2}{3}C_g(w_{g,2g}^{(1)} + 2w_{gd}^{(1)}) - \frac{2}{3}W_g^{(\tau)},\right.$$

$$F_2 = 3u_gv_gC_g + 3uv_gC_g + \frac{2}{3}uv_gC_{2g} + \frac{2}{3}u_{2g}v_gC_g$$

$$- \frac{2}{3}uv_{2g}C_{2g} - u_gv_{2g}C_g - 3uvC_g - \frac{2}{3}u_{2g}vC_g$$

$$+ \frac{4}{3}u_dv_gC_g - \frac{4}{3}uv_dC_d - 2u_gv_dC_g - \frac{4}{3}u_dvC_g$$

$$+ \frac{4}{3}uv_gC_d + \frac{1}{2}uv_g - \frac{1}{2}uv + C_g^2(v_{2g} + v + 2v_d - 4v_g)$$

$$+ v^2\left(-\frac{11}{3}C_g^2 - \frac{1}{4}C_g + \frac{2}{3}C_{2g}C_g + \frac{4}{3}C_gC_d\right)$$

$$+ v_g^2\left(-\frac{50}{3}C_g^2 - \frac{7}{3}\left(C_gC_{2g} - \frac{14}{3}C_gC_d - \frac{1}{4}C_g\right)\right.$$

$$+ vv_g\left(\frac{52}{3}C_g^2 + \frac{1}{2}C_g - \frac{2}{3}C_gC_{2g} - \frac{4}{3}C_gC_d\right)$$

$$+ v_gv_{2g}\left(\frac{8}{3}C_g^2 + \frac{4}{3}C_gC_{2g}\right) + vv_{2g}\left(-\frac{5}{3}C_g^2 + C_gC_{2g}\right)$$

$$+ v_gv_d\left(\frac{8}{3}C_gC_d + \frac{16}{3}C_g^2\right)$$

$$+ vv_d\left(-\frac{10}{3}C_g^2 + 2C_gC_d\right) + \frac{1}{6}W_g^{(J)}.$$

$\tilde{J}^{(2)}$ Matrix

$$J_{33} = -4C_d,$$

$$J_{34} = 2C_d(v_d - v) + \cos k_xC_d(v_g + v_f - 2v_g),$$

$$J_{35} = 2C_d(v_d - v) + \cos k_yC_d(v_g + v_f - 2v_g),$$

$$J_{36} = 2C_d(v - v_d) + 2u_d + \frac{4}{3}\gamma_g(C_d(C_gv_g + C_fv_f)$$

$$- 2C_gC_dv_g + C_g(u_g + u_f) - u_g(C_g + C_f)),$$

$$J_{44} = 2C_d(v_d - v) + \cos k_x(C_d(v_d + v_f) - 2C_dv_g),$$

$$J_{55} = 2C_d(v_d - v) + \cos k_y(C_d(v_d + v_f) - 2C_dv_g),$$

$$J_{64} = \frac{4}{3}uC_d(v - v_d) + \frac{4}{3}C_d(w_d - w) + 2v^2C_d$$

$$+ v_d^2\left(C_d + \frac{4}{3}C_d^2\right) + \frac{4}{3}W_{d4}^{(J)} - vv_d\left(3C_d + \frac{4}{3}C_d^2\right)$$

$$+ \cos k_x\left\{\frac{4}{3}C_d(uv_g - u_gv_d) + \frac{4}{3}u_gu_d - \frac{4}{3}C_dw_{2g}\right.$$

$$+ v_gv_d\left(C_d - \frac{4}{3}C_d^2\right) + 2C_dv_g(v - v_d) - \frac{2}{3}u(u_g + u_f)$$

$$\left. + \frac{2}{3}C_d[v_d(u_g + u_f) - u_d(v_g + v_f)]\right.$$

$$+ \left(\frac{2}{3}C_d^2 + \frac{1}{2}C_d\right)v_d(v_g - v_f) - C_dv(v_g + v_f)$$

$$+\frac{2}{3}C_d(w_{gd}^{(2)}+w_{gd}^{(1)})+\frac{4}{3}W_{d3}^{(J)}\Big\},$$

$$J_{65}=\frac{4}{3}uC_d(v-v_d)+\frac{4}{3}C_d(w_d-w)+2v^2C_d$$

$$+v_d^2\Big(C_d+\frac{4}{3}C_d^2\Big)+\frac{4}{3}W_{d4}^{(J)}-vv_d\Big(3C_d+\frac{4}{3}C_d^2\Big)$$

$$+\cos k_y\Big\{\frac{4}{3}C_d(uv_g-u_gv_d)+\frac{4}{3}u_gu_d-\frac{4}{3}C_dw_g$$

$$+v_gv_d\Big(C_d-\frac{4}{3}C_d^2\Big)+2C_dv_g(v-v_d)-\frac{2}{3}u(u_g+u_f)$$

$$+\frac{2}{3}C_d[v_d(u_g+u_f)-u_d(v_g+v_f)]$$

$$+\Big(\frac{2}{3}C_d^2+\frac{1}{2}C_d\Big)v_d(v_g+v_f)-C_d(v_g+v_f)$$

$$+\frac{2}{3}C_d(w_{gd}^{(2)}+w_{gd}^{(1)})+\frac{4}{3}W_{d3}^{(J)}\Big\},$$

$$J_{66}=F_0+\gamma_g(F_1+F_2),$$

$$F_0=-4uC_d-\frac{8}{3}uu_d+8C_duv+\frac{8}{3}wC_d-\frac{8}{3}w_dC_d$$

$$+v_du_d\Big(\frac{32}{3}C_d-4\Big)+\frac{8}{3}v_duC_d+3v_dC_d-3vC_d$$

$$-8v^2C_d-6v_d^2C_d-\frac{8}{3}v_d^2C_d^2+2vv_dC_d-8C_d^2vv_d$$

$$+2w_d^{(1)}+\frac{8}{3}(W_1^{(\tau)}-W_{d1}^{(J)}-W_{d2}^{(J)}),$$

$$F_1=-\frac{4}{3}w_g(C_g+C_f)+\frac{4}{3}C_g(w_{gd}^{(1)}+w_{fd}^{(1)})$$

$$+uv\Big(4C_d+\frac{4}{3}(C_g+C_f)\Big)+\frac{4}{3}u_gv(C_g+C_f)$$

$$-\frac{8}{3}u_dvC_g-4uv_dC_d+\frac{4}{3}uv_g(C_g+C_f)$$

$$-\frac{4}{3}u(v_gC_g+v_fC_f)-\frac{4}{3}v_dC_g(u_g+u_f)+\frac{4}{3}u_gv_gC_d$$

$$+\frac{8}{3}u_dv_gC_g-\frac{4}{3}u_dC_g(v_g+v_f)-\frac{2}{3}u_gC_d(v_g+v_f)$$

$$+v^2\Big(\frac{8}{3}C_gC_d-\frac{4}{3}C_g(C_g+C_f)\Big)-\frac{32}{3}vv_gC_gC_d$$

$$+2vC_gC_d(v_g+v_f)-\frac{10}{3}vC_d(v_gC_g+v_fC_f)$$

$$+v_gv_d\Big[\frac{20}{3}C_gC_d-2C_d(C_g+C_f)\Big]$$

$$+\frac{8}{3}v_gC_g(v_gC_g+v_fC_f)+\frac{16}{3}v_dC_d(v_gC_g+v_fC_f)$$

$$-\frac{8}{3}v_g^3C_g(C_g+C_f)$$

$$+vv_d\Big[\frac{4}{3}C_g(C_g+C_f)-\frac{4}{3}C_d(C_g+C_f)\Big],$$

$$F_2=v^2\Big[-\frac{8}{3}C_gC_d+\frac{4}{3}C_g(C_g+C_f)\Big]+16C_gC_dvv_g$$

$$-4C_duv+2vv_gC_gC_d+2vv_fC_dC_f$$

$$-\frac{10}{3}vC_dC_g(v_g+v_f)-\frac{4}{3}C_g(u_g+u_f)v+\frac{4}{3}C_du_gv_g$$

$$+\frac{8}{3}C_gu_dv_g+\frac{4}{3}(C_g+C_f)uv_g$$

$$-v_gv_d[20C_gC_d+2C_d(C_g+C_f)]$$

$$-\frac{8}{3}C_g(C_g+C_f)v_g^2+vv_d\Big(\frac{8}{3}C_gC_d-\frac{4}{3}C_g(C_g+C_f)\Big)$$

$$+4C_duv_d+v_dC_d[2C_d(v_g+v_f)-2(v_gC_g+v_fC_f)]$$

$$+\frac{4}{3}C_g(u_g+u_f)v_d-\frac{2}{3}C_du_g(v_g+v_f)$$

$$-\frac{4}{3}u(C_gv_g+C_fv_f)-\frac{4}{3}C_gu_d(v_g+v_f)$$

$$+\frac{10}{3}C_gC_dv_d(v_g+v_f)+\frac{1}{6}W_d^{(J)}$$

$$+2C_dv_d(C_gv_g+C_fv_f)+\frac{8}{3}C_gv_g(C_gv_g+C_fv_f)$$

$$-4C_gC_dv_g+2C_gC_d(v_g+v_f).$$

$\tilde{h}$ Matrix

$$h_{12}=2\Big(\frac{1}{2}+\gamma_g+\frac{1}{2}\gamma_d\Big),$$

$$h_{16}=2u\Big(\frac{1}{2}+\gamma_g+\frac{1}{2}\gamma_d\Big),$$

$$h_{26}=h_{16},$$

$$h_{33}=2\Big(\frac{3}{4}+2C_g\gamma_g+C_d\gamma_d\Big),$$

$$h_{34} = 2\left(\frac{1}{2}u + u_g\gamma_g + \frac{1}{2}u_d\gamma_d\right),$$

$$h_{35} = h_{34},$$

$$h_{36} = 2[-u + 2C_g\gamma_g(\nu_g - \nu) + C_d\gamma_d(\nu_d - \nu)],$$

$$h_{45} = h_{34},$$

$$h_{46} = 2\left[-u\nu + \gamma_g(w_g - u_g\nu - \nu_g u) + \frac{1}{2}w + \frac{1}{2}\gamma_d(w_d - u\nu_d - \nu_d u)\right],$$

$$h_{56} = h_{46},$$

$$h_{66} = 2\left\{\frac{3}{4}u + 2u\nu - w + \gamma_g\left[2C_g u_g + 2C_g(\nu^2 + 2\nu_g\nu) + \frac{4}{3}C_g^2(\nu^2 + 3\nu_g^2) + 2u^2 - \frac{2}{3}u_g^2 - 8C_g\left(\frac{1}{3}u_g\nu + u\nu_g\right)\right]\right.$$

$$-\frac{4}{3}\gamma_g W_g^{(\tau)} + \gamma_d\left[C_d u_d + C_d(\nu^2 + 2\nu\nu_d) + \frac{2}{3}C_d^2(\nu^2 + 3\nu_d^2) - 4C_d\left(\frac{1}{3}u_d\nu + u\nu_d\right)\right.$$

$$\left.\left. + u^2 - \frac{1}{3}u_d^2\right] - \frac{2}{3}\gamma_d W_d^{(\tau)}\right\}.$$

The following notation has been used:

$$\nu = \frac{1}{N}\sum_{\boldsymbol{\kappa}} 1, \quad \frac{1}{N}\sum_{\boldsymbol{\kappa}} \equiv \frac{1}{N}\sum_{\boldsymbol{\kappa}\in\Omega},$$

$$\nu_{\mathbf{l}} = \frac{1}{N}\sum_{\boldsymbol{\kappa}} e^{i\boldsymbol{\kappa}\cdot\mathbf{l}}, \quad \mathbf{l} = \mathbf{g}, \mathbf{d}, 2\mathbf{g}, \mathbf{f},$$

$$u = \frac{1}{N}\sum_{\boldsymbol{\kappa}} C_{\boldsymbol{\kappa}},$$

$$u_{\mathbf{l}} = \frac{1}{N}\sum_{\boldsymbol{\kappa}} e^{i\boldsymbol{\kappa}\cdot\mathbf{l}} C_{\boldsymbol{\kappa}}, \quad \mathbf{l} = \mathbf{g}, \mathbf{d}, 2\mathbf{g}, \mathbf{f},$$

$$w = \frac{1}{N^2}\sum_{\boldsymbol{\kappa}_1,\boldsymbol{\kappa}_2} C_{\boldsymbol{\kappa}_1-\boldsymbol{\kappa}_2},$$

$$w_{\mathbf{l}} = \frac{1}{N^2}\sum_{\boldsymbol{\kappa}_1,\boldsymbol{\kappa}_2} e^{-i\boldsymbol{\kappa}\cdot\mathbf{l}} C_{\boldsymbol{\kappa}_1-\boldsymbol{\kappa}_2}, \quad \mathbf{l} = \mathbf{g}, \mathbf{d}, 2\mathbf{g},$$

$$w_{\mathbf{l}_1\mathbf{l}_2}^{(m)} = \frac{1}{N^2}\sum_{\boldsymbol{\kappa}_1,\boldsymbol{\kappa}_2} e^{-i\boldsymbol{\kappa}_1\cdot\mathbf{l}_1} e^{-i\boldsymbol{\kappa}_2\cdot\mathbf{l}_2} C_{\boldsymbol{\kappa}_1-\boldsymbol{\kappa}_2}, \quad \mathbf{l}_{1,2} = \mathbf{g}, \mathbf{d}, 2\mathbf{g}, \mathbf{f},$$

$$m = 1: \mathbf{l}_1\cdot\mathbf{l}_2 > 0, \quad m = 2: \mathbf{l}_1\cdot\mathbf{l}_2 < 0,$$

$$m = 3: \mathbf{l}_1\cdot\mathbf{l}_2 = 0,$$

$$w_{\mathbf{l}}^{(m)} = w_{\mathbf{l}_1\mathbf{l}_2}^{(m)}(\mathbf{l}_1 = \mathbf{l}_2),$$

$$W_{\mathbf{l}}^{(\tau)} = \frac{1}{N^2}\sum_{\boldsymbol{\kappa}_1,\boldsymbol{\kappa}_2,\boldsymbol{\rho}} \exp[-i(\boldsymbol{\kappa}_1-\boldsymbol{\kappa}_2)\cdot\boldsymbol{\rho}] \times \exp(-i\boldsymbol{\kappa}_2\cdot\mathbf{l}) C_{\boldsymbol{\rho}} C_{\boldsymbol{\rho}-\mathbf{l}}, \quad \mathbf{l} = \mathbf{g}, \mathbf{d}, 2\mathbf{g},$$

$$W_1^{(\tau)} = \frac{1}{N^2}\sum_{\boldsymbol{\kappa}_1,\boldsymbol{\kappa}_2,\boldsymbol{\rho}} \exp[-i(\boldsymbol{\kappa}_1-\boldsymbol{\kappa}_2)\cdot\boldsymbol{\rho}] C_{\boldsymbol{\rho}} C_{\boldsymbol{\rho}-\mathbf{g}},$$

$$W_2^{(\tau)} = \frac{1}{N^2}\sum_{\boldsymbol{\kappa}_1,\boldsymbol{\kappa}_2,\boldsymbol{\rho}} \exp[-i(\boldsymbol{\kappa}_1+\boldsymbol{\kappa}_2)\cdot\mathbf{g}] \times \exp[-i(\boldsymbol{\kappa}_1-\boldsymbol{\kappa}_2)\cdot\boldsymbol{\rho}] C_{\boldsymbol{\rho}} C_{\boldsymbol{\rho}-\mathbf{g}},$$

$$W_3^{(\tau)} = \frac{1}{N^2}\sum_{\boldsymbol{\kappa}_1,\boldsymbol{\kappa}_2,\boldsymbol{\rho}} \exp(-i\boldsymbol{\kappa}_1\cdot\mathbf{g}_x)\exp(-i\boldsymbol{\kappa}_2\cdot\mathbf{g}_y) C_{\boldsymbol{\rho}} C_{\boldsymbol{\rho}-\mathbf{g}},$$

$$W_{l1}^{(J)} = \frac{1}{N^2}\sum_{\boldsymbol{\kappa}_1,\boldsymbol{\kappa}_2} \exp(i\boldsymbol{\kappa}_2\cdot\mathbf{l}) C_{\boldsymbol{\kappa}_1-\boldsymbol{\kappa}_2} C_{\boldsymbol{\kappa}_2}, \quad \mathbf{l} = \mathbf{g}, \mathbf{d},$$

$$W_{l2}^{(J)} = \frac{1}{N^2}\sum_{\boldsymbol{\kappa}_1,\boldsymbol{\kappa}_2} \exp[i(\boldsymbol{\kappa}_1-\boldsymbol{\kappa}_2)\cdot\mathbf{l}] C_{\boldsymbol{\kappa}_1-\boldsymbol{\kappa}_2} C_{\boldsymbol{\kappa}_2}, \quad \mathbf{l} = \mathbf{g}, \mathbf{d},$$

$$W_{l3}^{(J)} = \frac{1}{N^2}\sum_{\boldsymbol{\kappa}_1,\boldsymbol{\kappa}_2} \exp(-i\boldsymbol{\kappa}_2\cdot\mathbf{g}) \times [\gamma_l(\boldsymbol{\kappa}_2) - \gamma_l(\boldsymbol{\kappa}_1-\boldsymbol{\kappa}_2)] C_{\boldsymbol{\kappa}_1-\boldsymbol{\kappa}_2} C_{\boldsymbol{\kappa}_2}, \quad \mathbf{l} = \mathbf{g}, \mathbf{d},$$

$$W_{l4}^{(J)} = \frac{1}{N^2}\sum_{\boldsymbol{\kappa}_1,\boldsymbol{\kappa}_2} [\gamma_l(\boldsymbol{\kappa}_2) - \gamma_l(\boldsymbol{\kappa}_1-\boldsymbol{\kappa}_2)] C_{\boldsymbol{\kappa}_1-\boldsymbol{\kappa}_2} C_{\boldsymbol{\kappa}_2}, \quad \mathbf{l} = \mathbf{g}, \mathbf{d},$$

$$W_{\mathbf{l}}^{(J)} = \frac{1}{N^2}\sum_{\boldsymbol{\kappa}_1,\boldsymbol{\kappa}_2}\sum_{\mathbf{l}}\left[\sum_{\boldsymbol{\rho}} \exp(-i\boldsymbol{\kappa}_1\cdot\boldsymbol{\rho}) \times (C_{\boldsymbol{\rho}+\mathbf{l}} C_{\boldsymbol{\rho}-\mathbf{g}} - C_{\boldsymbol{\rho}} C_{\boldsymbol{\rho}+\mathbf{l}-\mathbf{g}}) - \sum_{\boldsymbol{\rho}} \exp[-i(\boldsymbol{\kappa}_1-\boldsymbol{\kappa}_2)\cdot\boldsymbol{\rho}]\exp(-i\boldsymbol{\kappa}_1\cdot\mathbf{g}) C_{\boldsymbol{\rho}} C_{\boldsymbol{\rho}+\mathbf{l}+\mathbf{g}}\right.$$

$$\left. + \sum_{\boldsymbol{\rho}} \exp[-i((\boldsymbol{\kappa}_1-\boldsymbol{\kappa}_2)\cdot\boldsymbol{\rho})]\exp[-i\boldsymbol{\kappa}_1\cdot(\mathbf{l}-\mathbf{g})] C_{\boldsymbol{\rho}} C_{\boldsymbol{\rho}+\mathbf{l}-\mathbf{g}}\right], \quad \mathbf{l} = \mathbf{g}, \mathbf{d}.$$

REFERENCES

1. B. O. Wells, Z. X. Shen, A. Matsuura, *et al.*, Phys. Rev. Lett. **74**, 964 (1995).
2. D. S. Marshall, D. S. Dessau, A. G. Loeser, *et al.*, Phys. Rev. Lett. **76**, 4841 (1996).
3. F. Ronning, C. Kim, D. L. Feng, *et al.*, Science **282**, 2067 (1998).
4. J. G. Tobin, C. G. Olson, C. Gu, *et al.*, Phys. Rev. B **45**, 5563 (1992).
5. K. Gofron, J. C. Campuzano, H. Ding, *et al.*, J. Phys. Chem. Solids **54**, 1193 (1993).
6. A. A. Abrikosov, J. C. Campuzano, and J. C. Gofron, Physica C (Amsterdam) **214**, 73 (1993).
7. D. S. Dessau, Z. X. Shen, D. M. King, *et al.*, Phys. Rev. Lett. **71**, 2781 (1993).
8. D. M. King, Z. H. Shen, D. S. Dessau, *et al.*, Phys. Rev. Lett. **73**, 3298 (1994).
9. P. Aebi, J. Osterwalder, P. Schwaller, *et al.*, Phys. Rev. Lett. **72**, 2757 (1994).
10. V. Borisenko, M. S. Golden, S. Legner, *et al.*, Phys. Rev. Lett. **84**, 4453 (2000).
11. A. P. Kampf and J. R. Schrieffer, Phys. Rev. B **42**, 7967 (1990).
12. A. G. Loeser, Z. X. Shen, D. S. Dessau, *et al.*, Science **273**, 325 (1996).
13. H. Ding, T. Yokoya, J. C. Campuzano, *et al.*, Nature **382**, 51 (1996).
14. H. Ding, M. R. Norman, T. Yokoya, *et al.*, Phys. Rev. Lett. **78**, 2628 (1997).
15. M. R. Norman, H. Ding, M. Randeria, *et al.*, Nature **392**, 157 (1998).
16. K. J. Szczepanski, P. Horsch, W. Stephan, and M. Ziegler, Phys. Rev. B **41**, 2017 (1990).
17. E. Dagotto, R. Joynt, A. Moero, *et al.*, Phys. Rev. B **41**, 9049 (1990).
18. D. Duffy, A. Nazarenko, S. Haas, *et al.*, Phys. Rev. B **56**, 5597 (1997).
19. C. L. Kane, P. A. Lee, and N. Read, Phys. Rev. B **39**, 6880 (1989).
20. G. Martínez and P. Horsch, Phys. Rev. B **44**, 317 (1991).
21. T. Toyama, Y. Sibata, S. Maekawa, *et al.*, J. Phys. Soc. Jpn. **69**, 9 (2000).
22. J. Bala and A. M. Oleś, Phys. Rev. B **61**, 6907 (2000).
23. R. Eder and K. Becker, Z. Phys. B **78**, 219 (1990).
24. A. V. Chubukov and D. K. Morr, Phys. Rep. **288**, 355 (1997).
25. C. E. Carneiro, M. J. De Oliveira, S. R. A. Salinas, and G. V. Uimin, Physica C (Amsterdam) **166**, 206 (1990).
26. R. Hayn, A. F. Barabanov, and J. Schulenburg, Z. Phys. B **102**, 359 (1997).
27. A. F. Barabanov, V. M. Berezovskiĭ, E. Zhasinas, *et al.*, Zh. Éksp. Teor. Fiz. **110**, 1480 (1996) [JETP **83**, 819 (1996)]; A. F. Barabanov, R. O. Kuzian, L. A. Maksimov, *et al.*, J. Phys.: Condens. Matter **3**, 9129 (1991); Physica C (Amsterdam) **252**, 308 (1995); **212**, 375 (1993).
28. Y. Sibata, T. Tohyama, and S. Maekawa, Phys. Rev. B **59**, 1840 (1999).
29. R. Eder, Y. Ohta, and G. A. Sawatzky, Phys. Rev. B **55**, R3414 (1997).
30. C. Kim, P. J. White, Z. X. Shen, *et al.*, Phys. Rev. Lett. **80**, 4245 (1998).
31. W. Yin, C. Gong, and P. Leung, Phys. Rev. Lett. **81**, 2534 (1998).
32. A. F. Barabanov, A. A. Kovalev, O. V. Urazaev, and A. M. Belemouk, Phys. Lett. A **265**, 221 (2000); A. F. Barabanov, A. A. Kovalev, and O. V. Urazaev, Dokl. Akad. Nauk **365**, 197 (1999) [Dokl. Phys. **44**, 286 (1999)]; Pis'ma Zh. Éksp. Teor. Fiz. **68**, 386 (1998) [JETP Lett. **68**, 412 (1998)].
33. V. J. Emery, Phys. Rev. Lett. **58**, 2794 (1987).
34. V. J. Emery and G. Reiter, Phys. Rev. B **38**, 4547 (1988).
35. A. F. Barabanov, E. Zhasinas, O. V. Urazaev, and L. A. Maksimov, Pis'ma Zh. Éksp. Teor. Fiz. **66**, 173 (1997) [JETP Lett. **66**, 182 (1997)].
36. R. O. Kuzian, R. Hayn, A. F. Barabanov, and L. A. Maksimov, Phys. Rev. B **58**, 6194 (1998).
37. G. Shirane, Y. Endoh, R. J. Birgenau, *et al.*, Phys. Rev. Lett. **59**, 1613 (1987).
38. M. Inui, S. Doniach, and M. Gabay, Phys. Rev. B **38**, 6631 (1988).
39. J. F. Annet, R. M. Martin, A. K. McMahan, and S. Satpathy, Phys. Rev. B **40**, 2620 (1989).
40. S. Bacci, E. Gagliano, and F. Nori, Int. J. Mod. Phys. B **5**, 325 (1991).
41. A. Moreo, E. Dagotto, T. Jolicoeur, and J. Riera, Phys. Rev. B **42**, 6283 (1990).
42. J. R. Schrieffer, J. Low Temp. Phys. **99**, 397 (1995).
43. P. Prelovek, Phys. Lett. A **126**, 287 (1988).
44. A. Ramsak and P. Prelovek, Phys. Rev. B **42**, 10415 (1990).
45. J. Zaanen and A. M. Oleś, Phys. Rev. B **37**, 9423 (1988).
46. M. S. Hybertsen, M. Schlüter, and N. E. Christensen, Phys. Rev. B **39**, 9028 (1989).
47. K. T. Park, K. Terakura, T. Oguchi, *et al.*, J. Phys. Soc. Jpn. **57**, 3445 (1988).
48. A. K. McMahan, R. M. Martin, and S. Satpathy, Phys. Rev. B **38**, 6650 (1988).
49. A. F. Barabanov and V. M. Berezovskiĭ, Zh. Éksp. Teor. Fiz. **106**, 1156 (1994) [JETP **79**, 627 (1994)]; A. F. Barabanov and V. M. Berezovskii, J. Phys. Soc. Jpn. **63**, 3974 (1994); Phys. Lett. A **186**, 175 (1994).
50. Y. Takahashi, Z. Phys. B **71**, 425 (1988).
51. M. S. Golden, V. Borisenko, S. Legner, *et al.*, Adv. Solid State Phys. **40**, 697 (2000).
52. S. Lenger, V. Borisenko, C. Dürr, *et al.*, Phys. Rev. B **62**, 154 (2000).
53. V. G. Grigoryan, G. Paasch, and S. L. Drechsler, Phys. Rev. B **60**, 1340 (1999).
54. H. Shimahara, J. Phys. Soc. Jpn. **62**, 1317 (1993).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 92, No. 4, 2001, pp. 696–700.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 4, 2001, pp. 799–804.

SOLIDS
Electronic Properties

The Effect of High Pressure on the Structure and on the Magnetic and Electronic Properties of Nickel Monoxide

A. G. Gavrilyuk[a], I. A. Troyan[a], I. S. Lyubutin[b, *], and V. A. Sidorov[a]

[a]*Institute of High-Pressure Physics, Russian Academy of Sciences, Troitsk, Moscow oblast, 142190 Russia*
[b]*Shubnikov Institute of Crystallography, Russian Academy of Sciences, Moscow, 117333 Russia*
**e-mail: lyubutin@ns.crys.ras.ru*

Received November 29, 2000

Abstract—A diamond anvil cell is used to investigate the effect of high pressure (up to 37.5 GPa) on the optical absorption spectra of a single crystal of nickel oxide (NiO). In addition, strain-gage measurements are used to experimentally investigate the $V(P)$ equation of state at a hydrostatic pressure of up to 8.5 GPa in a high-pressure chamber of the "toroid" type. Measurements are performed at room temperature. Absorption bands are observed, which correspond to optical d–d transitions of Ni^{2+} ion in the crystal field of ligands ${}^3A_{2g} \longrightarrow {}^3T_{2g}$, ${}^3A_{2g} \longrightarrow {}^1E_{1g}$, ${}^3A_{2g} \longrightarrow {}^3T_{1g}(F)$, and ${}^3A_{2g} \longrightarrow {}^1T_{2g}$. The values of energy of these transitions increase linearly with pressure, and their pressure coefficients are 7.3 ± 0.2, 2.87 ± 0.9, 9.7 ± 0.5, and 8.9 ± 0.3 meV/GPa, respectively. The pressure derivative of the crystal field parameter $10Dq$ corresponding to the ${}^3A_{2g} \longrightarrow {}^3T_{2g}$ transition gives the pressure dependence of the magnitude of exchange integral J in the Anderson hybridization model. It is found that, in the pressure range from zero to 37.5 GPa, the behavior of the exchange integral J is largely defined by the hybridization parameter $b = (10Dq/3)$. At the same time, the Coulomb interaction parameter U_{eff} is independent of pressure and, therefore, has no effect on the variation of J. The Coulomb interaction $U_{\text{eff}} \approx 7.47 \pm 0.005$ eV is determined. The experimental data on the equation of state are used to derive the $J \propto V^{\varepsilon}$ correlation, where $\varepsilon = -2.99 \pm 0.15$, which is in good agreement with the predictions of Bloch's theory ($\varepsilon = -10/3$).

1. INTRODUCTION

Nickel monoxide NiO is an antiferromagnetic insulator with a Néel point of 523 K [1]. Depending on the mode of preparation, it may have either a purely cubic structure or a rombohedrically distorted crystal structure of the NaCl type. The stoichiometric composition has a distorted lattice, which is attributable to the magnetic contribution to the total crystal energy that leads the cubic structure of NaCl to instability. A pure NaCl structure is obtained under conditions of a slight deviation in stoichiometry with respect to oxygen, whose excess somewhat reduces the magnetic contribution to the system Hamiltonian and stabilizes the cubic NaCl structure.

In his pioneer study [2], Mott treated NiO as a typical example of "Mott" insulator with a wide d–d gap U. However, it is clear today that some intermediate situation is realized in the case of NiO, when the Mott–Hubbard d–d gap U is comparable in magnitude with the p–d gap of charge transfer Δ [3–5]. Therefore, it is still hard to say whether nickel monoxide is in fact a Mott insulator or an insulator with charge transfer. Under the effect of high pressure, the insulator–metal transition is expected to occur in such systems. Nevertheless, no such transition has been observed in NiO so far. It is possible that the value of pressure must be very high (apparently, above 1.5 Mbar), because recent investigations of x-ray diffraction [6] revealed the absence of any structural transitions in NiO up to a pressure of 147 GPa.

The foregoing points clearly to the importance of studying the optical properties of NiO under pressure, in particular, the optical absorption spectra. Such experiments produce a wealth of information about the electronic properties of crystals, in particular, about the effect of high pressure on theoretical parameters such as the Mott–Hubbard gap U and the charge transfer gap Δ. This makes for a better understanding of the properties of the material, which is a typical representative of strongly correlated electron systems.

We have investigated the optical absorption spectra of a single crystal of NiO at high pressures of up to 37.5 GPa and used strain-gage measurements to study the pressure dependence of the unit cell volume, $V(P)$, under hydrostatic conditions up to 8.5 GPa.

The optical absorption in NiO at normal pressure was studied in detail by Newman and Chrenko [7]. They have demonstrated that, in the energy range of 0.1–3.5 eV, the absorption spectrum consists of several absorption bands and a background. The peaks of different absorption bands are identified in accordance with certain optical d–d transitions of Ni^{2+} ion in the crystal field of ligands. The background increases monotonically with energy up to the edge of optical

absorption, and its behavior may be described by the expression $\alpha = \text{const}\,(E - E_0)^{-4}$, where E_0 is the absorption edge parameter equal to 4 eV [7]. In interpreting the optical absorption spectra obtained at high pressures, we will rely on the inferences made in [7].

2. EXPERIMENTAL PROCEDURE

Single crystals of NiO were grown by the Verneuil method at the Institute of Crystallography of the Russian Academy of Sciences. The results of x-ray structural analysis demonstrated that the crystal to be investigated had an undistorted NaCl structure with the lattice parameter $a_0 = 4.177 \pm 0.003$ Å.

The optical absorption spectra of NiO were measured at room temperature in a high-pressure diamond anvil cell. A polyethyl siloxane liquid (PES-5) was used as the pressure-transmitting medium. The diameter of the working surface of the anvils was approximately 400 μm. The measurements were performed at different values of pressure up to the maximum value of 37.5 GPa. A rhenium foil gasket was compressed between two diamond anvils from an initial thickness of 200 μm to a working thickness of 50 μm. Then, the electric-spark technique was used to burn a hole about 200 μm in diameter in the compressed part of the gasket. A single-crystal sample of NiO with characteristic dimensions of $5 \times 40 \times 20$ μm^3, chipped from a large crystal, was placed into this hole. In so doing, the (100) crystal plane was normal to the direction of light propagation. The value of pressure was determined using the standard procedure of measuring the luminescence line shift for ruby under the effect of pressure. For this purpose, several pieces of ruby each about 10 μm in size were placed in the high-pressure cell along with the NiO sample.

In order to compare experimental results with theory, one needs to know the dependence of the properties of the material on interion distances and bond angles rather than on pressure. For this purpose, experiments were performed involving measurements of the $V = f(P)$ equation of state for NiO crystal under hydrostatic conditions. The measurements were performed up to 8.5 GPa in a high-pressure chamber of the "toroid" type [8] using an ethanol–methanol alcohol mixture (4:1).

The pressure-related variation of volume was measured using the strain-gage procedure developed previously in [9]. A NiO sample of characteristic dimensions of $4 \times 4 \times 4$ mm^3 was chipped from the same crystal as the sample in the optical absorption experiment. In addition to calibration against NaCl and Al [9], the strain gage was calibrated with respect to equations of state for gold, tungsten, and diamond.

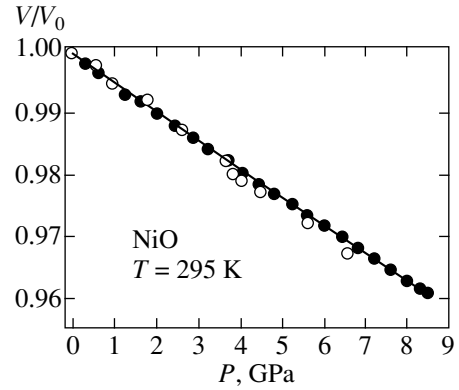


Fig. 1. The relative volume of a unit cell of NiO as a function of hydrostatic pressure. The points indicate the experimental data: solid dots, results of our strain-gage measurements; hollow dots, x-ray diffraction [10]. The line corresponds to Eq. (1).

3. EXPERIMENTAL RESULTS

Figure 1 gives the experimentally obtained dependence of relative variation of the unit cell volume for NiO on hydrostatic pressure. The experimental data were approximated by the Birch–Murnaghan equation of state with free parameters,

$$P = \frac{3}{2}\left[\left(\frac{V}{V_0}\right)^{-7/3} - \left(\frac{V}{V_0}\right)^{-5/3}\right] \times \left\{1 - \frac{3}{4}(4 - K')\left[\left(\frac{V}{V_0}\right)^{-2/3} - 1\right]\right\}. \quad (1)$$

The fitting results were used to find the bulk modulus $K = 197 \pm 1$ GPa and its derivative $K' = 3.4 \pm 0.3$. These values agree well with the data of measurements of x-ray diffraction at high pressure ($K = 199$ GPa [10] and $K = 187 \pm 7$ GPa [11]) and with the experimental data on ultrasound ($K = 193.8$ GPa [12]).

The optical absorption spectra of NiO, measured at different pressures at room temperature, are given in Fig. 2. The high-pressure cell used in the experiments enabled us to observe the optical *d-d* transitions between the electron states of Ni^{2+} ion from the ground level $^3A_{2g}$ to excited levels $^3T_{2g}$, ^1E_g, $^3T_{1g}(F)$, $^1T_{2g}$, and $^1A_{1g}$. The mathematical treatment of the spectra was based on the assumption of the presence in a spectrum of a number of wide absorption bands with a Lorentzian-shaped spectral line and the presence of a background whose shape is taken in the form $\text{const}(E - E_0)^{-4}$ (by analogy with [7]). The contribution made to absorption by excess oxygen was included in the form of a wide line of Lorentzian shape in the energy range 2.0–2.5 eV [7]. It turns out that such a nonlinear fitting procedure provides an excellent description of experimental spectra (see solid lines in Fig. 2). In addition to the position of absorption peaks, the mathematical treatment of spectra produced an estimation for the variation of the absorption edge E_0 with increasing pressure. However, the error of determining E_0 proved

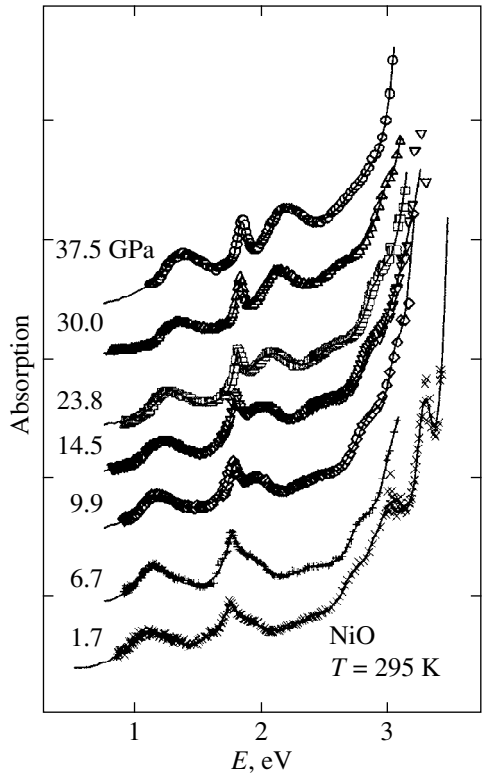


Fig. 2. Optical absorption spectra of NiO, measured at different pressures at room temperature. The symbols indicate the experimental points; the solid lines indicate nonlinear approximation (see text).

to be too high (of the order of 0.5 eV) to make exact inferences about the pressure behavior of the absorption edge. Nevertheless, within the experimental accuracy (about 0.5 eV), one can safely say that the parameter E_0 is almost invariable in the pressure range 0–37.5 GPa and approximately equal to 4 eV.

The table gives the values of energy of *d*–*d* transitions in NiO and the pressure derivatives (dE/dP), measured by us, compared with the results of other researchers [7, 13, 14]. Our data at atmospheric pressure are in good agreement with the data of Newman and Chrenko [7]. We have found that the energies of transitions increase with pressure (see Fig. 3). The ${}^3A_{2g} \longrightarrow {}^3T_{1g}(F)$ transition, which is weakly discernible at atmospheric pressure, shows up clearly at high pressures, while the ${}^3A_{2g} \longrightarrow {}^3T_{1g}(P)$ transition shifts toward the absorption edge and becomes inaccessible for observation. One can see in Fig. 3 that the variations of the transition energies with pressure are well approximated by linear dependences with pressure slopes of 7.3 ± 0.2, 2.87 ± 0.9, 9.7 ± 0.5, and 8.9 ± 0.3 meV/GPa for the ${}^3A_{2g} \longrightarrow {}^3T_{2g}$, ${}^3A_{2g} \longrightarrow {}^1E_{1g}$, ${}^3A_{2g} \longrightarrow {}^3T_{1g}(F)$, and ${}^3A_{2g} \longrightarrow {}^1T_{2g}$ transitions, respectively.

4. DISCUSSION OF THE RESULTS

Our results for the pressure dependence of the crystal field parameter $10Dq$, defined by the energy of ${}^3A_{2g} \longrightarrow {}^3T_{2g}$ transition [7, 15], are in good agreement with the previous experimental results of Stephens and Drickamer [13] (shown by solid dots in Fig. 3). The behavior of the parameter $10Dq$ at high pressure was discussed repeatedly, and a fundamental correlation was demonstrated between the crystal field parameter and the exchange integral *J*. According to the Anderson model [16], NiO is an antiferromagnetic insulator with spin $S = 1$. It is assumed that the 180-degree antiferromagnetic superexchange with ions in the second coordination sphere *J* is much stronger than the 90-degree ferromagnetic exchange with ions of the first coordination sphere; therefore, the latter may be ignored [17]. In the model of Ni-O-hybridization, Anderson derived the expression for the correlation between the exchange integral and the crystal field parameter [16],

$$J = \left(\frac{10Dq}{3}\right)^2 \frac{1}{U_{\text{eff}}}, \qquad (2)$$

The parameters of optical absorption spectra in a single crystal of NiO at normal pressure and room temperature, as well as the pressure coefficients dE/dP for respective electron transitions

Transition	Energy (eV) at normal pressure					Pressure coefficients dE/dP (meV/GPa)		
	Our results (exp.)	Ref. [7]		Ref. [13] (exp.)	Ref. [14] (calc.)	Our results (exp.)	Ref. [13] (exp.)	Ref. [14] (calc.)
		(exp.)	(calc.)					
${}^3A_2 \longrightarrow {}^3T_2({}^3F)$	1.10	1.13	1.07	1.103	1.103	7.27 ± 0.20	9.8	9.29
${}^3A_2 \longrightarrow {}^1E({}^1D)$	1.74	1.75	1.68	–	–	2.87 ± 0.90	–	–
${}^3A_2 \longrightarrow {}^3T_1({}^3F)$	1.86	1.95	1.83	1.735	1.748	9.7 ± 0.5	9.67	10.5
${}^3A_2 \longrightarrow {}^1T_2({}^1D)$	2.70	2.75	2.69	–	–	8.9 ± 0.3	–	–
${}^3A_2 \longrightarrow {}^3T_1({}^3P)$	2.97	2.95	3.04	2.839	2.806	–	13.2	11.9
${}^3A_2 \longrightarrow {}^1A_1({}^1G)$	3.27	3.25	3.21	–	–	–	–	–

where $U_{\text{eff}}^{-1} = U^{-1} + \Delta^{-1}$ is the Coulomb interaction parameter, Δ is the energy of electron transfer from the $2p$-orbital of a ligand to the e_g-orbital of an ion, U is the purely Coulomb d–d interaction of ion electrons, $b = 10Dq/3$ is the parameter of e_g-hopping in the Hubbard model for $3d$-electrons of Ni ion [16], and $10Dq$ is the crystal field parameter in the model of crystal field of ligands [15].

It has been demonstrated recently [18] that, for a NiO crystal of cubic structure in which the antiferromagnetic exchange predominates in magnetic interactions, the exchange integral J is proportional to the frequency Ω_M of two-magnon Raman scattering ($J \propto \Omega_M$), as well as to the Néel point ($J \propto T_N$).

In order to correctly analyze Eq. (2), one must estimate the pressure behavior of the Coulomb interaction parameter U_{eff}. For this purpose, we corrected the data of Massey *et al.* [18] on two-magnon Raman scattering to fit room temperature and determined the value of J at ambient pressure as $J \approx 18$ eV. The pressure dependence of the Néel point for NiO was found previously: $dT_N/dP = 7.33 \pm 0.06$ K/GPa [19]. We can use our data on the pressure dependence of the parameter $10Dq$ to calculate the correlation $U_{\text{eff}} = f(P)$ from Eq. (2),

$$U_{\text{eff}}(P) = \frac{1}{(1/A)T_N(P)}\left[\frac{10Dq(P)}{3}\right]^2. \qquad (3)$$

Figure 4 gives the results of calculating the function $U_{\text{eff}} = f(P)$. One can see in this figure that the parameter U_{eff} in the pressure range 0–37.5 GPa is almost independent of pressure and equal to 7.47 ± 0.05 eV. This result agrees with the inferences made by Massey *et al.* [18], who estimate U_{eff} at 7.5 eV. For the correlation of $T_N = AJ$, our data give $A = 2.44$. This is close to the value of 2.37 in [18]. The results of the foregoing analysis lead one to conclude that the exchange integral J is proportional to $(10Dq)^2$ (in accordance with Anderson's theory [16]) and its pressure dependence is fully defined by the variation of the hybridization parameter b.

We can further use our experimental data on the $V(P)$ and $J(P)$ dependences in NiO to verify the predictions of Bloch's theory [20], who suggested the empirical correlation $J \propto V^{\varepsilon}$ for oxides of transition metals, where $\varepsilon = -10/3$. Figure 5 gives the exchange integral as a function of the unit cell volume, $J(V)$, calculated by our experimental data. The inset at the top right of the figure gives, on a log scale, the relative values of these parameters $J/J_0 = f(V/V_0)$. A linear approximation of this correlation gives the value of the parameter $\varepsilon = (\partial \ln J/\partial \ln V) = -2.99 \pm 0.15$. The value of ε derived by us is slightly less than that derived by Bloch ($\varepsilon = -10/3$); however, within the error, it agrees with the experimental results for two-magnon Raman scattering ($\varepsilon = -3.3 \pm 0.2$) [18].

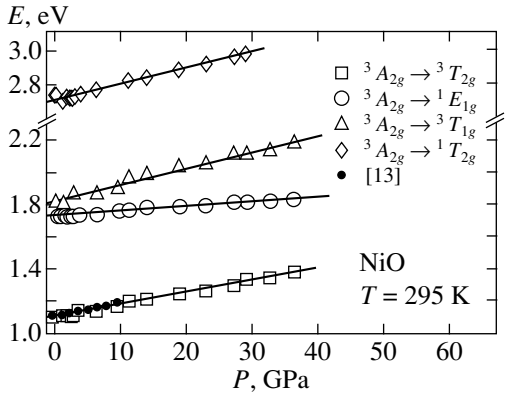


Fig. 3. The pressure dependences of the energies of the ${}^3A_{2g} \longrightarrow {}^3T_{2g}$, ${}^3A_{2g} \longrightarrow {}^3T_{1g}(F)$, ${}^3A_{2g} \longrightarrow {}^1E_{1g}$, and ${}^3A_{2g} \longrightarrow {}^1T_{2g}$ in the absorption spectrum of NiO at room temperature. The symbols indicate the experimental points; the solid lines indicate linear approximation.

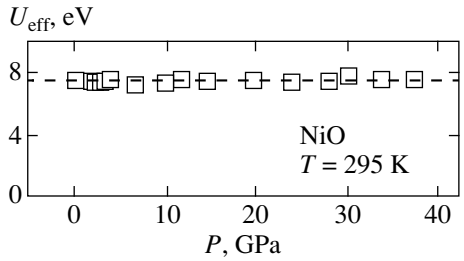


Fig. 4. The pressure dependence of the Coulomb energy parameter U_{eff} in the Anderson model, calculated from the experimentally obtained $T_N(P)$ and $10Dq(P)$ dependences.

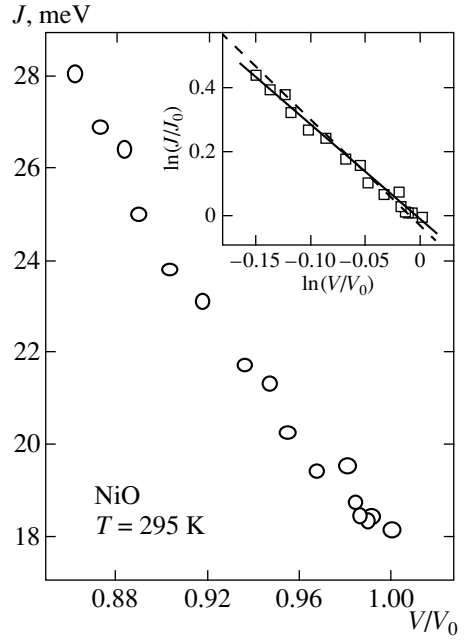


Fig. 5. The exchange integral J as a function of the relative volume of unit cell of NiO at room temperature. The inset gives the relative values of these parameters (□, experimental points; solid line, linear approximation) on a log scale for comparison with Bloch's theory (dashed line).

ACKNOWLEDGMENTS

We are grateful to Prof. P. Yu (University of California at Berkeley) for fruitful discussions and impetus for this study.

The study received support from the Russian Foundation for Basic Research (project no. 01-02-16769).

REFERENCES

1. G. A. Slack, J. Appl. Phys. **31**, 1571 (1960).
2. N. F. Mott, Proc. Phys. Soc. London, Sec. A **62**, 416 (1949); Can. J. Phys. **34**, 287 (1961); *Metal–insulator Transitions* (Taylor and Francis, London, 1990).
3. J. Zaanen, G. A. Sawatsky, and J. W. Allen, Phys. Rev. Lett. **55**, 418 (1985).
4. V. I. Anisimov, J. Zaanen, and O. K. Andersen, Phys. Rev. B **44**, 943 (1991).
5. S. Hufner, J. Osterwalder, T. Riesterer, and F. Hulliger, Solid State Commun. **52**, 793 (1984).
6. T. Eto, S. Endo, M. Imai, *et al.*, Phys. Rev. B **61**, 14984 (2000).
7. R. Newman and R. M. Chrenko, Phys. Rev. **114**, 1507 (1959).
8. L. G. Khvostantsev, L. F. Vereshchagin, and A. P. Novikov, High Temp.–High Press. **9**, 637 (1977).
9. O. B. Tsiok, V. V. Bredikhin, V. A. Sidorov, and L. G. Khvostantsev, High Press. Res. **10**, 523 (1992).
10. R. L. Clendenen and H. G. Drickamer, J. Chem. Phys. **44**, 4223 (1966).
11. E. Huang, High Press. Res. **13**, 307 (1995).
12. J. Wang, E. S. Fisher, and M. H. Manghnani, Chin. Phys. Lett. **8**, 153 (1991).
13. D. R. Stephens and H. G. Drickamer, J. Chem. Phys. **34**, 937 (1961).
14. Zheng Wen-Chen, Li Wei, and Wu Shao-Yi, Phys. Status Solidi A **173**, 437 (1999).
15. D. T. Sviridov, R. K. Sviridova, and Yu. F. Smirnov, *Optical Spectra of Transition Metals in Crystals* (Nauka, Moscow, 1976), p. 239.
16. P. W. Anderson, Phys. Rev. **115**, 2 (1959); Solid State Phys. **14**, 99 (1963).
17. M. T. Hutchings and E. J. Samuelsen, Phys. Rev. B **6**, 344 (1972).
18. M. J. Massey, N. H. Chen, J. W. Allen, and R. Merlin, Phys. Rev. B **42**, 8776 (1990).
19. V. A. Sidorov, Appl. Phys. Lett. **72**, 2174 (1998).
20. D. Bloch, J. Phys. Chem. Solids **27**, 881 (1966).

Translated by H. Bronstein

Journal of Experimental and Theoretical Physics, Vol. 92, No. 4, 2001, pp. 701–709.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 4, 2001, pp. 805–815.

SOLIDS
Electronic Properties

Magnetostriction Anomalies in $DyPO_4$ Caused by Interaction of Levels

Z. A. Kazei*, N. P. Kolmakova, A. A. Sidorenko**, and V. V. Snegirev***

** Moscow State University, Vorob'evy gory, Moscow, 119899 Russia*
*** Bryansk State Technical University, Bryansk, 241035 Russia*
e-mail: kazei@plms.phys.msu.su

Received November 30, 2000

Abstract—Anomalies of the linear magnetostriction caused by the interaction of energy levels of the rare-earth ion in $DyPO_4$ in a strong magnetic field along the [100] and [110] axes are investigated both experimentally and theoretically. Smeared jumps and maxima on the magnetostriction curve $\lambda(H)$ and its derivative $d\lambda(H)/dH$ along the three perpendicular axes in the critical fields H_c = 140 and 200 kOe, when the lower energy levels of the Dy^{3+} ion converge, are discovered. Under the assumption of the adiabatic character of the magnetization process in pulsed fields, the quadrupole moments $\langle O_2^0 \rangle$, $\langle O_2^2 \rangle$, and $\langle P_{xy} \rangle$ are calculated, which give a rather accurate description of the anisotropy of the linear magnetostriction $\lambda(H)$ and its dependence on temperature and magnetic field. It is discovered that a jumplike variation of the quadrupole interaction of α and $\gamma(\delta)$ symmetry in $DyPO_4$ under crossover leads, according to the experiment, to the decrease of the critical field by about 20 kOe and a sharper change of the linear magnetostriction $\lambda(H)$ and its derivative $d\lambda(H)/dH$ in the vicinity of the crossover. *© 2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Rare-earth paramagnets with a tetragonal zircon structure RXO_4 (X = P, V) are very convenient for the study of quadrupole ordering effects [1, 2] caused by quadrupole interactions, for example, Jahn–Teller structural phase transitions (spontaneous or stimulated by a magnetic field). These paramagnets are also convenient for the study of crystalline field effects, such as crossing of levels of rare-earth ions in a magnetic field. An analysis shows that the crossing or, in the more general case, convergence of the energy levels of rare-earth ions in a magnetic field (crossover) occurs for almost all rare-earth vanadates and phosphates with a zircon structure in strong and ultrastrong magnetic fields [3]. The crossover effect was discovered and studied in detail for rare-earth zircons $HoVO_4$ [4, 5] and $PrVO_4$ [6] with Van Vleck ions, and for $YbPO_4$ [7] with a Kramers ion when the magnetic field is oriented along the tetragonal axis. As a rule, such a geometry implies a true (without a gap) crossing of the lower levels, which is accompanied by sharp jumps on the magnetization curves at low temperatures. In the case when the magnetic field is oriented in the basal plane, the situation is qualitatively different. Indeed, the symmetry of the interacting levels is such that the magnetic field perpendicular to the tetragonal axis most often mixes their wave functions (mixes the states that differ by the projections $\Delta J_z = \pm 1$); i.e., it results in a repulsion of the levels and the occurrence of a gap. This enables one to assume that there exist qualitatively different dependences of the magnetic characteristics on the magnetic field and temperature when the levels cross.

The profound change of the electron structure of rare-earth ions (their spectra and wave functions) under crossover (in particular, the change of the ground state of the ion) is accompanied by a jumplike change not only of the magnetic moment, but also of various quadrupole moments. This must lead to anomalies in magnetoelastic characteristics (e.g., magnetostriction) that depend on quadrupole moments. Magnetostriction anomalies under crossover, which have never been studied, are of interest both from the experimental and theoretical points of view. In particular, crossover can lead to a variation of the contribution of quadrupole interactions to magnetoelastic effects, which depend not only on quadrupole constants but also on quadrupole moments determined by the electron structure of the rare-earth ion. This paper focuses on the experimental and theoretical study of magnetostriction anomalies under crossover by the example of $DyPO_4$, which has been under intensive study in relation to its ideal Ising properties at low temperatures [8, 9].

2. SAMPLES AND MEASURING TECHNIQUE

We measured the magnetostriction of the crystal $DyPO_4$ under low temperatures. For the field oriented along the axis of hard magnetization [100], deformations of the crystal along the three mutually perpendicular axes [100], [010], and [001] were measured. These

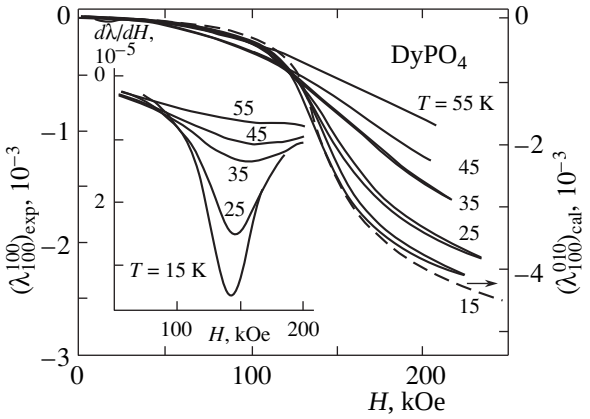


Fig. 1. Experimental (solid) and adiabatic calculated (dashed, T_{st} = 15 K) curves of the linear magnetostriction λ_{100}^{100} and its derivative $d\lambda_{100}^{100}/dH$ (the inset) for the $DyPO_4$ crystal measured along the [100] axis when the magnetic field is oriented along the [100] axis at various temperatures.

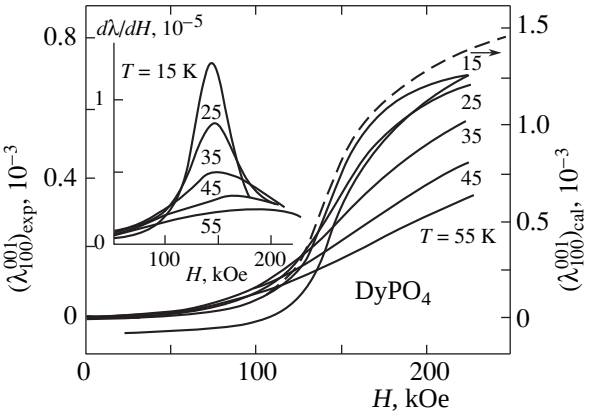


Fig. 2. Experimental (solid) and adiabatic calculated (dashed, T_{st} = 15 K) curves of the linear magnetostriction λ_{100}^{001} and its derivative $d\lambda_{100}^{001}/dH$ (the inset) for the $DyPO_4$ crystal measured along the axis [001] when the magnetic field is oriented along the [100] axis at various temperatures.

deformations make it possible to give a complete description of the volume, tetragonal, and orthorhombic deformations of the crystal in this geometry. For $H \parallel [110]$, the crystal deformations were measured along the $[1\bar{1}0]$, [110], and [001] axes. In order to measure the magnetostriction in pulsed magnetic fields up to 300 kOe (the pulse time was 3 ms) at temperatures ranging from 4.2 to 300 K, the method of a glued quartz piezoelectric sensor was used [10]. A thin monocrystal quartz (*x* cut) plate metallized on both sides was glued to the sample under study. The plate was 0.1–0.15 mm thick and had the dimensions of 1×2 mm^2. Under these conditions, only the linear magnetostriction along the *y* axis can be measured, since the *x*-cut quartz is not piezosensitive in the *z* direction, and the piezosensitivity shear modulus gives no contribution when the field is oriented along the symmetric directions for which no shear deformations occur.

The signal coming from the quartz sensor caused by the sample deformation was amplified by a specially designed electrometric amplifier. To eliminate the detrimental effect of the connecting cable capacitance, a capacitance compensating circuit with a doubly screened cable (triax) was used. In order to make temperature characteristics more homogeneous, the sample with the quartz sensor attached to it was placed in a nonmagnetic ampoule with a dry mix (1 : 1) of kerosene and diffusion oil. The absolute value of the magnetostriction was calibrated by the known value of the magnetostriction of electrolytic nickel. During the field pulse, the signals coming from the piezosensor $\lambda = \Delta l/l(t)$and the field coil $H(t)$ were digitized and registered at 0.02 ms intervals (about 500 points). Then, the smoothed linear magnetostriction function $\lambda(H)$ and its derivative $d\lambda(H)/dH$ were calculated using a computer.

Rare-earth zircon crystals are layered and brittle, which hinders measurements of their magnetostriction in pulsed fields. Preliminary study showed that the free $DyPO_4$ crystal with a sensor attached to it destroys along the cleavage planes after the first field pulse. To protect the sample against destruction, it was placed in a blob of araldite. When the field pulses were not stronger than the critical crossover field (in this case, the deformation of the crystal does not exceed 5×10^{-4}), the signal coming from the piezosensor was stable and reproducible, which testified that the sample did not destroy. For stronger fields, the signal at the start and the end of the pulse was different, and the maximal value of the signal decreased during two or three pulses; then, the situation became stable. In our opinion, this means that the crystal was partially laminated, and the thickness of the sample coupled to the sensor reduced to about 0.3–0.5 mm. The stability of the signal made it possible to investigate the dependence of magnetostriction on field and temperature. However, since elastic constants of quartz are greater than those of the samples under study, measurements performed on a thin crystal result in a noticeable reduction of the magnetostriction deformation. To estimate the absolute value of the magnetostriction in fields stronger than the critical one, we used referencing to the first measurements made in weak fields before the crystal lamination.

3. EXPERIMENTAL RESULTS

Experimental dependences of the magnetostriction deformation λ_{100}^{100}, λ_{100}^{001}, and λ_{100}^{010} along the [100], [001], and [010] axes and their derivatives $d\lambda_{100}^{100}/dH$, $d\lambda_{100}^{001}/dH$, and $d\lambda_{100}^{010}/dH$ for the monocrystal $DyPO_4$ on the strength of magnetic field oriented along the [100] axis are shown in Figs. 1, 2, and 3, respectively. The

behavior of λ_{100}^{100} and λ_{100}^{001} along the [100] and [001] axes is qualitatively the same. The deformation is small in the fields less than 100 kOe. Then, it increases jumpwise near the critical field H_c = 140 kOe and tends to saturation in the fields stronger than the critical one. The critical crossover field value H_c can be determined more accurately by derivatives $d\lambda_{100}^{100}/dH$ and $d\lambda_{100}^{001}/dH$ (see insets in the figures). The value of the critical field for the two geometries is the same; in this field, the magnetostriction reaches a huge value of about 10^{-3} and has different signs along two perpendicular axes. As the temperature increases, the dependences $\lambda_{100}^{100}(H)$, $\lambda_{100}^{001}(H)$ become smeared and the critical field value slightly increases.

The dependences $\lambda_{100}^{010}(H)$ are more complex: at low temperatures (T < 15 K), the magnetostriction is positive in weak fields and changes its sign in the vicinity of the critical field H_c. At higher temperature, the magnetostriction varies monotonically within the range of the field intensities examined. It will be shown below that such a behavior of $\lambda_{100}^{010}(H)$ is determined by the competition of different contributions to the magnetostriction.

A similar behavior of the magnetostriction is observed for the field orientation $H \parallel [110]$. By way of example, Fig. 4 presents the dependences $\lambda_{100}^{001}(H)$ for this geometry. It is seen that the behavior of these curves is basically the same as of those in Fig. 2; the difference is that the critical field value is increased from 140 to 200 kOe.

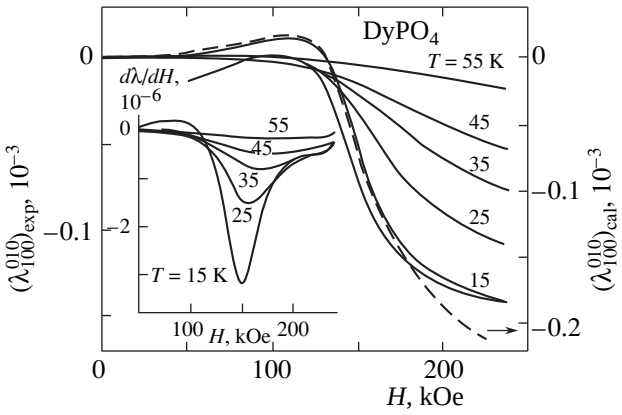


Fig. 3. Experimental (solid) and adiabatic calculated (dashed, T_{st} = 15 K) curves of the linear magnetostriction λ_{100}^{010} and its derivative $d\lambda_{100}^{010}/dH$ (the inset) for the $DyPO_4$ crystal measured along the axis [010] when the magnetic field is oriented along the [100] axis at various temperatures.

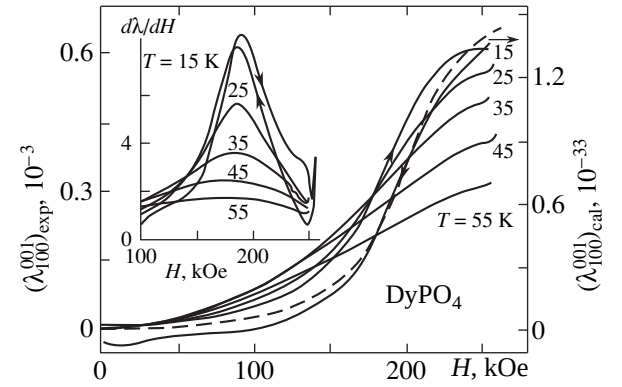


Fig. 4. Experimental (solid) and adiabatic calculated (dashed, T_{st} = 15 K) curves of the linear magnetostriction λ_{110}^{001} and its derivative $d\lambda_{110}^{001}/dH$ (the inset) for the $DyPO_4$ crystal measured along the axis [001] when the magnetic field is oriented along the [110] axis at various temperatures.

4. THEORETICAL TREATMENT

The complete Hamiltonian for a single 4*f* ion includes the Hamiltonian of the crystalline field H_{CF}, Zeeman's term H_Z, and the bilinear term H_B, which describe the interaction of the magnetic moment with the external field **H** and the exchange field $\mathbf{H}_B$, as well as the Hamiltonian of the quadrupole interaction H_{QT}:

$$H = H_{CF} + H_Z + H_B + H_{QT}. \tag{1}$$

Using the equivalent operator method and the approximation of the molecular field for the pair bilinear and quadrupole interactions, we can write these terms in the form (for details, see, e.g., [11])

$$H_{CF} = \alpha_J B_2^0 O_2^0 + \beta_J (B_4^0 O_4^0 + B_4^4 O_4^4) + \gamma_J (B_6^0 O_6^0 + B_6^4 O_6^4), \tag{2}$$

$$H_Z = -g_J \mu_B \mathbf{H} \cdot \mathbf{J}, \tag{3}$$

$$H_B = -g_J \mu_B \mathbf{H}_B \cdot \mathbf{J}, \quad \mathbf{H}_B = n g_J \mu_B \langle \mathbf{J} \rangle, \tag{4}$$

$$H_{QT} = -G^{\alpha} \langle O_2^0 \rangle O_2^0 - G^{\gamma} \langle O_2^2 \rangle O_2^2 - G^{\delta} \langle P_{xy} \rangle P_{xy}, \quad P_{xy} = \frac{1}{2}(J_x J_y + J_x J_y), \tag{5}$$

where n is the bilinear exchange parameter. Here O_m^n, α_J, β_J, and γ_J are the Stevens operator and parameters; B_m^n are the parameters of the crystalline field; g_J and μ_B are the Landé factor and the Bohr magneton, respectively; and

$$\langle O_m^n \rangle = \alpha_J \sum_i \langle i | O_m^n | i \rangle \frac{\exp(-E_i/k_B T)}{Z}$$

$$(O_m^n = O_2^0, O_2^2, P_{xy})$$

are the quadrupole moments. The total quadrupole constants

$$G^{\mu} = G^{\mu}_{ME} + K^{\mu} = (B_{\mu})^2/C^{\mu}_0 + K^{\mu} \quad (\mu = \alpha, \gamma, \delta)$$

include terms both from the one-ion magnetoelastic interaction B^{μ} and the pair quadrupole interaction K^{μ} (C^{μ}_0 is the background elastic constant in the absence of interactions). In the Hamiltonian H_{QT}, we dropped the ε-symmetry terms, which do not contribute to the orientation of the magnetic field in the basal plane and along the tetragonal axis.

Parameters of the pair interactions for $DyPO_4$ were determined from the measurements of the first- and third-order susceptibility and of the parastriction at low temperatures in relatively weak fields for all symmetry modes [12]. For $DyPO_4$, bilinear interactions are not very small, and lead to the antiferromagnetic ordering of Dy^{3+} ions at $T_N \sim 3.4$ K [13]. Estimates show that for the zircon structure the contribution of superexchange and dipole–dipole interactions to the total constant of bilinear interactions are comparable in magnitude. The bilinear interaction constant $n = \theta/C$ is reliably determined for the easy magnetization axis [001]; however, it is not clear whether or not bilinear interactions are completely isotropic. In the absence of magnetic field, the quadrupole interactions G^{μ} in $DyPO_4$ do not lead to sizable quadrupole effects, for example, to the spontaneous quadrupole ordering; however, their role increases in the vicinity of the crossover. In subsequent calculations, we use the values $\theta = -1.5$ K ($H \parallel [001]$), $G^{\alpha} = 1.5$ mK, $G^{\gamma} = 4.4$ mK, and $G^{\delta} = 16.6$ mK obtained in [12].

The eigenvalues and eigenfunctions, which are necessary for the calculation of thermodynamic properties, were determined by the numerical diagonalization of the complete Hamiltonian in which the quadrupole interactions of the α and γ(δ) symmetry (which depend on the electron configuration) were taken into account self-consistently. The symmetrized magnetoelastic deformations ε^{μ} ($\mu = \alpha1, \alpha2, \gamma, \delta$) are linearly related to the quadrupole moments $\langle O^n_m \rangle$ as

$$\begin{aligned}
\varepsilon^{\alpha1,2} &= A^{\alpha1,2}\langle O^0_2\rangle,\\
A^{\alpha1,2} &= \frac{B^{\alpha1,2}C^{\alpha2,1}_0 - B^{\alpha2,1}C^{\alpha1,2}_0}{C^{\alpha1}_0 C^{\alpha2}_0 - (C^{\alpha12}_0)^2},\\
\varepsilon^{\gamma} &= A^{\gamma}\langle O^2_2\rangle = \frac{B^{\gamma}}{C^{\gamma}_0}\langle O^2_2\rangle,\\
\varepsilon^{\delta} &= A^{\delta}\langle P_{xy}\rangle = \frac{B^{\gamma}}{C^{\delta}_0}\langle P_{xy}\rangle.
\end{aligned} \tag{6}$$

The coefficients $A^{\alpha1}$, $A^{\alpha2}$, A^{γ}, and A^{δ}, which depend on the ratio of the magnetoelastic coefficients B^{μ} and the elastic constants C^{μ}_0, were determined by measuring parastriction in relatively weak fields. In subsequent calculations, we use the values $A^{\alpha1} = 17 \times 10^{-6}$, $A^{\alpha2} = -27 \times 10^{-6}$, $A^{\gamma} = -66 \times 10^{-6}$, and $A^{\delta} = 281 \times 10^{-6}$, which were determined experimentally in [12].

The linear magnetostriction $\lambda^{\alpha1\alpha2\alpha3}_{\beta1\beta2\beta3}$ oriented along the direction (α1, α2, α3) induced by the magnetic field oriented along the direction (β1, β2, β3) is related to the symmetrized deformations as (see [14])

$$\begin{aligned}
\lambda^{001}_{100} &= \frac{1}{\sqrt{3}}\varepsilon^{\alpha1} + \frac{2}{\sqrt{6}}\varepsilon^{\alpha2}.\\
\lambda^{100,010}_{100} &= \left(\frac{1}{\sqrt{3}}\varepsilon^{\alpha1} - \frac{1}{\sqrt{6}}\varepsilon^{\alpha2}\right) \pm \frac{1}{\sqrt{2}}\varepsilon^{\gamma},\\
\lambda^{001}_{110} &= \frac{1}{\sqrt{3}}\varepsilon^{\alpha1} + \frac{2}{\sqrt{6}}\varepsilon^{\alpha2},\\
\lambda^{110,\bar{1}10}_{110} &= \left(\frac{1}{\sqrt{3}}\varepsilon^{\alpha1} - \frac{1}{\sqrt{6}}\varepsilon^{\alpha2}\right) \pm \frac{1}{\sqrt{2}}\varepsilon^{\delta}.
\end{aligned} \tag{7}$$

The orientation (β1, β2, β3) of the magnetic field appears implicitly in the formulas above by determining the symmetrized deformations $\varepsilon^{\alpha1}$, $\varepsilon^{\alpha2}$, ε^{γ} ($H \parallel [100]$), or ε^{δ} ($H \parallel [100]$).

5. DISCUSSION OF THE RESULTS

5.1. Crystalline Field and the Zeeman Effect

The crystalline field for $DyPO_4$ is believed to be reliably determined on the basis of numerous experimental data, including inelastic neutron scattering [15]. We used the parameters of the crystalline field from the paper [12] determined on the ground multiplet: $B^0_2 = 202$ K, $B^0_4 = 22$ K, $B^4_4 = 1024$ K, $B^0_6 = -57$ K, and $B^4_6 = 15$ K. The Dy^{3+} ion spectra calculated with these parameters in the magnetic field oriented along the [100] and [110] axes (the Zeeman effect) are presented in Fig. 5. The general multiplet splitting at $H = 0$ is ~ 430 K. The figure shows only the six lower levels, which make the major contribution to the thermodynamic properties at low temperatures. In the crystalline phosphate field, the ground multiplet $^6H_{15/2}$ of the Dy^{3+} ion is split in such a way that the ground state is represented by the "isolated" Kramers doublet with the almost maximal projection on the *z* axis (and the minimal projection on the *x* axis). The nearest excited doublet is separated by a gap of about 100 K, and its maximal component of the *g*-tensor is oriented along the *x* axis ($g^{ex}_x \gg g^{ex}_z$). In the J^-, $J_z$representation, the wave function of the lower doublet is $\pm 0.98|15/2\rangle$; for the excited doublet, it is $\{\pm 0.57|\mp 9/2\rangle, \pm 0.62|\pm 7/2\rangle, \pm 0.51|\mp 1/2\rangle\}$. It is exactly this specific feature of the spectrum and wave functions of the Dy^{3+} ion that determines its Ising prop-

erties and is favorable for the crossing of levels when the field is oriented in the basal plane.

As would be expected, the ground doublet with the maximal component g_z^{gr} for $H \parallel [100]$ and $H \parallel [110]$ splits weakly, whereas the first excited doublet splits much more strongly and differently for the two symmetric field directions in the basal plane. As a result, in the field of about 140 and 200 kOe, respectively, the split sublevels of excited doublets with large projections of the magnetic moment $\langle M_\perp \rangle$ approach the ground level with a small projection $\langle M_\perp \rangle$; this phenomenon results in a sharp increase of the magnetic moment.

A characteristic feature of the Zeeman effect in this geometry is the presence of a substantial gap of about 30 K between the interacting levels, which is retained in the critical field H_c. This is due to the fact that the field mixes the wave functions of those levels. For the field oriented along the tetragonal axis, there is usually no gap between the lower levels under crossover in the framework of the Hamiltonian used. The gap appears only at a small deviation of the field from the symmetry axis; i.e., when the component of the field in the basal plane becomes nonzero. The presence of a finite gap under crossover explains the specific features of the magnetic properties in this case, in particular, their dependence on temperature and magnetic field. It is clear that a disorientation of the field both in the basal plane and its going out of the basal plane is less important than for the field oriented along the tetragonal axis.

The comparison of spectra calculated with regard (solid curves) and without regard (dashed curves) to quadrupole interactions (Fig. 5) shows that their role becomes much more important in the fields stronger than the crossover one and that they cause a sizable change of the magnetic and magnetoelastic characteristics in the crossover region. In this case, even in the paramagnetic and quadrupole disordered phases, the spectrum depends not only on the magnetic field, but also on temperature, since the field dependences of the quadrupole moments $\langle O_2^0 \rangle (H)$, $\langle O_2^2 \rangle (H)$, and $\langle P_{xy} \rangle (H)$ vary with temperature. In the absence of the field, the dependence of the spectrum on temperature is caused by the quadrupole moment $\langle O_2^0 \rangle (T)$. We stress that the family of rare-earth zircons is a unique class of Jahn–Teller magnetics, in which full-symmetric quadrupole interactions cause the effects observed.

5.2. Magnetization Curves and the Magnetocaloric Effect

In order to interpret magnetic and magnetoelastic properties in pulsed fields with sufficiently small pulse duration, one must calculate adiabatic magnetization processes. In our experiment, the rate of the field increase was close to the estimate of the upper bound of the adiabaticity condition given in [16]; thus, we assume the magnetization process of the $DyPO_4$ monocrystal in our experiment to be adiabatic. The subsequent comparison of the calculation results with the experimental ones confirms this assumption.

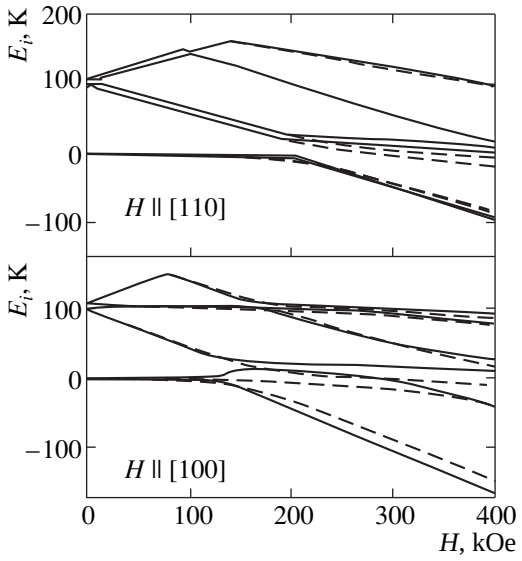


Fig. 5. Energy variation of the lower levels of the Dy^{3+} ion in $DyPO_4$ in a magnetic field (the Zeeman effect) along the [110] (above) and [100] (below) axes without regard (dashed curves) and with regard (solid curves) to the quadrupole interactions (G^α = 1.5 mK, G^γ = 4.4 mK, and G^δ = 16.6 mK) at T = 5 K.

When calculating the magnetic characteristics for every field value in the range from 0 to 400 kOe with the step ΔH = 2 kOe, the complete Hamiltonian was numerically diagonalized in order to determine the spectrum and the wave functions of the Dy^{3+} ion, and the "elementary" magnetocaloric effect ΔT was calculated when the field changed from H to $H + \Delta H$:

$$\Delta T = -\left(\frac{\partial M}{\partial T}\right)_H \frac{T \Delta H}{C_H}. \qquad (8)$$

In formula (8), the total heat capacity of the crystal, C_H, includes the lattice heat capacity $C_{\text{lat}} = (12\pi^4 k_B \nu/5)(T/T_D)^3$ (the Debye temperature for the phosphate lattice is T_D = 275 K [17] and ν = 6) and the magnetic heat capacity C_{mag} calculated for every value of the field and temperature on the basis of the rare-earth ion spectrum. These data make it possible to calculate the adiabatic magnetization of $DyPO_4$ and the sample temperature for given directions of the field. The isothermal and adiabatic magnetization curves along the three symmetric directions [001], [100], and [110] and the corresponding curves of the magnetocaloric effect ΔT for the initial temperatures T = 5 and 20 K are shown in Fig. 6. The smeared jumps of magnetization in the field oriented along the axes [100] and [110] correspond to approaching lower energy levels shown in Fig. 5.

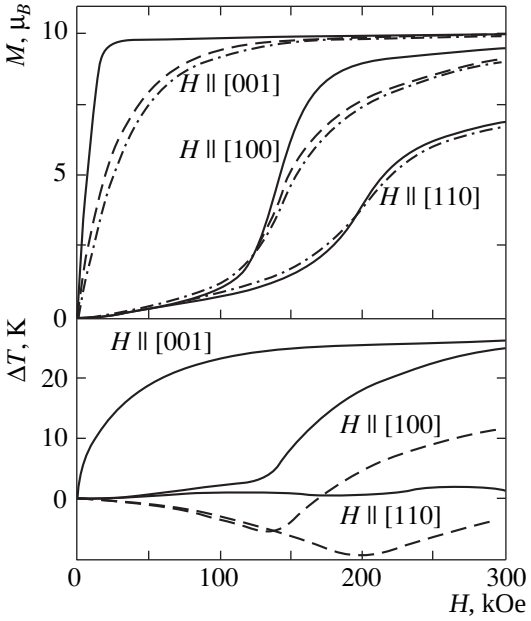


Fig. 6. Isothermal (solid curves, $T = 5$ K), adiabatic (dashed curves, $T_{st} = 5$ K and dot-and-dash curves, $T_{st} = 20$ K) magnetization curves $M(H)$ (above) and the magnetocaloric effect ΔT (below; solid curves at $T_{st} = 5$ K and dashed ones at $T_{st} = 20$ K) for $DyPO_4$ single crystal in the field oriented along the [001], [100], and [110] axes calculated with regard to quadrupole interactions.

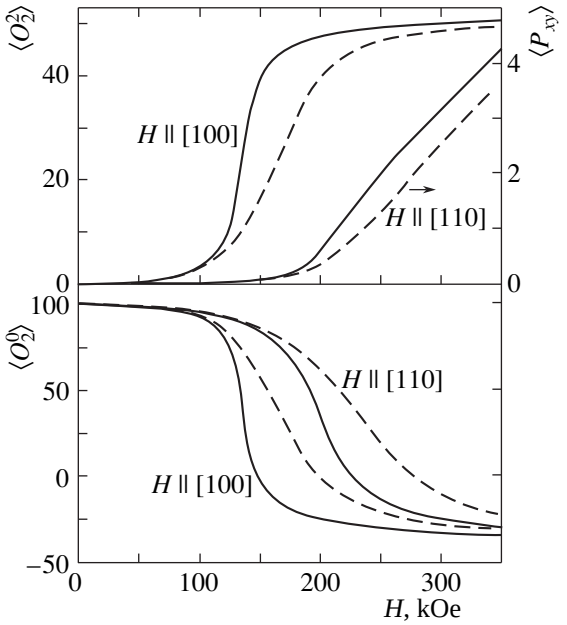


Fig. 7. The quadrupole moments $\langle O_2^0 \rangle$ (below) and $\langle O_2^2 \rangle$, $\langle P_{xy} \rangle$ (above) of $DyPO_4$ single crystal at $T = 5$ K and the field orientation along the [100] and [110] axes calculated in the isothermal mode with regard (solid curves) and without regard (dashed curves) to quadrupole interactions.

For $H \parallel [100]$ in the critical field $H_c \sim 140$ kOe, the magnetization of $DyPO_4$ increases in a jump from about $1\mu_B$ up to about $10\mu_B$, which is close to the saturation value. For $H \parallel [110]$, the critical field $H_c \sim 200$ kOe is stronger, the jump is smaller, and the magnetic moment does not reach the saturation level after the crossover. This fact gives reason to assume that there exists another crossover in stronger fields. Indeed, calculations show that for $H_{c2} \sim 3000$ kOe another crossing of the levels occurs, at which the magnetic moment increases in a jump by about $2\mu_B$ and reaches the saturation level. At low temperatures, the magnetization jumps in the isothermal mode vary insignificantly. Due to a finite magnitude of the spectrum gap, the jump on the curves $M(H)$ at $T = 5$ K for both field directions remains more smeared than, for example, for $HoVO_4$ under similar conditions [18]. In the case of the crossover without a gap, the magnetization jump and, respectively, the differential susceptibility maximum in the framework of the Hamiltonian used become infinitely sharp while approaching absolute zero.

For $H \parallel [001]$ and $H \parallel [100]$, the adiabatic magnetization curves are flatter than the isothermal ones; this is due to the heating of the $DyPO_4$ crystal in the field. The maximal magnetocaloric effect $\Delta T \sim 25$ K is observed in the field oriented along the easy magnetization axis [001]. For both orientations of the field in the basal plane, the variation of the sample temperature is small in fields weaker than the crossover field ($\Delta T < 5$ K), since the magnetization is small. In fields stronger than H_c, the sample heats noticeably for $H \parallel [100]$ and remains almost at the initial temperature for $H \parallel [110]$. Thus, for $H \parallel [100]$, the sample temperature varies monotonically with the field starting from the initial temperature $T_{st} = 5$ K; however, the rate of the temperature variation is different. For higher starting temperatures, the magnetocaloric effect becomes nonmonotonic for both field orientations in the basal plane, and it is accompanied by cooling of the sample in fields weaker than H_c (see dashed curves in Fig. 6). This is related to the complex behavior of $(\partial M/\partial T)_H$ in the crossover region.

5.3. Magnetostriction

The magnetostriction deformation $\lambda_{\beta 1 \beta 2 \beta 3}^{\alpha 1 \alpha 2 \alpha 3}$ for $H \parallel [100]$ is described by two quadrupole moments $\langle O_2^0 \rangle$ and $\langle O_2^2 \rangle$, whose linear combination with the corresponding coefficients $A^{\alpha 1}$, $A^{\alpha 2}$, and A^{γ} determines the crystal deformation along the three mutually perpendicular axes [100], [010], and [001] (see formulas (6) and (7)). The variation of the quadrupole moments $\langle O_2^0 \rangle$ and $\langle O_2^2 \rangle$ in the field $H \parallel [100]$ in the isothermal mode at $T = 5$ K is shown in Fig. 7 (solid curves). The quadrupole moment $\langle O_2^0 \rangle$ for the tetragonal crystal is nonzero even in the absence of the field; we see from

Fig. 7 that it slightly varies in weak fields, but decreases jumpwise in the critical field, changes its sign, and tends to saturation in fields stronger than the critical one. On the contrary, the quadrupole moment $\langle O_2^2 \rangle$ is zero in the absence of the field, but increases sharply in the critical field and reaches a magnitude of ~50 in fields stronger than H_c; this value is comparable with that observed under quadrupole ordering in rare-earth zircons. Since these quadrupole moments have different dependences on the field, their linear combinations can lead to nonmonotonic field dependences of the linear magnetostriction along certain directions in the crystal.

The deformation along the tetragonal axis λ_{100}^{001} is completely described by the quadrupole moment $\langle O_2^0 \rangle$, and its sign is determined by the relation of the coefficients $A^{\alpha 1}$ and $A^{\alpha 2}$, which have different signs. In addition, the deformations λ_{100}^{100} and λ_{100}^{010} along the [100] and [010] axes include a contribution of the quadrupole moment $\langle O_2^2 \rangle$, which is the same in magnitude but has different signs for two axes. Due to relations between various magnetoelastic and elastic coefficients, the linear magnetostriction in $DyPO_4$ is highly anisotropic and is maximal along the [100] axis when all three terms in (7) have the same sign; it is minimal along the [010] axis.

For $H \parallel [110]$ ($H_c \sim 200$ kOe), the quadrupole moments $\langle O_2^0 \rangle$ and $\langle P_{xy} \rangle$ exhibit a similar behavior. Moreover, the variation of the moment $\langle O_2^0 \rangle$ is almost the same, and that of the moment $\langle P_{xy} \rangle$ is by an order of magnitude less than the variation of $\langle O_2^2 \rangle$ for $H \parallel [100]$. We note that the quadrupole moment $\langle P_{xy} \rangle$ does not tend to saturation when $H > H_c$, in contrast to the moment $\langle O_2^2 \rangle$. The field dependences of the moments $\langle O_2^0 \rangle$, $\langle O_2^2 \rangle$, and $\langle P_{xy} \rangle$ become smeared as the temperature increases, and, in the process, the critical field H_c increases for both orientations of the field. Thus, the critical field is minimal along the [100] axis and increases when the field deviates from this axis both in the basal plane and in the direction of the tetragonal axis. For the deviation angles $\Delta\varphi = 5°$ and $\Delta\theta = 5°$, the increase of H_c is $\Delta H_c = 7$ and 10 kOe, respectively.

Since the approach of the energy levels in $DyPO_4$ is accompanied by a substantial increase or variation of the quadrupole moments, calculations must take into account quadrupole interactions. Accounting for quadrupole interactions of the α and $\gamma(\delta)$ symmetry results in a noticeable decrease of the critical field and makes the jumps sharper for both orientations of the field (cf. the solid and dashed curves in Fig. 7). We note that the dependences calculated with and without taking into account quadrupole interactions are almost identical in weak fields and begin to deviate in the vicinity of the critical field. This means that the contribution of quadrupole interactions is much greater in fields stronger than the critical one.

The adiabatic curves of linear magnetostriction along the [100], [010], and [001] axes calculated at $H \parallel [100]$ and along the [001] axis at $H \parallel [110]$ are presented in Figs. 1–4 with the purpose of comparing them with the experimental data. The calculation results are in good agreement with the field dependences of the linear magnetostriction along the three mutually perpendicular directions, including the nonmonotonic behavior in weak fields and the change of sign for $\lambda(H)$ at $H \parallel [100]$ (Fig. 3). Such nonmonotonic dependences of $\lambda(H)$ are observed only for a certain relationship of the contributions of the quadrupole moments $\langle O_2^0 \rangle$ and $\langle O_2^2 \rangle$, i.e., for a certain relationship between the coefficients $A^{\alpha 1}$, $A^{\alpha 2}$, and A^{γ}. In particular, to describe the experimental curves $\lambda(H)$, one must increase the coefficient A^{γ} by about 10% compared to that determined by the measurements made in weak fields.

The absolute values of the calculated linear magnetostriction are regularly greater than the experimental ones by a factor of 1.5–2. As has already been discussed above, this can be explained by the lamination of the sample along the cleavage planes in pulsed fields. The small thickness of the sample compared to the sensor results in the experimental underestimation of the magnetostriction magnitude.

5.4. Magnetostriction Susceptibility

Experimental data and theoretical analysis show that the anomalous behavior of the curves $\lambda(H)$, as well as $M(H)$, is very sensitive to quadrupole interactions. This dependence can be conveniently analyzed on the differential curves dQ_2^0/dH and dQ_2^2/dH, which are shown in Fig. 8. It is seen that the field dependences dQ_2^0/dH and dQ_2^2/dH are similar in the vicinity of the maximum, but are slightly different when the field is substantially less or greater than the critical one. We examine how the behavior of the dependences changes with temperature by the example of the quadrupole moment Q_2^0. At temperatures less than 1 K, the field dependences dQ_2^0/dH in the crossover region coincide; a small difference is observed only in weak fields. As the temperature increases, the maxima move to the region of stronger fields and become broader for both quadrupole moments Q_2^0 and Q_2^2. The dependences of the critical field $H_c(T)$ and the width of the maximum $\Delta W(T)$ (determined at the half height of the peak dQ_2^0/dH) on temperature are nonmonotonic (see the inset), but exhibit extrema at $T \sim 3$ 3 and 5 K, respectively. These dependences are practically identical to the similar curves determined on the basis of the anom-

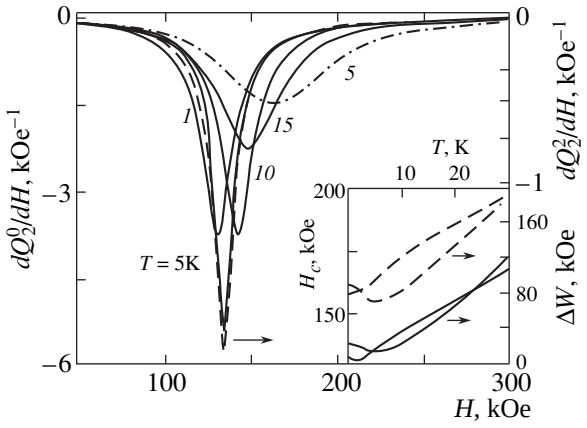


Fig. 8. Isothermal derivatives dQ_2^0/dH and dQ_2^2/dH (dashed curve at $T = 5$ K) of $DyPO_4$ at the field orientation $H \parallel [100]$ and at various temperatures calculated with regard (solid curves, $G^\alpha = 1.5$ mK, $G^\gamma = 4.4$ mK) and without regard (dash-and-dot curve at $T = 5$ K) to quadrupole interactions. The inset shows the dependences of the critical field H_c and the breadth of the maximum ΔW on temperature calculated with regard (solid curves) and without regard (dashed curves) to quadrupole interactions.

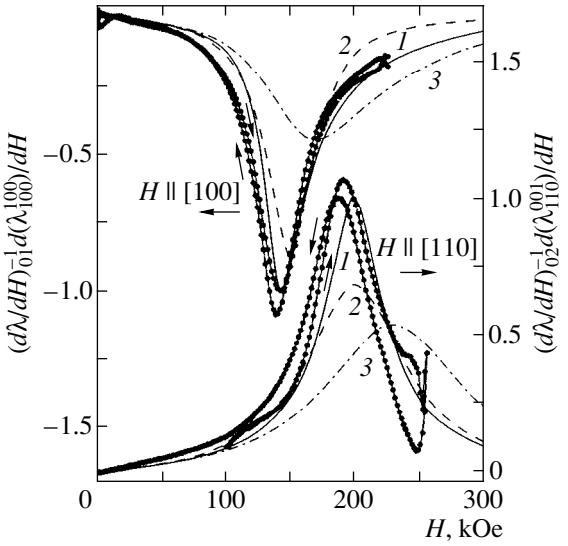


Fig. 9. Experimental (dots) and calculated adiabatic (*1*) and isothermal (*2*) curves of the magnetostriction susceptibility $d\lambda_{100}^{100}/dH$ and $d\lambda_{110}^{001}/dH$ for the $DyPO_4$ crystal normalized to their maximum values ($(d\lambda/dH)_{01} = (d\lambda_{100}^{100}/dH)_{max}$, $(d\lambda/dH)_{02} = (d\lambda_{110}^{001}/dH)_{max}$) in the fields oriented as $H \parallel [100]$ and $H \parallel [110]$, respectively at $T = 15$ K. The calculations are performed without regard (dash-dot curves, *3*) and with regard (curves *1*, *2*; $G^\alpha = 1.5$ mK, $G^\gamma = 4.4$ mK, and $G^\delta = 16.6$ mK) to quadrupole interactions.

alies of the magnetic moment dM/dH. The extrema of the curves $H_c(T)$ and $\Delta W(T)$ correlate with the nonmonotonic dependence of the initial magnetic susceptibility $\chi_a(T)$ along the [100] axis at low temperatures [12].

The comparison of the curves for $T_{st} = 5$ K calculated with regard (the solid curve) and without regard (the dashed curve) to the quadrupole interactions show that these interactions reduce the critical field and make it closer to the experimental value. In the process, the height of the maximum dQ_2^0/dH increases, and its width decreases almost by a factor of three. When the calculation is performed without taking into account the quadrupole interactions, the dependence $H_c(T)$ moves almost in parallel by 20 kOe to the region of greater fields, and the width of the peak $\Delta W(T)$ at low temperatures increases from 25 to 75 kOe. At high temperatures, the effect of quadrupole interactions is smaller.

The differential curves also make it possible to compare the experimental data with the calculated ones in the isothermal and adiabatic modes. Figure 9 shows the experimental and calculated adiabatic and isothermal derivatives $d\lambda_{100}^{100}/dH$ and $d\lambda_{100}^{001}/dH$ normalized to their maximum values at $T = 15$ K for $H \parallel [100]$ and $H \parallel [110]$. The calculated adiabatic curve $d\lambda_{100}^{100}/dH$ at $T_{st} = 15$ K describes the experimental curve rather accurately. A small excess of the critical field H_c over the experimental data can be eliminated by increasing the quadrupole constants by about 30%. We stress that all calculations were performed without any adjustable parameters and were based only on the magnitudes determined by measurements in relatively small fields. For the isothermal curve $d\lambda_{100}^{100}/dH$, the maximum is slightly shifted to stronger field values compared to the adiabatic curve. The main difference of the isothermal and adiabatic curves $d\lambda_{100}^{100}/dH$ is in their shape. The isothermal curve is almost symmetric; i.e., the location of the peak center is the same at any height. In the adiabatic mode, the curve is asymmetric and is more extended for $H > H_c$; this is due to the heavy heating of the sample in fields stronger than the critical one. The comparison shows that the experimental curve is as asymmetric as the adiabatic curve $d\lambda_{100}^{100}/dH$. Thus, the experiment testifies that the magnetization process of the samples under consideration in pulsed fields of the indicated duration is close to adiabatic.

For the field orientation $H \parallel [110]$, the curves $d\lambda_{100}^{001}/dH$ exhibit a similar behavior. Quadrupole interactions also decrease the critical field by about 20 kOe, and the maximum of the curves $d\lambda_{110}^{001}/dH$ becomes greater and narrower approximately by a factor of two, which is in much better agreement with the experiment. When the calculations are performed with the known parameters of the quadrupole interactions determined from independent experiments, the critical field H_c is

almost the same or is very close to the experimental value both for $H \parallel [100]$ and $H \parallel [110]$. We stress that the values of the constants G^μ were determined on the basis of the magnetoelastic coefficients B^μ found by measuring the parastriction and the elastic constants C^μ averaged over the rare-earth phosphate series. In the process, only the contribution to G^μ caused by the one-ion magnetoelastic interaction was taken into account, since no reliable data on the parameters of the pair quadrupole interaction (which usually yields a considerably smaller contribution in zircons) are available. Taking into account this fact, we should admit that the agreement with the experimental data is very good.

6. CONCLUSION

In conclusion, we formulate the main results obtained in this study. The investigation of $DyPO_4$ shows that the crossover in rare-earth compounds at low temperatures is accompanied not only by a jump of the magnetic moment, but also causes anomalies of the linear magnetostriction, which are determined by variations of quadrupole moments of the rare-earth ion. In our opinion, investigation of the magnetostriction anomalies under crossover is rather informative, since it makes it possible to determine or refine the values of magnetoelastic coefficients in addition to determining the critical field H_c. In addition, the magnetostriction anomalies, which are directly related to the variation of the quadrupole moments, enable one to estimate the role of quadrupole interactions in the effects under study.

A specific feature of $DyPO_4$ is that there exists a finite gap of ~30 K between the approaching levels under crossover when the field H is oriented in the basal plane. Due to this fact, the magnetization and magnetostriction curves remain rather smeared down to very low temperatures.

A considerable advantage of the systems under study is that they are relatively simple, and there is reliable information on the parameters of interactions available. This enables one to compare theoretical and experimental data not only qualitatively, but also quantitatively. In particular, this comparison shows that under pulsed fields the magnetization process is close to the adiabatic one and is accompanied by a considerable magnetocaloric effect.

It has been established that the contribution of the quadrupole interactions of the α and $\gamma(\delta)$ symmetry for $H \parallel [100]$ and $H \parallel [110]$ greatly increases under crossover, which opens up a new method for investigating these interactions and determining or refining the quadrupole interaction constants. In the absence of the field, the $DyPO_4$ crystal does not belong to the class of Jahn–Teller compounds; however, for $H \parallel [100]$ in the vicinity of the crossover, the contribution of the quadrupole interactions of the γ symmetry becomes comparable in magnitude with that observed for the Jahn–Teller magnetic $DyVO_4$, which exhibits spontaneous quadrupole ordering. The magnetic field forms an electronic structure such that the quadrupole effects for the Dy ion in the phosphate in the vicinity of the crossover field H_c and in the vanadate at $H = 0$ become comparable. In this case, not only the magnetic moment increases jumpwise in the critical field H_c, but also the quadrupole moment $\langle O_2^2 \rangle$, which is characteristic of the stimulated Jahn–Teller transition. It follows from the calculations that the Zeeman effect for lower levels in the presence of quadrupole interactions is of a more complex nature and considerably depends on the constants of the quadrupole interactions. In this connection, not only is the study of thermodynamic characteristics of interest, but also the direct investigation of the Zeeman effect in $DyPO_4$ under crossover.

ACKNOWLEDGMENTS

This work was partially supported by the Russian Foundation for Basic Research, project no. 00-02-17756.

REFERENCES

1. G. A. Gehring and K. A. Gehring, Rep. Prog. Phys. **38**, 1 (1975).
2. P. Morin and Z. Kazei, Phys. Rev. B **55**, 8887 (1997).
3. Z. A. Kazei, N. P. Kolmakova, V. V. Platonov, *et al.*, in *Abstracts of 3rd International Conference on f-Elements, Paris, 1997.*
4. J. E. Battison, A. Kasten, M. J. M. Leask, and J. B. Lowry, J. Phys. C **10**, 323 (1977).
5. T. Goto, A. Tamaki, T. Fujimura, and H. Unoki, J. Phys. Soc. Jpn. **55**, 1613 (1986).
6. Z. A. Kazei, N. P. Kolmakova, V. V. Platonov, *et al.*, Physica B (Amsterdam) **284**–**288**, 1483 (2000).
7. Z. A. Kazei, N. P. Kolmakova, R. Z. Levitin, *et al.*, Physica B (Amsterdam) **246**–**247**, 483 (1998).
8. J. C. Wright, H. W. Moos, J. H. Colwelle, *et al.*, Phys. Rev. B **3**, 843 (1971).
9. I. R. Jahn and S. H. Smith, Phys. Status Solidi B **68**, 531 (1975).
10. R. Z. Levitin, V. N. Milov, Yu. F. Popov, and V. V. Snegirev, Physica B (Amsterdam) **177**, 59 (1992).
11. P. Morin, J. Rouchy, and D. Schmitt, Phys. Rev. B **37**, 5401 (1988).
12. P. Morin and Z. Kazei, J. Phys.: Condens. Matter **11**, 1289 (1999).
13. J. Hoffmann, Phys. Status Solidi B **165**, 517 (1991).
14. E. de Lacheisserie, Ann. Phys. **5**, 267 (1970).
15. C.-K. Loong, L. Soderholm, J. Simon Xue, *et al.*, J. Alloys Compd. **207**, 165 (1994).
16. R. Z. Levitin, V. V. Snegirev, A. V. Kopylov, *et al.*, J. Magn. Magn. Mater. **170**, 223 (1997).
17. A. Kasten, H. G. Kahle, P. Klofer, and D. Schafer-Siebert, Phys. Status Solidi B **144**, 423 (1987).
18. P. Morin, J. Rouchy, and Z. Kazei, Phys. Rev. B **51**, 15103 (1995).

Translated by A. Klimontovich

Journal of Experimental and Theoretical Physics, Vol. 92, No. 4, 2001, pp. 710–714.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 4, 2001, pp. 816–821.

SOLIDS
Electronic Properties

The Theory of Coherent Resonance Tunneling of Interacting Electrons

V. F. Elesin

Moscow State Institute of Engineering Physics, Kashirskoe sh. 31, Moscow, 115409 Russia

Received December 5, 2000

Abstract—Analytical solutions of the Schrödinger equation for a two-barrier structure (resonance-tunnel diode) with open boundary conditions are found within the model of coherent tunneling of interacting electrons. Simple expressions for resonance current are derived which enable one to analyze the current–voltage characteristics, the conditions of emergence of hysteresis, and singularities of the latter depending on the parameters of resonance-tunnel diode. It is demonstrated that the hysteresis is realized if the current exceeds some critical value proportional to the square of resonance level width. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Resonance tunneling through nanostructures attracts ever increasing attention in view of its use in microwave-frequency devices. Resonance tunneling is closely associated with the effect of quantum interference of electrons and with the emergence of resonance levels of spatial quantization.

In a real situation, the need arises to include the effect of the ever present interaction between electrons on the processes of quantum interference and resonance tunneling. This is the more important because one can expect a high sensitivity of resonance tunneling to electron–electron interaction. This expectation follows from the fact that the resonance level shift due to interaction by a value that is small compared with the electron energy ε_R but comparable with the resonance level width $\Gamma \ll \varepsilon_R$ causes a sharp variation of the resonance current.

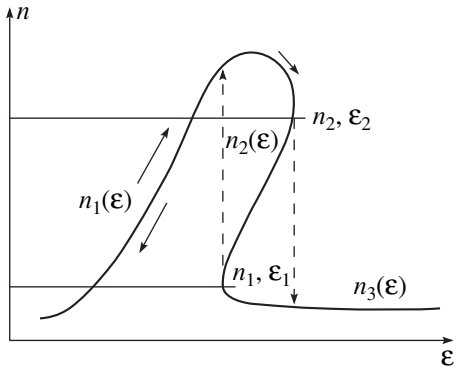
The simplest nanostructure in which resonance tunneling shows up is a two-barrier quantum well, the so-called resonance-tunnel diode. Its current-voltage characteristic (*IVC*) exhibits a descending portion with a negative differential resistance [1]. In addition, as was first observed by Goldman *et al.* [2], the *IVC* of a resonance-tunnel diode sometimes exhibits a hysteresis (see figure). The voltage dependence of current has three branches, and transitions between them are performed in jumps.

At least two interpretations of the hysteresis effect exist. According to the first version [2, 3], the hysteresis arises because of the accumulation of charge or interelectron interaction. The second interpretation [4] is based on the effect of the external circuits on the current in a resonance-tunnel diode characterized by a negative differential conduction. No consensus of opinion exists at present as regards the mechanism of hysteresis (see, for example, [5, 6]).

The theory of the effect of charge accumulation and interelectrode interaction on the *IVC* of a resonance-tunnel diode was treated in a number of studies which were performed using largely numerical methods. The phenomenological model was first studied by Sheard and Toombs [3]. Incoherent tunneling was assumed, and a *IVC* was found that exhibited hysteresis due to charge accumulation.

A more rigorous model was treated by Jun Zang and Birman [6], who used the Keldysh technique within the approximation of incoherent tunneling to derive expressions for the current and equations for the electron concentration in the quantum well. The Hartree approximation with local interelectron potential was used. The equation was solved numerically for the typical parameters of a quantum well and for two values of width $\Gamma = 2$ meV and $\Gamma = 5$ meV. A correlation was observed between the hysteresis region and the width Γ.

Note that both studies [3, 6] were performed within the model of incoherent tunneling (which is in fact semiphenomenological). Indeed, this model assumes



The electron concentration in a tunnel diode as a function of electron energy at a higher-than-critical current.

1063-7761/01/9204-0710$21.00

the existence of resonance level and does not involve explicit treatment of quantum interference. In addition, it is the perturbation theory that is in fact used instead of open boundary conditions for the probability of tunnel transitions from contacts to the quantum well (the so-called tunneling Hamiltonian method).

Therefore, it appears useful to study the effect of interelectron interaction within the model of coherent tunneling. In this model, quantum interference and resonance levels of spatial quantization arise automatically, and the open boundary conditions provide for a rigorous formulation of the problem. One can expect that this approach will help introduce additional arguments to the above-mentioned discussion (see [5, 6]).

It is the objective of this study to construct a consistent model of coherent tunneling of interacting electrons. The self-consistent Hartree–Fock approximation was used to find analytical solutions to the Schrödinger equation with open boundary conditions. Simple expressions for resonance current have been derived which enable one to analyze the conditions of the emergence of hysteresis and singularities of the latter depending on the parameters of resonance-tunnel diode.

It has been demonstrated for the first time ever that hysteresis is realized if the current exceeds some critical value J_c. Correlations have been found between J_c and the region of hysteresis and the structure parameters, in particular, the resonance level width.

This paper is organized as follows. Section 2 contains the formulation of the problem and derivation of the basic equations. Section 3 deals with the solution of the Schrödinger equation and derivation of the general expression for current and of the equation for electron concentration in the quantum well. In Sections 4 and 5, the dependence of current on the electron energy, the conditions of the emergence of hysteresis, and singularities of the latter are analyzed.

2. FORMULATION OF THE PROBLEM: BASIC EQUATIONS

We will study the following model of coherent tunneling of interacting electrons (see [7]). A steady electron flow proportional to q^2, with an energy ε approximately equal to ε_R, is delivered from the left to a one-dimensional quantum well with δ-functional barriers at points $x = 0$ and $x = a$ ($x = -\infty$). It is assumed that the electrons interact with one another inside the quantum well.

In the Hartree–Fock approximation, the Schrödinger equation has the form

$$-\frac{d^2\psi}{dx^2} + \alpha[\delta(x) + \delta(x-a)]\psi + U(x)\psi = \varepsilon\psi, \quad (1)$$

$$U(x) = \int dx' |\psi(x')|^2 V(x-x'), \quad (2)$$

where $U(x)$ is the self-consistent potential, $V(x)$ is the energy of interaction between electrons, and α is the barrier power. Here and below, $\hbar = 2m = 1$. The potential $U(x)$ may describe the effect of charge accumulation as well (see [6]). For simplicity, we will assume that the interelectron interaction is local (as in [6]),

$$V(x-x') = g\delta(x-x'). \quad (3)$$

Then, Eq. (1) takes the final form

$$\frac{d^2\psi}{dx^2} + \varepsilon\psi - g\psi|\psi|^2 - \alpha[\delta(x) + \delta(x-a)]\psi = 0. \quad (4)$$

We use the boundary conditions (see, for example, [7])

$$\psi(0)\left(1 - \frac{\alpha}{ip}\right) + \frac{1}{ip}\frac{d\psi(0)}{dx} = q,$$
$$\psi(a)\left(1 - \frac{\alpha}{ip}\right) - \frac{1}{ip}\frac{d\psi(a)}{dx} = 0, \quad \varepsilon = p^2, \quad (5)$$

which describe the electron flow from the left, their reflection, and departure to the region $x > a$.

The resonance current $J(\varepsilon)$ through the quantum well is found from the ordinary expression

$$J(e) = -ie\left[\psi^*\frac{d\psi}{dx} - \psi\frac{d\psi^*}{dx}\right]. \quad (6)$$

In the absence of interaction, the resonance current is given by the known expression (see, for example, [7])

$$J(\varepsilon) = \frac{Q\Gamma^2}{2[(\varepsilon - \varepsilon_R)^2 + \Gamma^2]},$$
$$\Gamma = \frac{4p^3}{\alpha^2 a}, \quad Q = q^2 p, \quad (7)$$

and a is the size of the quantum well.

Note that formula (7) was derived for the most interesting case of a quantum well with "strong barriers," when $\Gamma \ll \varepsilon_R$. It is in this limit that the resonant properties of the quantum well are realized most effectively.

3. GENERAL EXPRESSION FOR RESONANCE CURRENT OF INTERACTING ELECTRONS

The analytical solution of Eq. (4) with open boundary conditions (5) in the general case presents certain difficulties. However, the derivation of the solution may be simplified if one takes into consideration the high sensitivity (already noted above) of resonance current to the shift of resonance level due to interelectron interaction.

One can expect (as will be confirmed below) that even a small (compared with ε_R) energy shift $\Delta\varepsilon \sim \Gamma$ will cause a considerable variation of the current $J(\varepsilon)$. Therefore, the term that is nonlinear in (4) may be assumed to be small, and, in order to find the solution

of Eq. (4), one can use the methods described, for example, in [8]. We will seek the solution of Eq. (4) in the form

$$\psi(x) = A(x)e^{ipx} + B(x)e^{-ipx}, \tag{8}$$

where $A(x)$ and $B(x)$ are slowly varying complex functions of x. In so doing, the rate of variation of the functions $A(x)$ and $B(x)$ is proportional to the smallness of the nonlinear term.

If one substitutes solution (8) into Eq. (4) and performs averaging over fast-varying functions $\exp(\pm ipx)$ (for more detail, see [8]), the following equations are derived for $A(x)$ and $B(x)$:

$$\begin{aligned}\frac{dA}{dx} &= \frac{gA}{2ip}[|A(x)|^2 + 2|B(x)|^2],\\ \frac{dB}{dx} &= -\frac{gB}{2ip}[|B(x)|^2 + 2|A(x)|^2].\end{aligned} \tag{9}$$

One can readily demonstrate that the quantities $|A|^2$ and $|B|^2$ are independent of the coordinate,

$$\frac{d}{dx}|A|^2 = \frac{d}{dx}|B|^2 = 0. \tag{10}$$

In view of this fact, we find the solution of the set of equations (9),

$$\begin{aligned}A(x) &= A_0\exp\left\{-\frac{ixg}{2p}(|A_0|^2 + 2|B_0|^2)\right\},\\ B(x) &= B_0\exp\left\{\frac{ixg}{2p}(|B_0|^2 + 2|A_0|^2)\right\},\end{aligned} \tag{11}$$

where A_0 and B_0 are constants which may be found from the boundary conditions (5).

After some computations, we derive expressions for A_0, B_0, and determinant Δ_0,

$$A_0 = \frac{q}{\Delta_0}\left(2 - \frac{\alpha}{ip} - g_B\right), \tag{12}$$

$$B_0 = \frac{q}{\Delta_0}\left(\frac{\alpha}{ip} - g_A\right)\exp(2ipa - ipa\tilde{g}), \tag{13}$$

$$\begin{aligned}\Delta_0 &= \left(2 - \frac{\alpha}{ip} - g_A\right)\left(2 - \frac{\alpha}{ip} - g_B\right)\\ &-\left(\frac{\alpha}{ip} - g_A\right)\left(\frac{\alpha}{ip} - g_B\right)\exp(2ipa - ipa\tilde{g}),\end{aligned} \tag{14}$$

where

$$\begin{aligned}g_A &= \frac{g}{2p^2}(|A_0|^2 + 2|B_0|^2),\\ g_B &= \frac{g}{2p^2}(|B_0|^2 + 2|A_0|^2),\\ \tilde{g} &= \frac{3g}{2p^2}n_0, \quad n_0 = |A_0|^2 + |B_0|^2.\end{aligned} \tag{15}$$

Generally speaking, the quantities g_A, g_B, and $\tilde{g}$ themselves depend on A_0 and B_0. Therefore, the set of Eqs. (12)–(15) is self-consistent.

We substitute solutions (8) and (11) into Eq. (6) to find the current,

$$\begin{aligned}J(\varepsilon) &= e[|A_0|^2 - |B_0|^2]\\ &\times\left\{2p + \frac{gn_0}{p} + \frac{ig}{2p}(A_0B_0^*e^{2ipx} - \text{c.c.})\right\}.\end{aligned} \tag{16}$$

4. EQUATION FOR THE CONCENTRATION OF INTERACTING ELECTRONS

As we will see, the quantity that defines the behavior of current (16) is the resonant denominator Δ_0. In the most interesting case of "strong barriers", when $p/\alpha \ll 1$, $\Gamma/\varepsilon_R \ll 1$, and $\varepsilon \approx \varepsilon_R$, expression (14) for Δ_0 may be represented as

$$\begin{aligned}|\Delta_0|^2 &= \frac{16}{\Gamma^2}[(\varepsilon - \varepsilon_R - p^2\tilde{g})^2 + \Gamma^2],\\ \varepsilon_R &\approx \left(\frac{\pi}{a}\right)^2\left(1 - \frac{4}{\alpha a} - \frac{12}{\alpha^2a^2}\right).\end{aligned} \tag{17}$$

In deriving the latter expression, it was further assumed that the resonance level shift

$$\Delta\varepsilon = p^2\tilde{g} = \frac{3}{2}gn_0 \tag{18}$$

was small compared with ε_R. At the same time, it follows from (17) that the shift $\Delta\varepsilon$, which is commensurable with Γ, causes a strong variation of the resonant denominator and, consequently, of the current $J(\varepsilon)$ given by Eq. (16). Therefore, our initial assumption that the nonlinear term in Eq. (4) may be regarded as small proves to be justified. The smallness of the parameter $\Delta\varepsilon/\varepsilon_R$ enables one to simplify the expressions for A_0, B_0, and $J(\varepsilon)$ by omitting the respective small corrections. As a result, we derive the expression for current

$$J(\varepsilon) = \frac{eQ\Gamma^2}{2\left[\left(\varepsilon - \varepsilon_R - \frac{3}{2}gn_0\right)^2 + \Gamma^2\right]} \tag{19}$$

and the self-consistent equation for concentration

$$n_0(\varepsilon) = \frac{Q\Gamma}{2a\left[\left(\varepsilon - \varepsilon_R - \frac{3}{2}gn_0\right)^2 + \Gamma^2\right]}. \quad (20)$$

One can readily see from Eqs. (19) and (20) that the current $J(\varepsilon)$ and the concentration $n_0(\varepsilon)$ are related by

$$J(\varepsilon) = en_0(\varepsilon)a\Gamma. \quad (21)$$

Therefore, the current is expressed in terms of the electron concentration $n_0(\varepsilon)$ in the quantum well, which is to be found from Eq. (20).

5. ANALYSIS OF THE RESONANCE CURRENT OF INTERACTING ELECTRONS

Equation (20) may be conveniently reduced to the dimensionless form

$$n = f/[1 + (x - n)^2], \quad (22)$$

$$n = \frac{3gn_0}{2\Gamma}, \quad x = \frac{\varepsilon - \varepsilon_R}{\Gamma}, \quad f = \frac{3Qg}{4a\Gamma^2}. \quad (23)$$

The equation of type (22) is well known in the theory of nonlinear oscillation (see, for example, [8, 9]). In our case, the equation describes the dependence of the electron concentration n on the "pumping" current f and "detuning" current x. Expression (22) is a cubic equation,

$$n^3 - 2xn^2 + n(1 + x^2) - f = 0, \quad (24)$$

relative to n with the determinant

$$D = \frac{x^2(2 + x^2)}{27} + \frac{1}{27} + \frac{f^2}{4} - \frac{fx}{3}\left(1 + \frac{x^2}{9}\right). \quad (25)$$

For small values of f, which correspond to a small level shift, the concentration n is a symmetric function of x. As f increases, the curve $n(x)$ is deformed. When the value of f exceeds some critical value f_c, three solutions of $n(x)$ emerge (see figure), which correspond to one and the same value of x. Because the solution $n_2(x)$ increasing with x is unstable (see [8]), a hysteresis is observed. Indeed, as x increases, the concentration at $x > x_2$ decreases abruptly and is described by the third solution $n_3(x)$. When the increase of x changes to a decrease, the concentration increases abruptly to $n_1(x)$ when x becomes less than x_1.

First of all, we will find f_c from the condition of the confluence of minima of $x(n_1, n_2)$,

$$\frac{dx}{dn} = 3n^2 - 4xn + x^2 + 1 = 0. \quad (26)$$

The solution of this equation,

$$n_{1,2} = \frac{2x}{3} \pm \sqrt{\frac{x^2}{9} - \frac{1}{3}}, \quad (27)$$

gives two values of n at the extrema which merge on the condition that

$$x_c = \sqrt{3}, \quad n_c = 2/\sqrt{3}. \quad (28)$$

We substitute these quantities into Eq. (22) to find the critical value

$$f_c = 8/3\sqrt{3}. \quad (29)$$

This result agrees with the requirement of the emergence of three valid solutions of cubic Eq. (24). It is known to consist in that the determinant $D(x)$ (25) should be zero. One can readily check that, on substituting x_c and f_c into Eq. (25), we derive $D(x_c, f_c) = 0$. In addition, the derivative $\partial D/\partial x = 0$ also goes to zero, because D reaches a minimum at $x = x_c$.

Therefore, three solutions (i.e., three-valuedness of $n(x)$) and, consequently, hysteresis, emerge only when the condition

$$Q > Q_c, \quad Q_c = \frac{4a\Gamma^2}{3g} f_c = \frac{32a\Gamma^2}{9\sqrt{3}g} \quad (30)$$

is valid.

Otherwise, no hysteresis is present. This explains why hystereses are not always observed experimentally.

Another important conclusion is that the critical value of Q_c is proportional to the square of the level width Γ^2. This is associated with the fact that, the less the value of Γ, the smaller the value of concentration required for appreciable variation of resonance current. For example, in the model of [2], where $\Gamma = 0$, we have $Q_c = 0$.

The region of hysteresis, i.e., the values of n_2 and n_1 (x_2 and x_1) at extrema (see figure), will be found from the condition $dx/dn = 0$. In turn, the function $x(n)$ will be determined from Eq. (22) written as

$$x^2 - 2xn + n^2 + 1 - f/n = 0. \quad (31)$$

The solution of Eq. (31) has the form

$$x = n + \sqrt{f/n - 1}.$$

Hence, we take the derivative

$$\frac{dx}{dn} = 1 - \frac{f}{2n^2\sqrt{f/n - 1}} = 0$$

to derive

$$n^3 f - n^4 = \frac{f^2}{4}, \quad n \le f. \quad (32)$$

If we turn to more convenient variables

$$n = \frac{2N}{\sqrt{3}}, \quad f = \frac{8}{3\sqrt{3}}F, \quad F_c = 1, \tag{33}$$

we have, instead of (32),

$$4N^3F - 3N^4 = F^2, \quad N < \frac{4}{3}F. \tag{34}$$

We will find the solution of Eq. (34) in the limiting cases of $F \gg 1$ and $F \geq 1$. In the former case, we have

$$N_1 \approx \left(\frac{F}{4}\right)^{1/3}, \quad N_2 \approx \frac{4}{3}F,$$
$$x_1 \approx \frac{2}{\sqrt{3}}\left(\frac{F}{4}\right)^{1/3}, \quad x_2 \approx \frac{8F}{3\sqrt{3}}. \tag{35}$$

One can see that the region of hysteresis $x_2 - x_1$ increases proportionally with F; in so doing, x_2 increases much faster than x_1.

On turning to dimensional quantities, we derive

$$\varepsilon_2 - \varepsilon_R = \frac{3Q_g}{4a\Gamma},$$
$$\varepsilon_1 - \varepsilon_R = (Q\Gamma)^{1/3}\left(\frac{3g}{16a}\right)^{1/3}.$$

Therefore, as Γ increases, the value of ε_2 rapidly decreases and that of ε_1 slowly increases, so that the region of hysteresis decreases.

In the second limiting case of $F \geq 1$ (i.e., in the vicinity of the threshold), we have

$$N_{2,1} \approx F \pm \sqrt{\frac{F-1}{3}}, \tag{36}$$

$$x_{2,1} \approx \sqrt{3} + \frac{8(F-1)}{3\sqrt{3}} \pm \frac{16}{3}(F-1)^{3/2}. \tag{37}$$

One can see that the region of hysteresis

$$x_2 - x_1 = \frac{32}{3}(F-1)^{3/2} \tag{38}$$

slowly decreases with an increase in F.

At the same time, the extreme values of concentration diverge at a very high rate,

$$\frac{\partial}{\partial F}(N_2 - N_1) = \frac{1}{\sqrt{3(F-1)}} \longrightarrow \infty, \quad F \longrightarrow 1. \tag{39}$$

6. CONCLUSION

According to the foregoing results, the suggested model of coherent tunneling of interacting electrons predicts the emergence of singularities of the *IVC*, in particular, hysteresis, that are observed experimentally and agree qualitatively with the results obtained using the model of incoherent tunneling [3, 6].

At the same time, fundamentally novel results have been obtained. Firstly, these results include the prediction of the critical value of current through a resonance-tunnel diode, below which no hysteresis is present. Secondly, it is demonstrated that the value of critical current is defined by the width of resonance level (is proportional to Γ^2 according to Eq. (30)). Hence follows the important inference that a hysteresis is observed if certain conditions are valid (see (30)). This enables one to check the validity of the interpretations discussed above.

Note that, according to the simplified model of [3], in which it was assumed that $\Gamma = 0$, the critical current is zero, and a hysteresis must always be observed. In the more rigorous model of incoherent tunneling [6], the dependence of the hysteresis width on the level width Γ was indicated; however, no inference was made as to the existence of critical current. The *IVC* was calculated numerically for only two values of Γ.

The analytical solution of the set of Eqs. (11)–(15) enables one to fully analyze the dependence of the *IVC* on the parameters of resonance-tunnel diode and on the magnitude of current.

ACKNOWLEDGMENTS

I am grateful to Yu.V. Kopaev for useful discussions of this study.

The study was performed within the framework of the Program on the Physics of Solid-State Nanostructures of the Russian Ministry of Science and Technology (project no. 99-1140) and of the Integratsiya Russian Federal Program (project AO 133).

REFERENCES

1. L. Esaki and R. Tsu, Appl. Phys. Lett. **22**, 562 (1973).
2. V. J. Goldman, D. C. Tsui, and J. E. Gunningham, Phys. Rev. Lett. **58**, 1256 (1987); **59**, 1623 (1987).
3. F. W. Sheard and G. A. Toombs, Appl. Phys. Lett. **52**, 1228 (1988).
4. T. Sollner, Phys. Rev. Lett. **59**, 1622 (1987).
5. K. L. Jensen and F. A. Buot, Phys. Rev. Lett. **66**, 1078 (1991).
6. Jun Zang and J. L. Birman, Phys. Rev. B **46**, 5020 (1992).
7. V. F. Elesin, Zh. Éksp. Teor. Fiz. **116**, 704 (1999) [JETP **89**, 377 (1999)].
8. N. N. Bogolyubov and Yu. A. Mitropol'skii, *Asymptotic Methods in the Theory of Nonlinear Oscillations* (Fizmatgiz, Moscow, 1963; Gordon and Breach, New York, 1962).
9. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 1: *Mechanics* (Fizmatgiz, Moscow, 1958; Pergamon, New York, 1988).

Translated by H. Bronstein

Journal of Experimental and Theoretical Physics, Vol. 92, No. 4, 2001, pp. 715–743.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 4, 2001, pp. 822–852.

NONLINEAR PHYSICS

On Stochastic Mixing Caused by the Rayleigh–Taylor Instability

N. A. Inogamov[a], A. M. Oparin[b, *], A. Yu. Dem'yanov[c], L. N. Dembitskiĭ[c], and V. A. Khokhlov[a]

[a] *Landau Institute for Theoretical Physics, Russian Academy of Sciences, Chernogolovka, Moscow oblast, 142432 Russia*
[b] *Institute for Computer-Aided Design, Russian Academy of Sciences, Moscow, 123056 Russia*
[c] *Moscow Institute of Physics and Technology, Institutskiĭ per. 9, Dolgoprudnyĭ, Moscow oblast, 141700 Russia*
**e-mail: oparin@landau.ac.ru*

Received September 5, 2000

Abstract—The mixing of contacting substances is considered. The evolution of the mixing layer over a long time period from multimode initial perturbations is investigated numerically in the short-scale and wide-range cases. In the case of a short-scale initiation, the flow is stochastic in the sense that the time of the considered evolution exceeds the period of correlation. The effect of the amplitude of wide-range perturbations on the dynamics of mixing is analyzed. The scale-invariant properties of the spectral and statistical parameters of turbulent mixing are investigated for the first time. The universal spectra characterizing the turbulence mixing in the entire self-similar interval on a unified basis are obtained. The simulation is based on the effective algorithms with high approximating qualities, which have been tested earlier. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Rayleigh–Taylor instability is the term applied to the hydrostatic equilibrium instability for which the density in horizontal layers increases with height (heavy above light) [1–4]. The problem has found wide applications in astrophysics and the physics of high energy densities [1–4]. Suffice it to mention the experiments on mixing carried out and planned to be staged using the available high-power laser systems such as Iskra-5 (Russia), NIF (USA), LIL and LMJ (France), as well as devices being constructed. In view of its importance and complexity, this problem is considered by many authors (see the reviews in [1–4]). We will briefly describe here the prevailing ideas to emphasize new aspects of the present publication. The horizontal density jump between two pure substances is blurred into a turbulent transition region by the Rayleigh–Taylor instability. If the "microscopic scales" λ_{micr} (viscosity, diffusion, and capillarity) are small, while the external scales (compressibility and height of the homogeneous atmosphere, the size of a cuvette or the rated region) are large, the turbulent mixing becomes self-similar. The thickness of the mixed zone increases asymptotically according to the quadratic law:

$$h = \alpha \mathrm{At} g t^2, \quad h = h_+ + h_-,$$
$$\alpha = \alpha_+ + \alpha_-, \quad \mathrm{As} = \frac{h_-}{h_+} = \frac{\alpha_-}{\alpha_+}, \qquad (1)$$

where $\mathrm{At} = (1 - \mu)/(1 + \mu)$ is the Atwood number, $\mu = \rho_l/\rho_h < 1$ is the ratio of the densities of light and heavy liquids, g is the acceleration, h_+ and h_- are the penetration depths for the heavy and light liquids, respectively, and As is the asymmetry coefficient. The heavy and light liquids are supplied to the mixing zone through its upper and lower boundaries, respectively. The volume fraction of the heavy substance averaged over the zone is equal to 1/As.

2. PREVALENCE OF TRANSVERSE STRUCTURES

One of the trends in studying mixing is based on the theory of the turbulent mixing length and on the one-dimensional K and $K - \varepsilon$ models [5, 6]. The turbulent transport is approximated by a diffusion flow in the form of Fick's law. Phenomenological considerations are used to derive the expressions for the turbulent mixing length l_t and the turbulent velocities $w_t = \sqrt{(l_t \partial \ln \rho/\partial z) g l_t}$ appearing in the turbulent diffusion coefficient $D_t = l_t w_t/3$, where z, w, and x, y, u, and v are the vertical and horizontal (perpendicular to acceleration g) coordinates and velocities. In the one-dimensional approach, the instantaneous profile of density ρ depends only on coordinate z. It resembles the arctangent function with a bell-shaped profile of derivative ρ'_z typical of diffusion.

In the molecular-kinetic theory, an important characteristic is the Knudsen number[1] whose analogue in the mixing length theory is the ratio $\mathrm{Kn}_t = l_t/h_+$. The phenomenological coefficients are chosen from a comparison with experimental data. Length l_t is of the order

[1] The ratio of the mean free path to the geometrical scale.

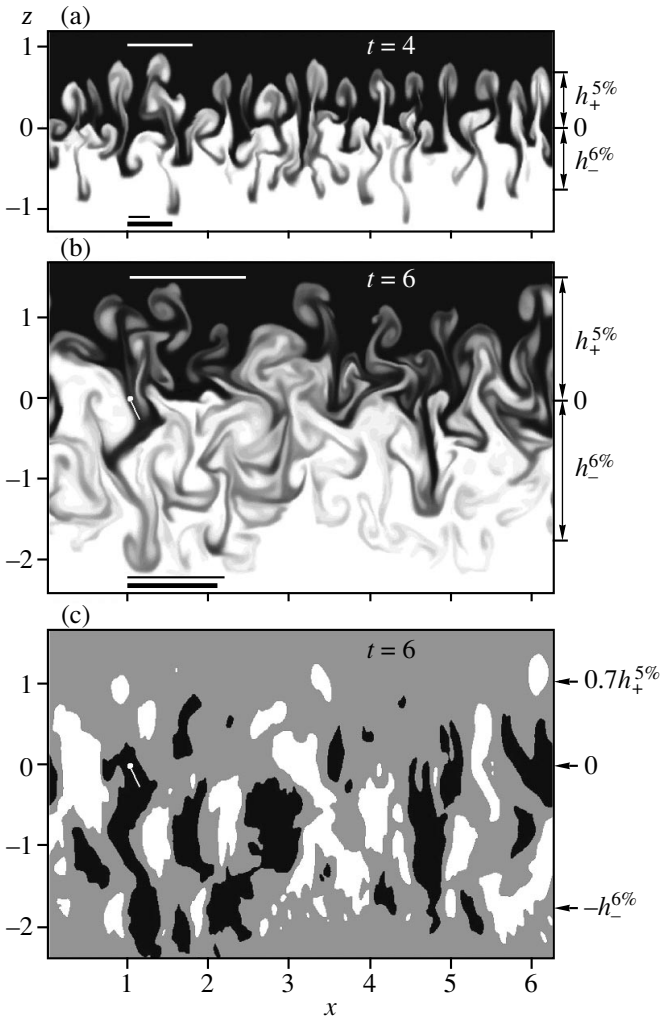


Fig. 1. (a, b) Evolution of a 2D flow (after a short-scale initiation) and the correspondence between (b) the instantaneous distribution $\rho(x, z, t)$ and (c) the velocity field: (a, b) Densities $\rho_h = 1$ (black) and $\rho_l = 0.1$ (white); the intensity of grey color (palette of 30 shades) indicates the value of density, $\rho_l \le \rho \le \rho_h$. Black segments of various thickness at the bottom indicate correlators K_w (thin segment) and K_H (thick segment) (see Fig. 4). (c) Tricolored field of the vertical velocity component $w(x, z, t = 6)$. Grey color corresponds to regions of rest or relatively slow motion ($-w_s < w < w_s$, $w_s = 0.8\dot{h}_+^{5\%}$). White color corresponds to the regions in which $w > w_s$, while black colors marks the regions with $w < -w_s$.

of the scale of typical structures in the mixing layer. An analysis of the experimental results shows that $\mathrm{Kn}_t \sim 1$, while the criterion for the applicability of the diffusion approximation is $\mathrm{Kn} \ll 1$. The rate of expansion of the mixing zone (or layer) is determined by transverse structures. Indeed, the motion is generated by the exchange or transposition of the heavy and light liquids. Such transpositions are impossible without the formation of transverse structures. It turns out that in order to ensure the required rate of the vertical expansion, the characteristic scale of the structures cannot be smaller than thickness h_+. The present work is devoted to the study of these important coarse or prevailing structures. In the self-similarity limit (1), the structures also become self-similar.

The geometry of the mixing layer is described in Section 3. The main definitions and the brief information on the numerical methods used are given in Sections 4, 5, and 6. In Sections 7 and 8, transient processes are considered. The problems of scale invariance are considered in Sections 9, 10, and 11. Self-similar substitutions are described in Sections 14, 15, and 16. The separation of the random component is discussed in Sections 17 and 18. Sections 19 to 21 are devoted to a comparison of theoretical estimates with the results of calculations. The dynamic estimates connected with the scales of transverse structures are given in Sections 22 to 24. A comparative analysis of two- and three-dimensional cases is carried out in Section 25. The spectra and statistics of random periodic functions simulating the dependence of fluctuating functions on the horizontal coordinate are considered in Section 26. The stimulation of mixing and deceleration by lateral boundaries are analyzed in Sections 27 and 28.

3. EXCHANGE IN COUNTER JETS. CASCADE EVOLUTION

Figures 1 and 2 show the results of typical calculations. The horizontal structure formed by alternating columns in which a heavy substance precipitates at the bottom and a light substance rises is clearly manifested. In order to compare the distributions of ρ and $\mathbf{v}\{u, v, w\}$ and to emphasize the correlation between ρ and w, the position of one of the jets of the sinking heavy substance ($x \approx 1.03$, $z \approx 0$) is marked in Fig. 1. The horizontal line $z = 0$ indicates the initial position of the interface. The zero-flow conditions are set at the upper and lower boundaries of the rated region. The effect of compressibility is insignificant: the Mach number $\dot{h}_+/c_h \sim 0.1$, where c_h is the velocity of sound in the heavy substance. The periodicity conditions are satisfied at the lateral boundaries.

Figure 1 emphasizes the existence of transverse variability (the horizontal chains of columns extended along the vertical). The finite acceleration $\ddot{h}_+$ of the mixing zone expansion is ensured by the increase in the column thickness with time. Let $\bar{\lambda}$ be the average thickness of a pair of adjacent columns,[2] $\bar{\lambda} \sim h_+$. The length of the vertical segments $h_+^{5\%}$ and $h_-^{6\%}$ in Figs. 1a and 1b and of white horizontal segment $h_+^{5\%}$ at the top of Figs. 1a and b were calculated using formula (1) with

[2] The averaging procedure and the straggling relative to the mean value will be considered below.

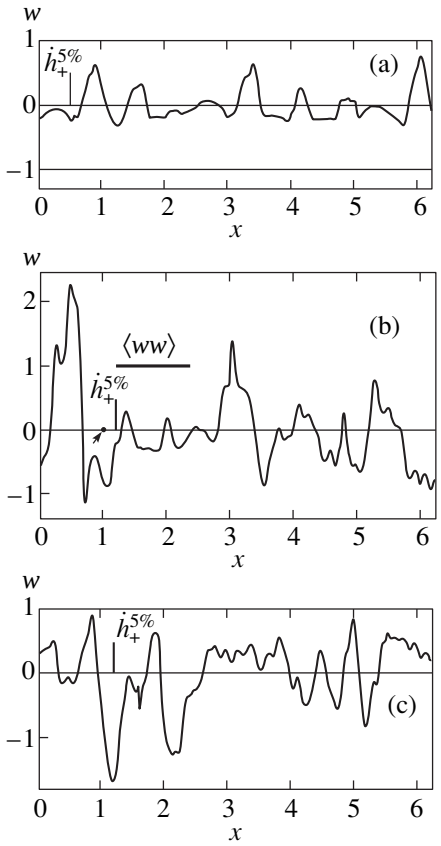


Fig. 2. Vertical velocity component $w(x, z, t = 6)$ in horizontal cross sections at levels z marked by arrows in Fig. 1c: (a) upper cut $z = 0.7h_+^{5\%}$; (b) middle cut $z = 0$, and (c) lower cut $z = -h_-^{6\%}$.

$\alpha_+ = 0.05$ and $\alpha_- = 0.06$. Figure 2 presenting the profiles of the vertical velocity component in various cross sections supplements Fig. 1c. In the upper part (the zone of penetration into the heavy substance), the velocities are low. Fast flows are observed in the lower part of the mixing zone. This is clearly seen in Fig. 1c. Note the asymmetry of the jets of the heavy and light liquids at the level $z = 0$. Ascending jets are narrower and faster, $\mu = 1/10$. This asymmetry decreases as the values of ρ_l and ρ_h become closer. In calculations with $\mu = 1/3$, distribution $w(x)$ is almost invariant to the inversion relative to the vertical direction ($w \longrightarrow -w$).

The most interesting feature of the evolution is the increase in the size of structures with time. It is manifested in the increase of the transverse scale $\bar{\lambda}$ with time t. In our calculations, we put $g = 1$ and $L_\perp = 2\pi$, where $L_\perp$ is the transverse size of the computational "box". This determines the units of time measurement. The hierarchy of successive enlargements is called the inverse cascade [3, 4]. This term is adopted to emphasize the difference from the Kolmogorov cascade of fragmentations towards high frequencies.

4. SPECTRA AND CORRELATORS: CORRELATION LENGTH

Turbulent mixing is on the average invariant to translations in the horizontal direction. Consequently, it would be interesting to construct the spectra of variables along the spatial coordinate x. The "bumps" on these spectra indicate the presence of well-developed transverse structures. Figure 3 shows the Fourier spectra of H, ρ and w at instant $t = 6$ to which Figs. 1c and 2 correspond. Here, $H(x, t) = [\int \rho(x, z, t)dz]/(\rho_h - \rho_l)$ is the mass contained in a vertical column (the integral is taken from the "bottom," $z = -h_{\text{down}}$, to the "top," $z = h_{\text{up}}$, of the box). The spectrum $w_n(z = 0, t = 6)$ was calculated from the function $w(x, z = 0, t = 6)$, whose graph is presented in Fig. 2b:

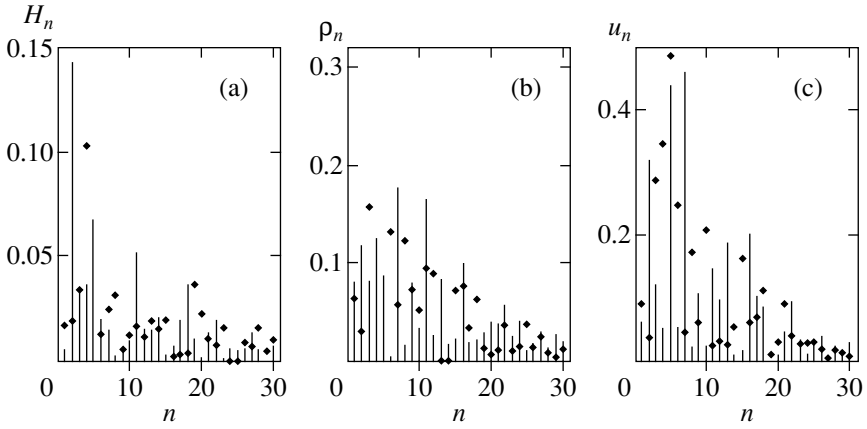


Fig. 3. Spectra of H, ρ, and u in the middle plane $z = 0$ in which the interface was at $t = 0$: (a) the amplitudes $H_n(t = 6)$ of the Fourier expansion of the mass $H(x, t)$ of the column characterizing the density fluctuations associated with the columns of heavy (large values of H) and light (small H) liquids; (b) density spectrum $\rho(x, z = 0, t = 6)$, and (c) vertical velocity spectrum $w(x, z = 0, t = 6)$. The spectrum formed by vertical segments corresponds to calculations with the number N_x of nodes along the horizontal equal to 400, while the spectrum of markers corresponds to $N_x = 600$.

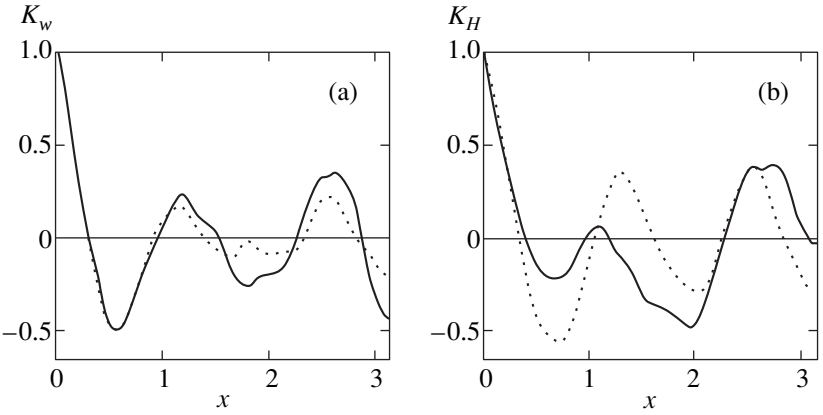


Fig. 4. Graphs of the vertical velocity correlators $K_w(x, z = 0, t = 6)$ and column mass correlators $K_H(x, t = 6)$. The solid curves correspond to $N_x = 400$ and the dashed curves to $N_x = 600$. In order to estimate the correspondence of scales, the correlation lengths are shown in Figs. 1a and b in the form of horizontal segments.

$$w_n = \sqrt{a_n^2 + b_n^2}, \quad a_n = \pi^{-1}\int_0^{2\pi} wc_{nx}dx,$$

$$b_n = \pi^{-1}\int_0^{2\pi} ws_{nx}dx,$$

$$w(x, \ldots) = \sum (a_n c_{nx} + b_n s_{nx}), \quad c_{nx} = \cos nx,$$

$$s_{nx} = \sin nx.$$

Figure 4 shows the graphs of correlators (correlation between columns)

$$K_w(x, z, t) = \langle ww \rangle$$

$$= \frac{\int_0^{2\pi} w(x', z, t)w(x' - x, z, t)dx'}{\int_0^{2\pi} [w(x', z, t)]^2 dx'}, \qquad (2)$$

$$K_H(x, t) = \langle HH \rangle$$

for $z = 0$, $t = 6$. Correlator (2) is an even function, $K(x) = K(-x)$, with a period $L_\perp = 2\pi$. For this reason, it is meaningless to plot its graph on an interval larger than $0 < x < L_\perp/2$. The properties of parity and periodicity of function $K(x)$ on a "long" segment $L_\perp$ are "imposed" by the periodic boundary conditions on the lateral boundaries of the rated region. Segment $L_\perp$ is referred to as long here since we are interested in the correlator characteristics over one or two correlation lengths l_{corr}, which will be defined below as long as the correlation length is small as compared to the rated segment $L_\perp$. At a later stage, complete correlation across the rated region sets in, $l_{corr} \sim L_\perp$. In this case, the flow is strongly "restricted" by lateral boundaries. At a late stage, the mixing along the vertical is correlated on a scale of the order of $L_\perp$. An interesting example of a late state of mixing in a box extended along the vertical and containing a zigzag structure is considered in [7].

Functions ρ and w are well correlated along the vertical (correlation along the column). This is clear from Fig. 1 and from a comparison of Figs. 2a, 2b, and 2c, which are cuts at various horizontal lines marked by the arrows in Fig. 1c. At a stage far from the "restricted" flow, the height of the columns is smaller than the transverse scale $L_\perp$.

We will define the correlation length (λ_K) as the distance between the center $x = 0$ and the second peak (the first peak is at the center). This length is associated with one of the definitions of the average size of a transverse structure. With such a definition, the average size is characterized by the total thickness of two adjacent columns along the horizontal (the substance precipitates in one column and rises in the other column).[3] It is important to note that in the case of a strictly periodic alternation of columns, this length is equal to the period of the flow. Thus, the above definition is a natural generalization to the randomly periodic case, when the exact period cannot be defined, but a typical scale exists (the description of random periodic functions can be found in [3, 4]). The choice of the scale from the total width of adjacent columns appears as more justified than its determination based on the half-width of the first correlator peak with the maximum at the center $x = 0$. As a matter of fact, with such a definition the scale has a correct asymptotic form in the limit of periodic functions.

The lengths λ_K determined from functions K_w and K_H at instants $t = 4$ and 6 are presented respectively in the form of thin and bold horizontal segments at the bottom of

[3] We are speaking of coupled columns and of direct and backward flows. In an incompressible liquid, the backward flow compensates the sinking volume.

Figs. 1a and 1b. This is done to ensure the cross comparison of the sizes determined in different ways.

5. VARIATIONS OF THE "BOX" HEIGHT, DENSITY RATIO, AND MESH STEP

The Fourier transforms (Fig. 3) and correlators (2) (Fig. 4) provide information on the statistics of transverse structures. Before going over to the spectral statistics, we will briefly describe the numerical procedure. In our series of computations, the amplitude and the horizontal scale of initial noise were varied. Thus, we investigated the transition from the initial to the self-similar stage. The calculations were made in the 2D and 3D geometries. The results presented below were obtained in the 2*d* case for which rich information has been accumulated. In our computations, the step of the computational mesh was varied (the number of nodes N_x on a segment of the x axis varied from 100 to 600). The box height $h_{\text{box}} = h_{\text{down}} + h_{\text{up}}$ was also varied (the versions with $h_{\text{box}}/L_\perp$ varying from 2 to 14 were used; the ratio $h_{\text{down}}/h_{\text{up}}$ was usually 3:2 or 5:4). The ratio of densities was varied as follows: the versions with a large ($\text{At} \approx 1$, $\mu = 1/10$), intermediate ($\text{At} < 1$, $\mu = 1/3$) and small contrast ρ ($\text{At} \approx 0$, $\mu = 1/1.2$, Boussinesq's case) were used. The computations were made using two numerical codes. In one of these codes, the Belotserkovskii–Davydov method of large particles [8–10] was used, while in the other code, use was made of the quasi-monotonic grid-characteristic scheme with the second-order approximation [8–10]. The grid-characteristic method was verified by comparing with two- and three-dimensional analytic solutions of the one-mode problem [11] and by comparing with two- and three-dimensional test problems proposed at the Marseille [12] and previous Workshops on Physics of Compressible Turbulent Mixing.

6. SHORT-SCALE PERTURBATIONS AND INERTIAL GAP

Figures 1–4 illustrate the evolution of mixing from short-scale initial perturbations after the start. At $t = 0$, the surface velocity field of the form

$$\mathbf{v} = -\nabla\varphi, \quad \varphi = \sum \varphi_n,$$
$$\varphi_n = \operatorname{sgn} z (a_n^0 \cos nx + b_n^0 \sin nx) e^{-n|z|}/n \tag{3}$$

was specified. At the initial instant, the boundary of the contact coincides with the horizontal $z = 0$. The short-scale modes were characterized by the wave numbers n ($k_n = 2\pi/\lambda_n = n$ for $L_\perp = 2\pi$) in the interval $k_{sw1} < n < k_{sw3}$ with the edges $30 < n < 120$ or $50 < n < 100$. The amplitudes a_n^0 and b_n^0 appearing in formula (3) were cast by a random number generator (superscript 0 on the amplitudes indicates that $t = 0$). The modulo maximum value of the average amplitude was at point k_{sw2} with $k_{sw2} = k_{sw1}$ or with k_{sw2} from the interval between k_{sw1} and k_{sw3}.

It is important that there is a gap free of initial perturbations between mode k_1 with $n = 1$ and with wavelength λ_1, which is equal to the "box" width $L_\perp$, and the short-scale range. Its logarithmic width in doublings was $N_{\text{doubl}} = \ln_2(k_{sw1}/k_1) = 5$–$6$. This width defines the number of generations of pair "confluences" of columns or the number of scale doublings [3, 4] and, accordingly, the width of the inertial interval in which the quadratic dependences (1) are valid. An inverse cascade in this inertial interval follows the scale-invariant mode.

7. TRANSIENT PROCESSES AT LOW AND HIGH FREQUENCIES AT THE EDGES OF THE SELF-SIMILAR RANGE

The stabilization of the self-similar mode for $k \sim k_{sw}$ and its termination for $k \sim k_1$ occur gradually. The departure from the quadratic self-similarity for $k \sim k_1$ is associated with the "restriction" of the flow by the lateral boundary conditions. The restriction of mixing leads to depletion of the spectrum and to a decrease in the number of dynamically significant transverse modes in the distribution of amplitudes over harmonics $f_n(z, t)$, where f_n is the Fourier transform of one of the functions $f(x, z, t) = (\rho, u, w, p)$ being calculated.

The lateral boundaries of the rated region or experimental cuvette generate a number of dynamic effects. We will confine our analysis to the case when the upper and lower boundary conditions are immaterial (the cuvette is extended along the vertical). In connection with lateral boundaries, we will speak of the "bottleneck," its early effect on the flow approaching it, and the passage through the bottleneck. Let us specify the approach of the bottleneck (we are speaking of the approach in the space of wave numbers). In the mode language, the mode with k_1 ($n = 1$) corresponds to the bottleneck. The instantaneous fields of variables in the mixing zone can be described by spectra f_n in the horizontal wave modes k_n. These spectra turn out to be concentrated (see Fig. 3). Let $\bar{k}_n$ (or $\bar{n}$) be the inverse scale near which the spectrum is localized; i.e., the amplitudes f_n which are significant at the given instant are concentrated (a "bump" in the spectrum). The value of $\bar{k}_n$ decreases with time (inverse cascade). Accordingly, the spectrum approaches the bottleneck k_1.

Owing to transient processes for $k \sim k_{sw}$ and $k \sim k_1$, the real width of the self-similar interval $N_{\text{doubl}}^{\text{real}} = N_{\text{doubl}} - (\Delta N_{sw} + \Delta N_1)$ is smaller than N_{doubl}. The quantities ΔN_{sw} and ΔN_1 characterize the losses associated with short- and large-scale limitations. The increase in the quantity $N_{\text{doubl}}^{\text{real}}$ required for improving the accuracy of the analysis can be achieved, first, by the enlargement

of computer resources and, second, by analyzing transient processes on small and large cutoff scales.

8. "FORGETFULNESS" OF SHORT-SCALE PHASES (ESTABLISHMENT OF SELF-SIMILARITY): DELAYS DUE TO SCHEMA EFFECTS

Let us consider a transient process for large wave numbers $k \sim k_{sw}$. The number of meshes in the grid over wavelength λ_{sw1}/Δ varied from 4 to 15, where $\lambda_{sw} = 2\pi/k_{sw}$ and Δ is the mesh spacing. For example, in the calculations depicted in Fig. 1, we had $\lambda_{sw1}/\Delta \approx 8$ (400×1000 mesh). The establishment of self-similar mode after short-scale initiation occurs as a result of a complex process involving the enhancement of the initial perturbations to the level of nonlinear saturation and then the enhancement of the combination harmonics forming the "infrared" wing in the distribution of amplitudes f_n.

In the case when the initial amplitudes are linear, the rough estimation of the duration t_{trans} of the transient process gives

$$t_{\text{trans}} \approx (\gamma_{sw}^{\text{code}})^{-1} \ln(w_{sw}^{\text{nonl}}/\bar{w}_{sw}^0). \quad (4)$$

Here $w_{sw}^{\text{nonl}} = 0.6\sqrt{(1-\mu)g/k_{sw}}$ is the limiting velocity of the ascent of 2D bubbles in the one-mode regime, which is taken for estimation [11], $\gamma_{sw}^{\text{code}} = \beta(\lambda_{sw}/\Delta)\gamma_{sw}$, $\gamma_{sw} = \sqrt{\text{At}gk_{sw}}$ is the increment of the Rayleigh–Taylor instability, and $\beta < 1$ is the correction factor taking into account the stabilizing effect of mesh discretization. Its value depends on the number of nodes per wavelength. For $\lambda/\Delta \sim 1$, the instability is strongly stabilized, $\beta \sim 1$. For $\lambda_{sw1}/\Delta = 3$ to 6, the classical increment decreases by a factor of 1.5–3.

The average initial velocity of short-wave perturbations is defined as

$$\bar{w}_{sw}^0 = \sqrt{(2\pi)^{-1}\int_0^{2\pi}[w(x, z=0, t=0)]^2 dx}.$$

If the initial spectrum contains N modes, we have $\bar{w}_{sw}^0 \sim \sqrt{N}\bar{c}_n^0$, where $\bar{c}_n^0 = \sqrt{\overline{(a_n^0)^2} + \overline{(b_n^0)^2}}$, $\sqrt{\overline{(a_n^0)^2}} = \sqrt{\overline{(b_n^0)^2}}$ is the root-mean-square amplitude of the initial short-wave noise, and $N = k_{sw3} - k_{sw1}$ for $L_\perp = 2\pi$. The significant factor $\sqrt{N}$ connecting the quantities $\bar{w}^0$ and $\bar{c}^0$ appears as a result of summation of N random modes. In the version depicted in Fig. 1, the initial perturbation $\bar{w}_{sw}^0$ amounts approximately to 2% of the nonlinear velocity w_{sw}^{nonl}.

For moderate values of number λ_{sw}/Δ, the effective value of the initial velocity amplitude decreases:

$$\bar{w}_{sw}^0 \longrightarrow w_{\text{eff}} = \beta'\bar{w}_{sw}^0,$$
$$\beta' = \beta'(\lambda_{sw}/\Delta) < 1,$$

which determines the evolution of instability in accordance with formula (4). This is associated with the 2π effect and discretization. The distance h_e along the vertical over which velocity (3) decreases by a factor of e amounts to $\lambda/2\pi$. Consequently, for $\lambda_{sw}/\Delta = 3$ to 6, we obtain $h_e = (0.5\text{–}1)\Delta$. This means that the surface layer of a thickness approximately equal to h_e, in which the velocity field is concentrated at $t = 0$, is very thin as compared to the mesh spacing Δ. Under these conditions, the effect of mesh enlargement is significant. The action of β and β' factors prolongs the transient process. Thus, the duration t_{trans} (4) of the transition is determined by number λ_{sw}/Δ. For a fixed value of k_{sw}, the process duration t_{trans} on coarse meshes is larger. In the grid-characteristic method, $t_{\text{trans}} \sim 1$ ($\gamma_{sw1} = 1/6.4$) for $N_x = 400$ (the number of nodes along the horizontal) and $k_{sw1} = 50$ (the edge of short-scale range). It should be noted that for identical parameters, the transition rate in the method of large particles is higher than in the grid-characteristic method.

9. SIMILARITY AND "STATIONARITY" (OR UNIVERSALITY)

In a transient process, the phase information contained in high-frequency modes is gradually lost. The process terminates when a self-similar flow is formed, which is similar, or universal (and "stationary" in this sense) in the corresponding self-similar variables. The advantage of self-similarity is exactly this stationarity or universality and, hence, the relative simplicity of mixing.

10. SELF-SIMILARITY OF VERTICAL EXPANSION

In the given case, self-similarity is of the power (quadratic) type. The stationary (self-similar) total thickness of the mixing layer is $H_{ss} = h/gt^2$, the subscript "ss" indicating the affinity to the self-similar mode. The stationary velocity and acceleration of the expansion of the layer as a whole are defined as $V_{ss} = \dot{h}/gt$ and $G_{ss} = \ddot{h}/g$. The thickness, velocity, and acceleration of the upward expansion of the mixed layer of a heavy liquid are specified by the time-independent constants H_{ss}^+, V_{ss}^+, and G_{ss}^+.

These quantities are determined by two universal functions of the density ratio μ. We can write

$$H_{ss}^+ = \alpha_+\frac{1-\mu}{1+\mu} = \alpha_+\text{At}, \quad V_{ss}^+ = G_{ss}^+ = 2H_{ss}^+. \quad (5)$$

The factor $1 - \mu$ in the numerator of relation (5) is exact. It follows from the Archimedes law, $g \longrightarrow (1 - \mu)g$. The factor $1 + \mu$ in the denominator is associated with the inertia of the light liquid in the mixture of light and heavy liquids. It follows from natural and numerical experiments that the coefficient α_+ in relations (1) and (5) remains approximately unchanged upon the variation of ratio μ. The accuracy of modern experiments does not allow us to refine the weak dependence of α_+ on μ.

Since we have chosen two geometrical characteristics h_+ and h_- (1) for describing the vertical structure, there also exist two universal functions associated with the distribution along the vertical. The second is the function $H_{ss}^-(\mu)$. Usually, the dependence of the asymmetry coefficient on μ, $h_-/h_+ = \mathrm{As}(\mu)$ is investigated instead ($H_{ss}^-(\mu) = \alpha_+ \mathrm{At}\,\mathrm{As}$) [3, 4].

11. SELF-SIMILARITY OF AN INVERSE CASCADE

A cascade is a sequence of logarithmic links (or generations) with a hierarchical structure resembling a geometrical progression. We fix a certain instant t_f. Let $\bar{k}_f = 2\pi/\bar{\lambda}_f$ be the characteristic wave number for horizontal structures at this instant. The time evolution for $t > t_f$ involves the saturation (restriction, termination of growth, and suppression) of higher harmonics with $k > \bar{k}_f$ and simultaneously the enhancement of subharmonics with $k < \bar{k}_f$ [3, 4]. This leads to a decrease in $\bar{k}$. A link in the cascade is regarded as passed when the inverse scale $\bar{k}$ decreases by a factor of q, where $q \sim 1$ is the common ratio (step) of a geometric progression. Usually, we are dealing with doubling and assume that $q = 2$, which is justified to a certain extent [3, 4]. In view of the quadratic form of self-similarity (1), the progression step in time amounts to $t_{n+1}/t_n = \sqrt{q}$.

The scale invariance of a cascade indicates the similarity of generations. If we compare the instantaneous fields of variables $(\rho_1, u_1, v_1, w_1, p_1)$ and $(\rho_2, u_2, \ldots)$ at instants t_1 and t_2, they are on the average identical (to within the phase) after the similitude transformation:

$$\langle\{\rho_1(x, y, z), u_1(x, y, z), \ldots, p_1(\ldots)\}\rangle = \left\langle\left\{\rho_2\left[\left(\frac{t_2}{t_1}\right)^2 x, \left(\frac{t_2}{t_1}\right)^2 y, \left(\frac{t_2}{t_1}\right)^2 z\right], \frac{t_1}{t_2} u_2\left[\left(\frac{t_2}{t_1}\right)^2 x, \ldots\right], \ldots, \left(\frac{t_1}{t_2}\right)^2 p_2[\ldots]\right\}\right\rangle. \quad (6)$$

Examples of the compression transformations (6) are presented in Figs. 5–8. Disregarding the phase information, we see that the approximate similitude of density fields (Figs. 5–7) and velocity fields (Fig. 8) takes place. A certain residual effect of the mesh is worth noting. Self-similarly scaled spatial sizes of the columns along the vertical and horizontal are approximately identical; cf. the upper ($t = 4$) and lower ($t = 6$) fields in Fig. 5. In view of the mesh effect, the fine substructure in prevailing columns for $t = 6$ is sharper for $t = 6$; cf. the 400×1000 mesh (Fig. 5) and the 600×1500 mesh (Fig. 6). Figure 7 corresponds to the density ratio $\mu = 1/3$; in Figs. 5 and 6, we have $\mu = 1/10$.

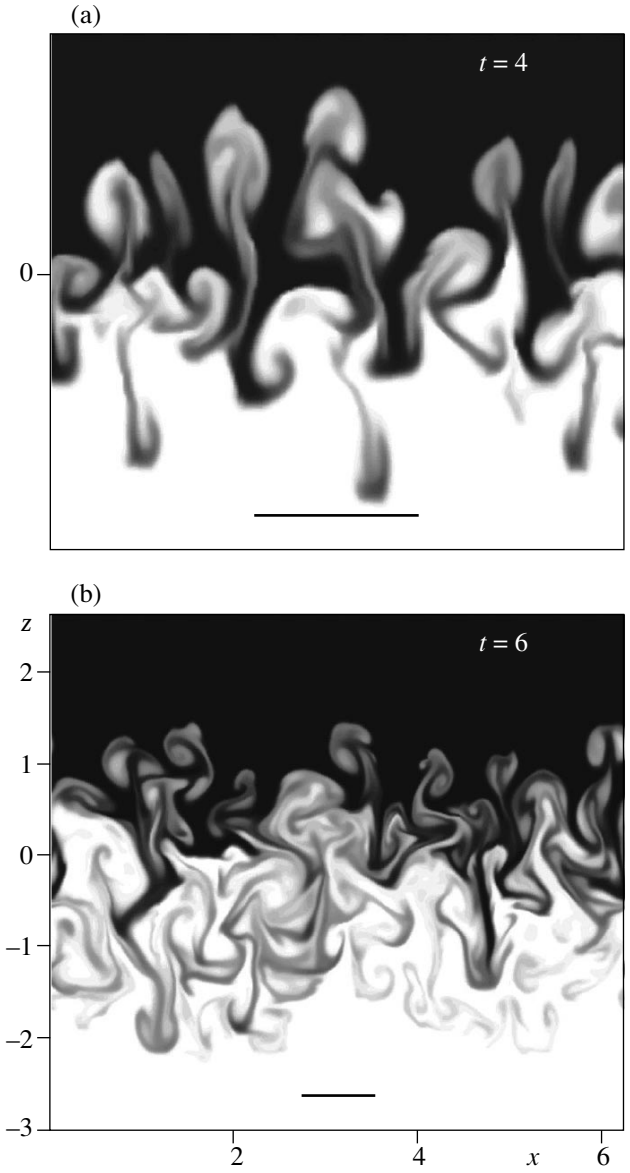


Fig. 5. Comparison of $\rho(x, z, t)$ fields at instants $t = 4$ and 6. Coordinates x and z are scaled in accordance with relations (6) (cf. Figs. 1a and 1b). The mesh size is 400×1000, $\mu = 1/10$; the grid-characteristic method is used. Horizontal segments indicate the size of 50 mesh spacings.

The numerical verification of the fulfillment of the similitude criteria carried out by us here proves that

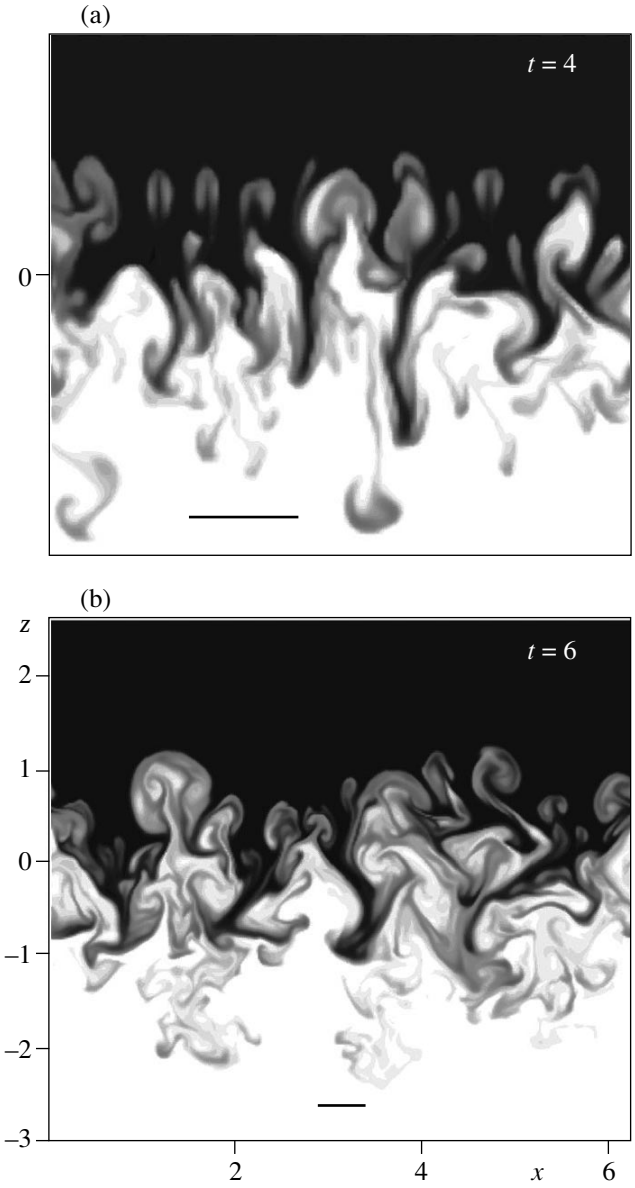


Fig. 6. The same as in Fig. 5 for a finer mesh of 600×1500.

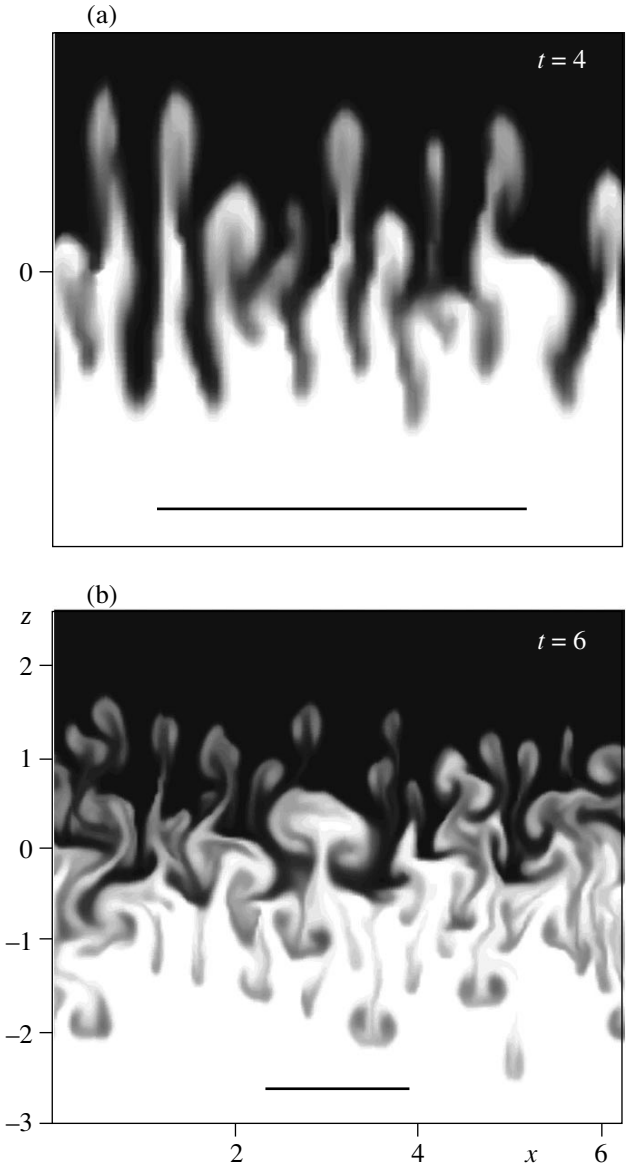


Fig. 7. The same as in Figs. 5 and 6 for a different density ratio 1/3; mesh size 200×500.

these criteria are approximately satisfied outside the initial (self-similarity stabilization) and final (effect of bottleneck, see Section 7) transient stages.

12. SPECTRA EVOLUTION: INITIAL REGION AND THE MEASURE OF NONLINEARITY OF THE INITIAL PERTURBATION

In the previous section, we considered the evolution in conventional coordinates (coordinate representation, coordinate space) not only along the vertical, as in Section 10, but along the vertical and horizontal. Let us now analyze the time evolution in the space of wave numbers (momentum representation, momentum space). Figure 9 shows a cinegram illustrating the gradual rearrangement of the spectrum. The spectrum has the shape of a bump localized near the characteristic inverse scale $\bar{k} = 2\pi/\bar{\lambda}$. In the case of small (and, hence, linear) initial perturbations, the amplitude of the initial modes are the first to increase. The position of the bump on the n axis remains unchanged in this case (the numbers of significant modes do not change) and only its "height" increases. As the certain value of the bump height (saturation height) is attained, its further increase slows down. A further increase in the height is accompanied by a displacement of the bump to the left along the n axis.

Let us suppose that saturation takes place for $t \approx t_{sat}$. For $t > 0$, the bump acquires an infrared, or subhar-

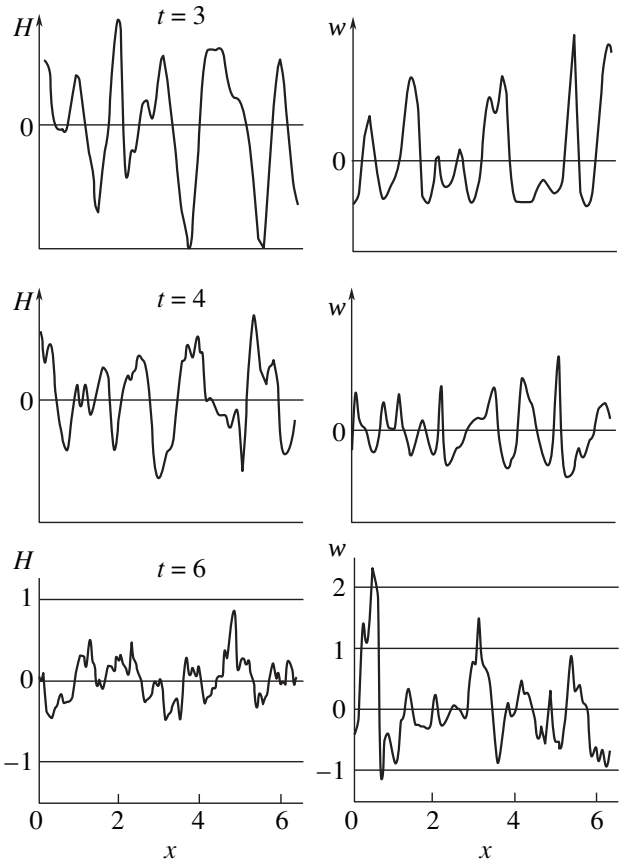


Fig. 8. Horizontal profiles of H and w at the middle section $z = 0$ at instants $t = 3$, 4, and 6. Profiles are scaled in accordance with (6).

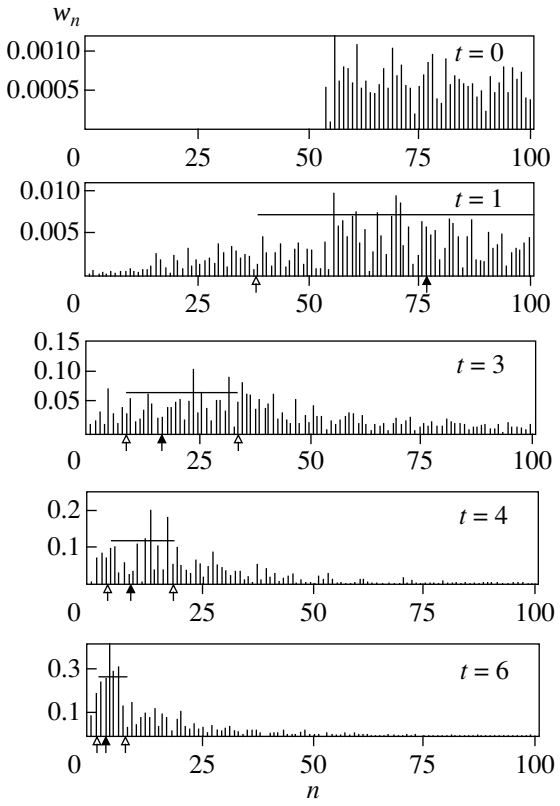


Fig. 9. Evolution of the vertical velocity spectrum. Amplitudes $w_n = \sqrt{a_n^2 + b_n^2}$ are shown. The wave number scale is the same. Dark arrows mark the characteristic inverse scale, while light arrows correspond to twice and half as large scales. The horizontal segments give the saturation amplitude.

monic wing on the left side. As long as $t \ll t_{sat}$, its amplitude is smaller than the amplitude of modes with $k \sim k_{sw}$. For $t \sim t_{sat}$, the amplitudes of the wing and of the bump itself become equal. This leads to a displacement of the bump to the left.

In the case of saturated initial amplitudes, the displacement to the left begins at the very outset. This corresponds to moderate (average) initial perturbations or amplitudes which are nonlinear from the very outset. The magnitude of the saturation amplitude depends on the shape of the bump and its width along n. Let us find out why the width of the spectral region in which the initial perturbation is concentrated affects the nonlinear saturation amplitude. We consider a rectangular bump. Such a bump is composed of modes having approximately identical amplitudes and filling the spectral region $k_{sw1} < k_n < k_{sw3}$, where $N = k_{sw3} - k_{sw1}$ is the number of modes in the initial bump. We assume that the perturbation is concentrated in the spectral region whose range is not logarithmically wide: $k_{sw3} \sim k_{sw1}$ rather than $\ln(k_{sw3}/k_{sw1}) \gg 1$. The phases change at random from mode to mode.

In the case of a single mode ($N = 1$), we take the elevating speed of periodic bubbles,

$$w_{sat} = w_b = F\sqrt{(1-\mu)g/k_{sw}}, \quad k_{sw} = k_{sw1} = k_{sw3}, \quad F \approx 0.6, \tag{7}$$

for estimating the saturation amplitude from velocity (see [11] and the literature cited therein), where F^2 is the Froude number. For the single mode, the elevating speed (7) coincides in order of magnitude with the amplitude of the harmonic [3, 4] (for $t \sim t_{sat}$).

In the case of many modes ($N \gg 1$), the root-mean-square velocity (over the space)

$$\bar{w} = \sqrt{\frac{1}{2\pi}\int_0^{2\pi} [w(x, z = 0, t)]^2 dx} \tag{8}$$

differs significantly from the bump-averaged amplitude $\bar{w}_n$ [13, 14, 3, 4]. For the sake of definiteness, the middle cut $z = 0$ is taken in (8). In view of the summation of a large number of harmonics, the velocity $\bar{w}$ averaged over x is clearly larger than the amplitude-averaged velocity $\bar{w}_n$. If the harmonics are phased, we have $\bar{w} \sim N\bar{w}_n$ due to interference. It is well known [13, 14, 3, 4] that the sum of harmonics with random phases

increases in proportion to the square root of the number of addends:

$$\tilde{w} \sim \sqrt{N}\bar{w}_n. \quad (9)$$

Indeed, it is known that

$$\left|\sum_{n=1}^{N} \exp(i\psi_n)\right| \sim \sqrt{N},$$

where ψ_n is the random phase on the segment $[0, 2\pi]$ (the problem on random walks or on diffuse spreading). Accordingly, a rough estimate of the saturation amplitude has the form

$$\bar{w}_{\text{sat}} \sim w_b, \quad \bar{w}_n \sim w_b/\sqrt{N}.$$

This expression follows from formula (9). The value of $\bar{w}_{\text{sat}}$ is calculated by averaging in accordance with formula (8), and the quantity w_b is defined by formula (7).

Apart from the initial perturbations with small and moderate amplitudes considered by us, large initial velocities $\bar{w}_0$ are also possible. If the value of $\bar{w}_0$ is much larger than w_{sat}, the effect of gravitation in the initial region is small. The motion occurs in the mixing mode associated with the Richtmyer–Meshkov instability [15–20, 3, 4]. In this mixing mode, the displacement in the wave numbers n is small. The duration of the initial interval is of the order of $1/k_{sw}\bar{w}_0$. During this time interval, the rates of the Richtmyer–Meshkov mixing decrease to values of the order of w_{sat}. After this, the Rayleigh–Taylor evolution begins.

13. EVOLUTION ON THE WHOLE: BOUNDARIES OF THE SELF-SIMILAR REGION

We single out the self-similar region from the evolution on the whole. The latter includes the initial, self-similar, and final regions (three regions in all) presented in Fig. 9. For the chosen values of mesh density (N_x = 200, 400, 600, $L_\perp = 2\pi$), wavelength, and perturbation amplitude ($k_{swl} \approx 50$, the perturbation is linear), the formation of the self-similar stage begins at $t_{\text{begin}} \approx 2.5/\sqrt{\text{At}}$. A noticeable deceleration of mixing in the bottleneck (see Section 7) begins at $t_{\text{end}} \approx (6 \text{ to } 7)/\sqrt{\text{At}}$. The value of t_{end} weakly depends on small-scale characteristics (mesh spacing and the parameters of short-scale perturbation).

14. SELF-SIMILAR SUBSTITUTIONS IN THE COORDINATE REPRESENTATION

We carry out the formal substitution

$$(x, z, t) \longrightarrow (\hat{x}, \hat{z}, t), \quad f(x, z, t) \longrightarrow \hat{f}(\hat{x}, \hat{z}, t),$$

$$f = (\rho, u, w, p), \quad \hat{x} = \frac{x}{gt^2}, \quad \hat{z} = \frac{z}{gt^2},$$

$$\rho(x, z, t) = \rho_h\hat{\rho}(\hat{x}, \hat{z}, t), \quad (10)$$

$$u(x, z, t) = gt\hat{u}(\hat{x}, \hat{z}, t), \quad w(x, z, t) = gt\hat{w}(\hat{x}, \hat{z}, t),$$

$$p(x, z, t) = \rho_h g^2 t^2 \hat{p}(\hat{x}, \hat{z}, t).$$

The sets of variables (x, z, t, f) and $(\hat{x}, \hat{z}, t, \hat{f})$ are equivalent. Both sets are suitable for describing a flow of the general (i.e., not necessarily self-similar) type. The self-similar coordinates $\hat{x}$ and $\hat{z}$ together with time $\hat{t}$ are required to study the establishment and breakdown of self-similarity. These processes occur in the initial and final regions, respectively (see Section 13).

Let us consider the evolution of the mixing zone in the self-similar variables. From the three regions, the middle is self-similar (see Section 13). In this region, there is no motion in the self-similar variables (on the average, to within fluctuating phases). This is the essence of "stationarity" (see Sections 9 and 10). In variables $(\hat{x}, \hat{z}, t, \hat{f})$, the motion (i.e., the dependence on t, or nonstationarity) is observed on the initial segment. In the self-similar region, the motion is "frozen." For $t_{\text{begin}} < t < t_{\text{end}}$, time t is omitted from the set of arguments $(\hat{x}, \hat{z}, t)$, and variables f become functions of the self-similar "coordinates" $(\hat{x}, \hat{z})$ alone. For $t > t_{\text{end}}$, the motion (nonstationarity) is resumed as we approach the bottleneck (see Section 7).

In the self-similar region in the physical variables (x, z, t, f), the coarsening of the structures $\bar{\lambda} \propto t^2$ and an increase in the amplitude of fluctuations of u, w, and p are observed (the amplitude of fluctuations of ρ remains unchanged). Substitutions (10) just establish the fact that the evolution is reduced to the scaling transformation (scaling). Let us consider, for example, a random periodic function corresponding to the vertical component of velocity $w(x, z, t)$. The step and amplitude of this function increase in proportion to t^2 and t, respectively. This is illustrated in Fig. 8.

It should be noted that the term random periodic function [3, 4] (or distribution) is more appropriate than the term quasiperiodic function for studying the instantaneous distributions fluctuating during a displacement along the horizontal we are dealing with. It correctly characterizes the emerging structures. The term quasiperiodic function is applied to a finite sum of periodic functions [21]. Its Fourier transform and correlator differ from those of a random periodic function.

15. SELF-SIMILAR SUBSTITUTIONS IN THE MOMENTUM REPRESENTATION

The coordinate and momentum representations are connected through the direct and inverse Fourier transformations:

$$f(k) = \int_{-\infty}^{\infty} \frac{dx}{2\pi} e^{ikx} f(x), \quad f(k) = a(k) + ib(k), \tag{11}$$

$$|f(k)| = \sqrt{a^2 + b^2},$$

$$f(x) = \mathrm{Re}\left[\left(\int_{-\infty}^{\infty} dk\right)_s e^{-ikx} f(k)\right]. \tag{12}$$

For example, we can write

$$w(x, 0, t) = \mathrm{Re}\left[\left(\int_{-\infty}^{\infty} dk\right)_s e^{-ikx} w(k, 0, t)\right]. \tag{13}$$

In formulas (11) and (12), as well as in the subsequent analysis, $f(x) \equiv f(x, z, t)$, $f(k) \equiv f(k, z, t)$. In the momentum representation, the equivalent of substitution (10) in the wave number has the form

$$k \longrightarrow \hat{k}, \quad \hat{k} = kgt^2. \tag{14}$$

Let us analyze the transformation of the spectral densities (Fourier transforms) of the required functions as a result of the self-similar substitution. Here, there is a delicate aspect associated with the delta-correlation of the Fourier transform of $f(k)$ and with peculiarities in the integration of delta-correlated functions. In the case of smooth functions $f(k)$, the Fourier expansion (13) in the self-similar mode has the form

$$\hat{w}(\hat{x}, 0) = \mathrm{Re}\int_{-\infty}^{\infty} d\hat{k} e^{-i\hat{k}\hat{x}} \hat{w}(\hat{k}, 0),$$

$$w(k, 0, t) = g^2 t^3 \hat{w}(\hat{k}, 0),$$

where $\hat{k}$ is defined by relation (14). The spectral functions $f(k)$ in the problem on mixing differ considerably from smooth functions, and the self-similar relation $w(k) \propto t^3 \hat{w}(\hat{k})$ is not observed. The correct relation $w(k) \propto t^2 \hat{w}(\hat{k})$ will be derived below.

In the problem under investigation, $w(x)$ is a random periodic function, while $w(k)$ is a delta-correlated (needle-shaped) function [3, 4]. In order to emphasize the peculiarities in the integration of such strongly nonsmooth functions, we have introduced the parentheses and subscript s into formulas (12) and (13). Replacing the integral by the limit of the sum, we must replace the linear differential dk by the root differential $\sqrt{dk}$ [3, 4]. Accordingly, the spectral density $f(k)$ has fractional dimensions (it contains a half-integer power of the dimension of length).

This circumstance has an important corollary associated with the scaling of self-similar representations of the required functions. Owing to the root differential, the spectral densities are transformed under the self-similar substitution according to the following rules

$$H(k, t) = g^{3/2} t^3 \hat{H}(\hat{k}, t),$$

$$\rho(k, z, t) = \rho_h \sqrt{g} t \hat{\rho}(\hat{k}, \hat{z}, t),$$

$$w(k, z, t) = g^{3/2} t^2 \hat{w}(\hat{k}, \hat{z}, t),$$

$$p(k, z, t) = \rho_h g^{5/2} t^3 \hat{p}(\hat{k}, \hat{z}, t),$$

where H is the mass of a column (see Section 4). In the self-similar mode, time is omitted, and hence we obtain

$$\hat{H}(\hat{k}) = \frac{H(k, t)}{g^{3/2} t^3}, \quad \hat{\rho}(\hat{k}, \hat{z}) = \frac{\rho(k, z, t)}{\rho_h \sqrt{g} t},$$
$$\hat{w}(\hat{k}, \hat{z}) = \frac{w(k, z, t)}{g^{3/2} t^2}, \quad \hat{p}(\hat{k}, \hat{z}) = \frac{p(k, z, t)}{\rho_h g^{5/2} t^3}. \tag{15}$$

These relations are significant. Among other things, relations (15) lead to the nontrivial conclusion that the maximum spectral amplitude of pressure fluctuations increases in proportion to t^3. These relations make it possible to analyze the universal spectra describing mixing in the self-similar mode. The spectra are universal in the $\hat{k}$, $\hat{f}$ coordinates.

Let us show how relations (15) are derived. By way of an example, we consider the velocity transformation at the middle cut, $z = 0$. In the self-similar mode, we have $w(x, 0, t) = gt\hat{w}(\hat{x}, 0)$, (10). The Fourier expansion (13) for the needle-shaped spectral density $w(k, 0, t)$ can be written as

$$w(x, 0, t) = \int_{-\infty}^{\infty} \sqrt{dk} e^{-ikx} w(k, 0, t). \tag{16}$$

In formula (16), we replace x by $\hat{x}$ and $w(x)$ by $\hat{w}(\hat{x})$ in accordance with (10) and dk by $d\hat{k}$ and k by $\hat{k}$ in accordance with (14). As a result, expansion (16) assumes the form

$$\hat{w}(\hat{x}, 0) = \int_{-\infty}^{\infty} \sqrt{d\hat{k}} e^{-i\hat{k}\hat{x}} \frac{w(k, 0, t)}{g\sqrt{gt^2}}. \tag{17}$$

Theoretically, the self-similar analogue of expansion (16) must have the form

$$\hat{w}(\hat{x}, 0) = \int_{-\infty}^{\infty} \sqrt{d\hat{k}} e^{-i\hat{k}\hat{x}} \hat{w}(\hat{k}, 0). \tag{18}$$

Comparing integrals (17) and (18), we find that the self-similar spectral density $\hat{w}(\hat{k}, 0)$ in (18) is equal to the fractional expression in the integrand of (17). This leads to the velocity transformation appearing in the set of transformations (15). The remaining expressions in this set can be derived similarly.

16. SELF-SIMILAR VARIABLES IN THE BOX

Let us support the formal derivation of the important relations (15) by a more visual derivation. We consider mixing in a box. There is a difference between the Fourier integral and the Fourier series. The integral is written for unbounded systems with $x \in (-\infty, \infty)$. We are interested in the case of a box and the corresponding series expansion in the box modes $\{k_n = n\}$. The derivation of relations (15) was based on the peculiarities of the integration of delta-correlated functions, $(\int dk)_s = \int \sqrt{dk}$. The main property of the spectral densities $f(k)$ and $\hat{f}(\hat{k})$ defined through integral $\int \sqrt{dk}$ is that they depend neither on the form of discretization of the k axis nor on the constraints $L_\perp$, $L_\perp = \infty$ or $L_\perp < \infty$ (this is demonstrated in [3, 4]). Consequently, they are suitable for constructing universal spectra. Let us define an analogue of the spectral densities in a box.

Far from the bottleneck ($\bar{\lambda} \ll L_\perp$), mixing in a box does not differ from mixing in an unbounded system. The Fourier spectrum in x in the mixing zone $-h_- < z < h_+$ has the form of a broad bump ($\Delta k \sim \bar{k}$) located near the characteristic wave number $\bar{k}$ (see Sections 4, 7, and 12), Fig. 3. The bump contains approximately $\bar{k}/(k_{n+1} - k_n) = L_\perp/\bar{\lambda}$ equidistant ($k_{n+1} - k_n$ = const) box modes. It is important that at a large distance from the bottleneck, this number is large ($L_\perp/\bar{\lambda} \gg 1$). Let us consider a large (l) and a small (s) box with the lateral size L_l and $L_s < L_l$, respectively. As long as $\bar{\lambda} \ll L_s$, the flows in the boxes starting from the same instant are statistically equivalent. Consequently, $f_l(x) = f_s(x)$, where the coordinate functions $f_l(x)$ and $f_s(x)$ correspond to different boxes. At the same time, the amplitudes of the Fourier expansion of functions $f(x)$ in different boxes are different: $(f_n)_l \neq (f_n)_s$, $(f_n)_l < (f_n)_s$. This makes these amplitudes inconvenient. Clearly, this is due to the difference in the number of modes constituting the bumps in different boxes since the value of $\bar{k}$ is the same, and the separation between levels k_n in the large box is smaller. An analogue of the spectral density (15), which would be invariant to the box width, can easily be obtained using the normalization

$$\tilde{f}_n = \frac{f_n}{\sqrt{\delta k}}, \quad \delta k = k_{n+1} - k_n = \frac{2\pi}{L_\perp}. \tag{19}$$

The situation with the evolution in time is similar. Let us consider a box with $L_\perp = 2\pi$. We compare the instants t_s and $t_l > t_s$, $\bar{\lambda}(t_l) \ll L_\perp$. Formulas (10) specify the time scaling for the self-similar functions $f(x)$ in the coordinate representation. For instance, $w(x, 0, t) \propto t$. Let us compare the spectral bumps of $w(x, 0, t)$ for t_s and t_l. The number of box modes covered by bump t_l is smaller. This causes an additional increase in amplitude w_n. In order to obtain an invariant quantity, which is the same in all self-similar generations of coarsening, we must compensate this increase.

Let us consider quantity $w_n(x, 0, t)/t$. In the self-similar mode, it is proportional to acceleration and constant (see formula (10)). Let us now consider the Fourier amplitude $w_n(0, t)/t$ of the Fourier expansion of this quantity. We construct the series of amplitudes not as a function of the harmonic number n, but as a function of the self-similarity index (number):

$$\hat{n} = nt^2. \tag{20}$$

We assume that $g = 1$ and $L_\perp = 2\pi$. Transformation (20) generalizes transformation (14) to the case of a box. In the self-similar mode, the position of the amplitude bump $w_{\hat{n}}(0, t)/t$ on the $\hat{n}$ axis is invariant. Let $\hat{n}_{\max}$ be the position of the peak of the bump and $\Delta\hat{n}$ be its spectral width. In the self-similar regime, these quantities do not depend on time.

Conversely, amplitudes $w_{\hat{n}}(0, t)/t$ increase with time t. This is due to a decrease in the number of modes on interval $\Delta\hat{n}$. The separation between adjacent modes on the $\hat{n}$ axis is given by

$$\delta\hat{n} = (n+1)t^2 - nt^2 = t^2. \tag{21}$$

It can be seen that this quantity increases with t. Consequently, the number of modes in the bump is of the order of

$$N_{\mathrm{mod}} = \frac{\Delta\hat{n}}{\delta\hat{n}} = \frac{\Delta\hat{n}}{t^2} \tag{22}$$

and decreases with increasing t. The sum of Fourier harmonics $w_n(0, t)/t$ constituting the bump gives the quantity $w(x)/t$,

$$\mathrm{Re}\sum \frac{w_n(0, t)}{t} e^{-inx} = \frac{w(x, 0, t)}{t}. \tag{23}$$

In the self-similar regime, this quantity is on the average constant in time. It is important that for $N_{\mathrm{mod}} \gg 1$, the phases of adjacent harmonics $w_{\hat{n}}/t$ with numbers n and $n+1$ are not mutually correlated. In order to ensure the constancy of $w(x)/t$ (23), amplitudes $w_{\hat{n}}/t$ must therefore increase in proportion to $1/\sqrt{N_{\mathrm{mod}}}$, where N_{mod} is defined by formulas (21) and (22). As in the case (19), this increase can be compensated and the required invariant quantity can be obtained by introducing the spectral density

$$\tilde{w}_{\hat{n}} = \frac{w_{\hat{n}}(0, t)/t}{\sqrt{\delta\hat{n}}} = \frac{w_{\hat{n}}(0, t)}{t^2}. \tag{24}$$

This formula is an analogue of the spectral densities (19) and (15).

Formula (24) defines the spectral self-similar variables in the case of a box. Consequently, the spectra in variables

$$\hat{n},\ \tilde{f} \tag{25}$$

are universal for a box. Generalizing the above derivation carried out for the special case of vertical velocity at the middle cut $z = 0$, $\hat{z} = 0$ to the general case of $\hat{f}(\hat{x}, \hat{z})$, we obtain the following expressions for $\tilde{f}_{\hat{n}}(\hat{z})$ and $\tilde{H}_{\hat{n}}$:

$$\begin{aligned} \tilde{H}_{\hat{n}} &= \frac{H_n(t)}{t^3}, \quad \tilde{\rho}_{\hat{n}}(\hat{z}) = \frac{\rho_n(z,t)}{\rho_h t}, \\ \tilde{w}_{\hat{n}}(\hat{z}) &= \frac{w_n(z,t)}{t^2}, \quad \tilde{p}_{\hat{n}}(\hat{z}) = \frac{p_n(z,t)}{\rho_h t^3}. \end{aligned} \tag{26}$$

We can conditionally state that these expressions are obtained from formulas (10) by dividing by $\sqrt{\bar{n}}$, $\bar{n} \propto 1/t^2$. For example, $H_n(t) \propto t^2 \tilde{H}_{\hat{n}} / \sqrt{\bar{n}}$.

17. ON THE POSSIBILITY OF A CONVERGING DESCRIPTION OF TURBULENT MIXING

The speed and memory of multiprocessor computers are rapidly increasing. This will permit in the future to use denser meshes (to increase N_x and to reduce the spacing Δ), to reduce the time step (to decrease the Mach number and compressibility effects at late stages) and, accordingly, to extend the self-similarity range. What is the best way to take advantage of growing resources (what must the future strategy be)? We are speaking of the fundamental, or theoretical, aspect of the problem (the convergence of the self-similarity description) rather than of engineering applications. The latter always have details requiring additional resources. We will propose below a procedure for processing the results of numerical simulation, which can be refined or improved (and, hence, converging in this sense). Successive refining is associated with a gradual increase in the computer power. The procedure is based on an analysis of the statistical features of mixing, which makes it possible to separate the random noise and to single out the main aspect of mixing. The latter is associated with the existence of clusters of columns (bubbles). The description of a cluster and the interaction of columns in its requires refinement.

The primary products of any simulation (including high-quality modeling) are the sequences of instantaneous fields (frames) $f(\mathbf{r}, t_i)$, $t_1, t_2, \ldots$, similar to those presented in Figs. 1a and 1b, recorded and stored in the computer memory. Let us find out which way of product processing is the most promising for obtaining a comprehensive description of self-similarity.

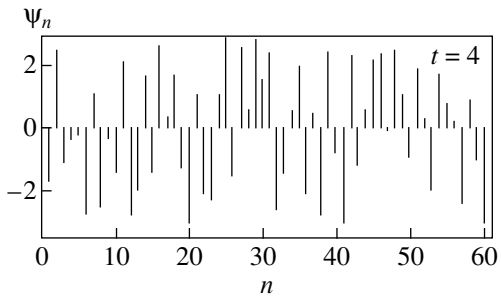


Fig. 10. Phases $\psi_n = \arctan(b_n/a_n)$ for vertical velocity $w(x, z = 0, t = 4)$.

The processing includes three stages. At the initial stage, we prove the self-similarity of mixing by verifying the conditions of statistical self-similarity of fields (6), (10) (i.e., self-similarity on the average; see Sections 5–8). It can be seen that the field recording is "long" (in the limit, $L_\perp \gg \bar{\lambda}$) and overloaded with phase information. The latter includes the recording of specific amplitudes and the positions of columns on the x axis.

At the second stage of data processing, we contract the recording in a broad ($L_\perp \gg \bar{\lambda}$) box to a recording on a finite segment. This is achieved by a transition to the momentum representation and Fourier transforms. The substitution is possible owing to the invariance to translations in a direction transverse to acceleration g. The excessive (phase) information in the spectra is contained in delta-correlated phase fluctuations (Fig. 10). In the case of such fluctuations, the phase of a harmonic in the spectrum changes by a large quantity $\sim\pi$ upon an infinitesimal displacement of the wave number k in the spectrum [3, 4]. In the case of numerical simulation, discrete Fourier series are obtained. The excessive information on the mutual phase arrangement of clusters formed by columns in the case of the series is contained in the fluctuations of phase ψ_n by a large value $\sim\pi$ from one of neighboring harmonics to another (see Fig. 10).

The phase information is of no importance. Its random nature is due to the attenuation of the correlator (see Fig. 4) upon a shift over a distance of the order of several column widths. Phases change by 100% from one realization to another. This is due to the exponential divergence of the trajectories of the corresponding dynamic system, which is responsible for the time correlation of phases over intervals of the order of a coarsening step (see Section 11 for the definition of steps q and $\sqrt{q}$). This explains the difference in Figs. 5 and 6, which were obtained as a result of computations with the same initial data and different meshes. It would be interesting in the future to extract information on the correlation time from numerical data. The ratio $q_{\text{corr}} = t_{\text{lost}}/t_{\text{obs}}$ characterizes the duration of a correlation. Here

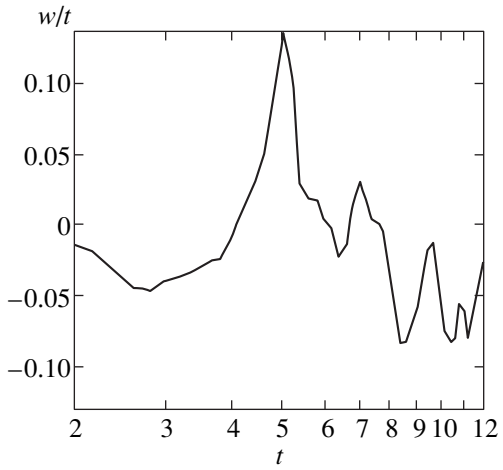


Fig. 11. Sign reversals of large-scale turbulent fluctuations. The curve describes the quantity $w(x_f, z = 0, t)/t$.

t_{obs} characterizes an arbitrary instant starting from which the correlation is measured and t_{lost} is the instant by which the correlation is lost, $t_{lost} > t_{obs}$. At the self-similarity stage, ratio q_{corr} is constant. The introduction of q_{corr} is required since the duration of a coarsening step itself increases with time (in proportion to $\propto t^2$).

The problem of time correlation is quite complicated. We fix an arbitrary point x_f on the cut $z = 0$. Let us consider the dependence of the function $W(t) = w(x_f, z = 0, t)/t$ on argument t. In the self-similarity region, this function is alternating (Fig. 11). Let Δt_1 be the duration of a time interval during which the function $W(t)$ does not change sign (e.g., is positive). The next time interval Δt_2 during which the function is positive is on the average longer than Δt_1 by a factor of q_t. Segment Δt_3 is still longer, and on the average we have $\Delta t_3 = q_t^2 \Delta t_1$. This sequence obviously forms a geometric progression. It would be interesting to find the step q_t of this progression and to compare it with quantities q_{corr} and $\sqrt{q}$ (see Section 11).

The function $W(t)$ is plotted in Fig. 11. The number of sign reversals gives an order of magnitude of the number of coarsening generation. We consider $W(t)$ as a function of $\ln t$. The quantity q_{corr} is defined by correlator

$$K(\ln t) = \langle W(\ln t' - \ln t) W(\ln t') \rangle,$$

where the averaging is carried out over $\ln t'$. In order to calculate this correlator, a long self-similar stage with many coarsening generations is required (see Sections 6 and 7 and the definitions of number N_{doubl}, ΔN_{sw}, and ΔN_1). The duration of this stage is determined by the computer power.

We considered above the initial and middle stages of data processing and determined the reasons behind phase fluctuations. It is important to get rid of the arbitrariness associated with these fluctuations since it prevails and, hence, masks the physically meaningful information. This is done at the final stage of processing, which will be described below.

18. SMOOTHING OF FOURIER TRANSFORM MODULI AND DETECTION OF CLUSTERS

Let us single out the meaningful information. For this purpose, we get rid of phase fluctuations, first, by going over to the modulus of the Fourier transform of $w_n = \sqrt{a_n^2 + b_n^2}$ (rejecting the random phase ψ_n) and, second, by smoothing the fluctuations of the modulus. The typical modulus w_n and phase ψ_n of harmonics are shown in Figs. 9 and 10. The smoothing of moduli means the replacement of w_n by the value $\bar{w}_n$ averaged over adjacent harmonics. The calculation of the mean value $\bar{w}_n$ is carried out over Δn_{intrn} neighboring harmonics. In order suppress the fluctuations associated with the random mixing of clusters, it is necessary that $\Delta n_{intrn} \gg 1$.

Clearly, high-performance computers capable of simulating a logarithmically long ($N_{doubl} = \ln_2 k_{sw1}/k_1$) self-similar cascade (see Sections 6 and 7) are required for obtaining exact mean values. As a matter of fact, the number of harmonics Δn in a bump (see Sections 7 and 12) is of the order of $L_\perp/\bar{\lambda}$. The averaging is exact in the limit $\Delta n \gg 1$ since $1 \ll \Delta n_{intrn} \ll \Delta n$.

A detailed description of self-similar mixing is contained in the smoothed spectra $\bar{H}_n$, $\bar{\rho}_n$, $\bar{u}_n$, $\bar{w}_n$, and $\bar{p}_n$ of the basic functions. It was mentioned above that these spectra are obtained as a result of three-stage processing of the results of numerical simulation. The quality of the spectra must be high enough to judge the shape of a bump in the spectrum. The examples of smoothed spectra will be given below. These spectra must be plotted in the self-similar coordinates defined in Sections 15 and 16.

An example with a middle section $z = \hat{z} = 0$ was considered by us earlier. In order to characterize the mixing zone, the spectra at several cuts are required, e.g., for $\hat{z} = \pm 0.5 h_+/gt^2$. By increasing computer resources, we can cover the mixing zone $As h_+ < z < h_+$ by a denser sequence of self-similar cuts ($z_{cut} \propto \hat{z}_{cut} t^2$, where z_{cut} is the vertical coordinate of the cut plane). An increase of computer power makes it possible to draw the spectral bump more exactly. The condensation of cuts and refining the drawing signifies the convergence of the advanced procedure of studying the self-simulating mixing. For the sake of comparison, it is worth noting that the convergence of the theory of the turbulent mixing length cannot be improved. This phenomenological theory can satisfactorily describe the vertical profile. Nevertheless, it always has an error which is generally of the order of unity.

The smoothed spectra obtained at different heights of the cut provide much richer information on the mechanics of mixing as compared to that contained in coefficient α_+ (1), (5). It is important to note that the smoothed spectra are closely related to correlators (2) [3, 4] and the clusterization of several adjacent columns. These spectra are the clusterization characteristic, or measure. For example, the length of the left wing of a self-similar bump determines the number of columns in the cluster. The rate at which the amplitude decreases as we move to the short-scale side is connected with fractal dimension. The spectrum is terminated at the viscous, diffusion, or mesh scale.

The smoothed self-similar spectra are universal (like coefficient α_+). For this reason, their computation must be used for the mutual testing of various numerical codes.

19. THEORETICAL SPECTRUM

The main parameters defining a smoothed self-similar bump in the spectrum are, first, the position $\hat{n}_{\max}$ of the bump peak on the axis of the wave numbers $\hat{n}$, second, the maximum amplitude $\hat{f}_{\max}$, and, third, the bump half-width $\Delta\hat{n}$. The spectral self-similar variables $\hat{n}$ and $\tilde{f}$ are defined in Sections 15 and 16 by formulas (14), (15), (20), 25), and (26). Let us estimate the position of the bump on the $\tilde{f}$ axis. The elevation speed w_b for a periodic bubble (i.e., the bubble in a vertical potential tube [11]) is given by formula (7). In a tube of constant width λ_{theor}, in the 2D case we have $F \approx 0.6$. We substitute k_{theor} for k_{sw} in (7) and equate speed $w_b(k_{\text{theor}})$ to velocity $\dot{h}_+$ (see formulas (1) and (5) in Section 10) [22, 20]. This gives the following theoretical estimate of the transverse scale:

$$\hat{k}_{\text{theor}}(\alpha_+, F) = \hat{n}_{\text{theor}} = n_{\text{theor}}t^2 = \frac{1}{4}\frac{F^2}{\alpha_+^2}\frac{(1+\mu)^2}{1-\mu}. \quad (27)$$

Let us estimate the typical bump amplitude $(w_n)_{\text{theor}}$ in the vertical velocity spectrum. The value of $\dot{h}_+$ gives the typical velocity (see Fig. 2, in which this velocity is compared with the velocities in columns). The scale-invariant velocity or acceleration are proportional to $\dot{h}_+/t$ (see Eq. (10) in Section 14). Let us distribute this self-similar acceleration among the N_{mod} modes with a random phase, which constitute the bump. In accordance with (9), we have

$$2\alpha_+ \text{At} g t \sim (w_n)_{\text{theor}}\sqrt{N_{\text{mod}}}. \quad (28)$$

We can write

$$N_{\text{mod}} \sim k_{\max} = 2\pi/\lambda_{\max}, \quad \lambda_{\max} = \alpha_H \text{At} g t^2, \quad (29)$$

where the coefficient α_H corresponding to the characteristic horizontal scale is introduced in analogy with coefficient α_+. Substituting N_{mod} from (29) into (28), we obtain

$$(\tilde{w}_{\hat{n}})_{\text{theor}} = \frac{(w_n)_{\text{theor}}}{t^2} \sim \frac{2}{\sqrt{2\pi}}\sqrt{\frac{\alpha_H}{\alpha_+}}\alpha_+^{3/2}\text{At}^{3/2}g^{3/2}. \quad (30)$$

Putting $g = 1$, $\alpha^+ = 5\%$, and $\alpha_H = \alpha^+$, we obtain $(\tilde{w}_{\hat{n}})_{\text{theor}} \approx 10^{-2}\text{At}^{3/2}$. The estimates similar to (30) can also be derived for the characteristic spectral amplitudes $\tilde{H}_{\hat{n}}$, $\tilde{\rho}_{\hat{n}}$, and $\tilde{p}_{\hat{n}}$ (26).

It is difficult to estimate the bump width theoretically. The problem of the bump width is connected with the problem of column clusters. We will now attempt to estimate this width by comparing the value following from formula (27) and the result of numerical simulation.

20. UNIVERSAL SPECTRUM: PRE-SELF-SIMILAR AND PAST-SELF-SIMILAR STAGES

The evolution of the spectra in self-similar variables $\hat{n}$ and $\tilde{w}_{\hat{n}}$ is illustrated in Fig. 12. These frames were obtained from the evolution frames presented in Fig. 9

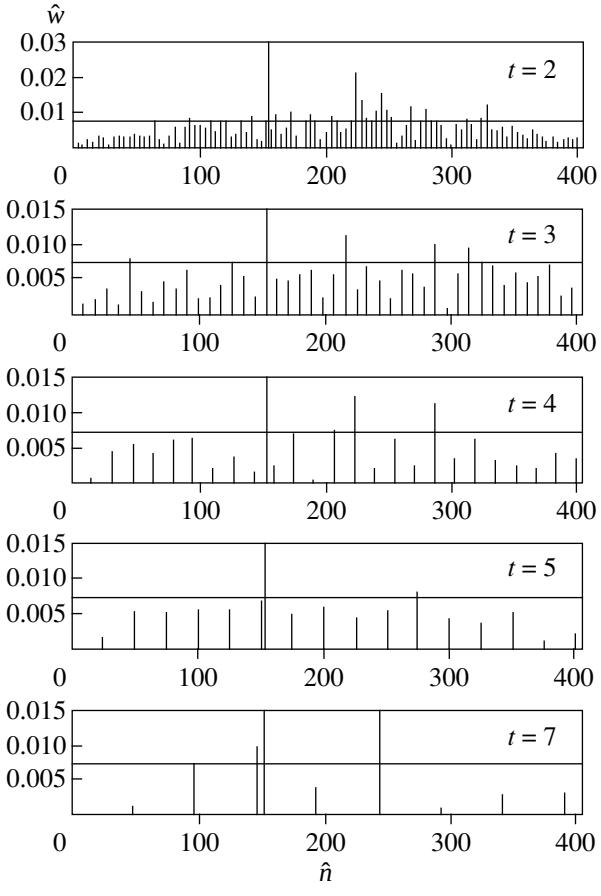


Fig. 12. Universal or "stationary" spectrum. The self-similar variables $\hat{n}$ and $\tilde{w}$ are used. The thin horizontal line represents formula (30) and the vertical line marks the characteristic inverse scale.

at the same instants through substitutions (20) and (26). The intervals between harmonics on the $\hat{n}$ axis increase with time (21), and the number N_{mod} of modes in the bump increases (22) (transition from a dense to a thin spectrum). In its evolution, the mixing process passes through the initial, self-similar, and final stages (see Section 13). At the initial stage, the spectrum is "condensed" in the "long-wave" (relative to self-similar variables) region (near point $\hat{n} = 0$). Since the velocities at $t = 0$ are finite, this spectrum differs considerably from the self-similar spectrum: $\tilde{w}_{\max} \longrightarrow \infty$ and $\hat{n}_{\max} \longrightarrow 0$ for $t \longrightarrow 0$ (quantities $\tilde{w}_{\max}$ and $\hat{n}_{\max}$ are defined in Section 19). As a result of the transient process, the long-wave spectrum is gradually displaced to the middle region of the wave numbers $\hat{n}$ and is transformed into a self-similar spectrum.

It is very important to note that a stationary bump is formed in the spectrum at the self-similar stage (see Section 9). This is illustrated in Fig. 12. The duration of the self-similarity stage and, hence, the number of generations in the inverse cascade can generally be infinitely large. These quantities are determined by the number N_x of nodes and by the relation $N_{\mathrm{doubl}} = \ln_2(k_{sw1}/k_1)$ (see Section 6). In the above calculations, we observed $N_{\mathrm{doubl}}^{\mathrm{real}} = 1.5$ to 2.5 doublings (see Section 7). During this whole time, the spectrum remains stationary on the average. This means that the smoothed spectrum is also stationary. Individual harmonics in this case live their own life. A harmonic arrives at the stationary bump in Fig. 12 from the long-wave region on the left. During a time of the order of the lifetime of one generation, the harmonic traverses the bump from left to right and then is lost in the short-scale region.

The reserve of harmonics in the long-wave region is exhausted with time. The bump approaches the scale k_1 determined by the width of the box (bottleneck, see Section 7). The number of modes decreases thereby to $N_{\mathrm{mod}} \sim 1$ (22). The quadratic self-similarity is terminated, and the final stage begins, in which the mixing rate is determined by the lateral boundary conditions (the restriction by lateral boundaries and the retardation of mixing; see Sections 4, 7, and 13). The bump loses its stationarity and becomes short-scale ($\hat{n}_{\max} = t^2 \longrightarrow \infty$), broad ($\hat{\Delta} n \sim t^2$), and low ($\tilde{w}_{\max} \longrightarrow 0$).

21. COMPARISON OF THEORETICAL AND NUMERICAL SPECTRAL AMPLITUDES

At the self-similar stage, the average shape of the bump in the spectrum is fixed. Amplitude (30) is shown on the spectra in Fig. 12 by a thin straight line. It can be seen that the amplitudes and the position of the self-similar bump are approximately fixed on the time interval $2.5 < t < 6$. A comparison of the theoretical and numerical spectra involves the comparison of amplitudes and wave numbers. It is significant that the self-similar amplitude is in good agreement with the theoretical estimate (30). Let us go over to wave numbers.

22. MECHANICS OF THE RAYLEIGH–TAYLOR TURBULENCE

Before we go over to wave numbers and transverse dimensions, let us analyze the relation between the horizontal and vertical scales. The most important parameter of turbulent mixing is coefficient α_+. It defines the displacement, velocity, and acceleration of the upper boundary of the mixing zone (5) along the vertical. It should be noted, by the way, that in applications, it determines the path length $S_{ff}(t_{perf}) = gt_{perf}^2/2$ traversed by a foil of thickness h_{foil} being accelerated by the instant of perforation t_{perf} [23]. The ratio $h_{foil}/S_{ff}(t_{perf}) \approx 2\alpha_+ \mathrm{At}$ gives the relative shell width in the problems on inertial fusion. The relation $h_{foil} \approx \alpha_+ \mathrm{At} g t_{perf}^2$ leads to an estimate of t_{perf}. It is expedient to have as large values of S_{ff}/h_{foil} as possible for optimizing compression and for increasing the energy yield [24].

Coefficient α_+ is determined by a complex combination of random and nonlinear processes. Let the wavelength λ_{theor} appearing in formula (27) be given by formula (29) ($\lambda_{\mathrm{theor}} = \alpha_H \mathrm{At} g t^2$). Here, coefficient α_H is connected with the horizontal scale determining the rate of penetration of the mixing zone into the heavy liquid and, hence, the primary (coarse) fragmentation of this liquid. Substituting expression (29) with the coefficient α_H defined in this way into formula (27), we obtain

$$\alpha_+ = \frac{F^2}{8\pi}(1+\mu)R_{HV}, \quad R_{HV} = \frac{\alpha_H}{\alpha_+}. \tag{31}$$

We will refer to quantity R_{HV} in (31) as the horizontal–vertical ratio. Putting $R_{HV} = 1$ and assuming that $F = 1/\sqrt{3}$ (2D analysis in the Layzer approximation [2–4, 11]; see also [25–29]), we obtain

$$\alpha_+ = \frac{1}{24\pi}(1+\mu) \approx 1.3(1+\mu)\%. \tag{32}$$

Estimates (31) and (32) explain the reasons behind the numerical smallness of coefficient α_+. This is associated with geometrical limitations ("2π effect", the ratio of k to λ, see Section 8), the dynamics of separation of the homogeneous flow of a heavy liquid by a counterflow of bubbles (coefficient F), and the fabrication of horizontal scales through nonlinear and random processes (λ and ratio R_{HV}). An analysis of the density fields presented in Figs. 1 and 5–7 shows that the position of the upper front of the mixing zone is determined by the current positions of the tops of columns (bubbles) intruding into the heavy liquid. In order to "raise" the bubble tops, say, to $\alpha_+ = 4\%$ from the too low value

given by (32), we must increase coefficient F and ratio R_{HV}. The product F^2R_{HV} must be increased approximately by a factor of three as compared to the value of $F^2R_{HV} = 1/3$ adopted in (32).

Relation (31) contains the vertical (α_+) and the horizontal (R_{HV}, clusterization) measures as well as the dynamic factor F. The latter is associated with the recalculation of the horizontal scale into the vertical one. An analysis of the mechanics of mixing involves the study of the dynamic (F) and scale (R_{HV}) factors. The information on coefficient α_+ and ratio R_{HV} is extracted from numerical simulation. An estimate of factor F follows from the theoretical limitations imposed on its possible values. We will consider factors F and R_{HV} separately.

23. COLLIMATION EFFECT AND THE DYNAMIC ACTION OF A FAST ASCENDING JET OF A LIGHT LIQUID ON A CONTINUOUS HEAVY LIQUID

The dynamic resistance to the intrusion of bubbles into a heavy liquid, which is responsible for the smallness of coefficient F, is mainly concentrated in the vicinity of the bubble top (where the column "pricks" the heavy liquid). The bubble separates the initial one-phase (without a light liquid) heavy liquid (initial dispersion). The separation of the "approach steam" of the heavy liquid takes place at the top of the bubble. The elevating speeds of bubbles in the region of the mixing zone front determine the velocity of the front itself (coefficient α_+).

It is usually assumed that columns in the turbulent mixing zone are approximately similar to periodic bubbles [20, 22]. In estimates, it is assumed that the light liquid under the dome of a bubble near its top moves as a single whole together with the dome. The dynamic pressure (ram) p^l_{ram} of the light liquid is equal to zero. In this case, the balance of the buoyancy and the ram p^h_{ram} of the heavy liquid leads to formula (7) with the Archimedean factor $\sqrt{1-\mu}$ and with $F \approx 0.6$ in the 2D geometry and with $F \approx 1$ in the 3D geometry [11].

Let us consider in greater detail the interaction between the light and heavy liquids in the upper part of the mixing zone. An analysis of the results of calculations shows that the pressure p^l_{ram} in the turbulent case is significant for the bubble dynamics and for estimating α_+ if the density ratio μ is not very small, $\mu > \mu_{thr}$, $\mu_{thr} = 10^{-3}$–10^{-2} (the range of $\mu < \mu_{thr}$ corresponds to ultralow values of parameter μ). This means that in an analysis of bubbles, the inertia of the light liquid must be taken into account along with the gravitation and inertia of the heavy liquid.

The decisive role of pressure p^l_{ram} in the formation of "mushrooms" at the jet wakes is obvious. Under the action of this pressure, the jet of the heavy liquid decelerates, splits (2D) (or assumes the shape of a disk in the 3D case), and turns back. The drag and splitting effect is closely related to the formation of vortex dipoles (2D) and rings (3D) and to the problem of collision of jets [3, 4, 30]. It is typical of any jet (gravitation may be zero) propagating in a medium.

The analyses of periodic flows prove that the maximum velocity of a light liquid can be high. For example, in some calculations [3, 4, 30], it was an order of magnitude larger than the elevation speed of a bubble. High velocities are attained at late stages. For $\mu = +0$, the light liquid in a bubble ascends with the velocity of the bubble $+w_b$ (the z axis is directed upwards). Only a small volume of the light liquid near the end of the heavy jet moves with the velocity $-w_j$ of this jet. At late stages, we have $w_j \gg w_b$.

We will confine our analysis to moderate and small values of parameter μ, $\mu_{thr} < \mu < \mu_{mid}$, where $\mu_{mid} \approx 0.3$. In a steady-state bubble, the light liquid ascends with a velocity w_b. In our earlier calculations [30], for $\mu = 1/10$ at relatively late stages, the heavy jet traverses considerable distances in the gravitational field, and the substance in it is accelerated to high velocities w_j. This jet is decelerated with the formation of a mushroom. The heavy liquid in the jet is accelerated gradually, while the deceleration by the dynamic pressure of the light liquid is quite sudden. With such a drag, the gravitational acceleration is quite small as compared to the acceleration due to the dynamic pressures exerted by the light and heavy liquids. Large positive velocities of the light liquid $w_{\text{coll}} \gg w_b$ are attained in the vicinity of the mushroom-shaped formation. This is associated with the narrowing of the duct for the light liquid into a bubble. This is essentially the collimation of the light liquid jet. In a periodic bubble, the segment of the light jet collimated by the mushroom is far from the bubble top. In the vicinity of the top, pressure p^l_{ram} is small, and estimate (7) is approximately valid.

A new circumstance following from an analysis of the turbulent fields of ρ and w is associated with the collimation of light jets by long tubes leading a fast jet to the region of the bubble top (see Figs. 1, 2, 5 to 8). The formation of narrow ducts for the lifting of the light liquid to a bubble is a common phenomenon which is important for the dynamics of mixing. Such ducts are often constricted and form "trapped" bubbles separated from the main volume of the light liquid. This resembles the bubbling observed during emptying bottles. Such phenomena are accompanied with noticeable pressure jumps ("booms") and are important for the transformation of a small part of the mixing energy into the acoustic energy. As a result, the field of acoustic vibrations is formed around the mixing zone. The prob-

lem of acoustic generation is interesting and requires further investigations.

Let us estimate the effect of pressure $p^l_{\rm ram}$ on the dynamics of bubbles under these conditions. Let us analyze Figs. 1, 2, 5–8. In Figs. 1b and 2b, fast narrow jets of the light liquid can be clearly seen at $x \approx 0.4$, 3, and 5.3. High-intensity well-collimated jets pass to large bubbles located at $x \approx 1.36$ and 3.6 in Fig. 6b. An analysis of the graphs similar to those shown in Fig. 2 leads to the conclusion that the lifting velocity $w_{\rm coll}$ can attain high values $4\dot{h}^{5\%}_+$ (see, for example, Fig. 2b with a light jet at $x \approx 0.4$).

The acceleration of the light liquid in ducts or channels to high velocities is due to a considerable difference $\Delta p_{mb} = p_m - p_b$ between the main volume of the light liquid and a bubble (pulling or suction of the light liquid by the bubble). Bubbles ascend to a considerable height in the heavy liquid. The pressure p_b at such a height is low. The pressure difference Δp_{mb} is maintained since the blocks of the heavy liquid with a large size along x "rest" on the underlying main volume of the light liquid. The blocks are conglomerates, or clusters of finer jets. Such clusters can clearly be seen in Figs. 5 and 6. For example, intense conglomerates are located at $x \approx 1.6$ and 4.6 in Fig. 5b, at $x \approx 0.6$, 2.8, and 5.6 in Fig. 6b, and to the right and left of the central bubble in Fig. 6a. The block structure is also manifested in Fig. 7b corresponding to the calculations made for a smaller value of the Atwood number.

Heavy blocks are quite broad. The distances between large bubbles separated by these blocks are of the order of $(1.5–2)h^{5\%}_+$. The heavy substance filling the major part of the block volume precipitates slowly (see Fig. 2b). Fast heavy jets are narrow and descend in the light liquid at large depths under the blocks (see Figs. 1b, 1c and 2c). The fragments associated with the previous generations are present near the upper and lower fronts of the mixing zone. This succession can be easily revealed from an analysis of a sequence of evolution frames. At the upper front, such fragments are individual relatively small (as compared to the current scale) bubbles which have lost their connection with the main volume of the light substance. At the lower front, these are strongly extended and tortuous fragments of heavy jets, which were separated from the main volume at the previous evolution stages. The fragments gain in position but lose in velocity in favor of new generations and, hence, are gradually absorbed in the central mixing zone. The "resting" of the blocks on the underlying volume indicates that the downward acceleration of the heavy substance in them is much smaller than the Archimedean value. Indeed, for ratios $\mu > \mu_{thr}$ that are not very small, the value of $2\alpha_- {\rm At} g = 2\alpha_+ {\rm As}\,{\rm At} g$ amounts to 1/10–1/5 of $(1 - \mu)g$. As a result of the resting, the conglomerate tends to shut off the access to bubbles. For this reason, the channels become quite narrow and are blocked from time to time by dams of heavy liquid.

Tubes in a bubble are sustained owing to a high speed of transportation. The ram pressure $p^l_{\rm ram}$ is significant. Under this pressure, heavy dams blocking the duct are pulled into a bubble or are pricked. The light liquid flowing into the bubble "catches" the heavy substance from the tube walls. This can clearly be seen in all density fields depicted in Figs. 1 and 5 to 7. As a result of such a "catching," a fine mixture of the heavy and light phases is formed.

Let us now estimate the effect of the inertia of the light substance on factor F. We consider the problem of intrusion of a jet into the half-space (the problem of penetration of a solitary cumulative jet through an armor jacket) [31, 32]. A generalization to the periodic case is given in [3, 4, 30]. In these publications, a chain of heavy jets penetrating into the half-space filled with a light substance is considered. The density ratio is insignificant. The solution of the problem can be transformed into the case we are interested in, where light jets penetrate into a heavy substance. If we disregard gravitation, the Bernoulli's integral leads to the equality of the pressures exerted by the heavy and light liquids at the point of stagnation (top) of the jet, $\rho_h (w^c_h)^2 = \rho_l (w^c_l)^2$, where w^c are velocities in the reference frame associated with the top. Returning to the laboratory reference frame in which the liquids are at rest outside the mixing zone, we obtain

$$w_{\rm coll} = (1 + 1/\sqrt{\mu}) w_b, \qquad (33)$$

where w_b and $w_{\rm coll}$ are the elevation speeds for the top of a bubble and for a light jet in the laboratory frame.

In the calculations with parameter $\mu = 1/10$, the typical value of the density $\rho_{\rm mix}$ of the phase mixture sucked into a bubble through a tube is of the order of $(2 \text{ to } 3)\rho_l$. Accordingly, we have $1 + 1/\sqrt{\mu_{\rm mix}} \approx 3$ to 4. Using the velocities $w_{\rm coll}$ from the graphs similar to those presented in Figs. 2b and 8 in formula (33), we find that the typical values of $w_{\rm coll}$ are sufficient to ensure $w_b \approx \dot{h}_+$ with $\alpha_+ = 5\%$.

The values of w_b, h_+, and α_+ associated with complex, distributed, or integrated action of the gravitational load exerted by the heavy substance on the light one were estimated by us earlier. With such an action, the heavy substance sags under the broad blocks or conglomerates which separate coarse bubbles ascending with the maximum speed. The transverse size of the bubbles are much smaller than the width of the conglomerate. Under the weight of large blocks, fast jets of the light liquid injected into the bubbles are dispersed. The momentum of these jets is transferred to the upward motion of the heavy liquid with velocity (33) in relatively small neighborhoods of bubble tops. In such

a complex way, the exchange of the heavy and light substances is executed, and the center of mass of the system is lowered in the case of the integrated mechanism.

A simple or local estimate of the gravitational values of w_b, h_+, and α_+ (disregarding the momentum of light jets) on the basis of the steady-state periodic solution is given by formula (7). Substituting into this formula the cluster width $\bar{\lambda}_{\mathrm{dyn}} \approx 2h_+^{5\%}$ typical of Figs. 1 and 5–7 and taking $\mu_{\mathrm{mix}} \approx 0.2$, we obtain $\alpha_+ \approx 3\%$. For example, the speed of a bubble located between the clusters of heavy jets at $x \approx 1.2$ in Fig. 6b is $w_b \approx 0.45$, $\dot{h}_+^{5\%}(t = 6) \approx 0.5$. The local estimate (7) gives $w_b \approx 0.36$ for $\bar{\lambda}_{\mathrm{dyn}} \approx 2h_+^{5\%}$. It follows hence that the local gravitational effect is slightly smaller than the integrated effect. It should be noted that estimate (7) is valid for steady-state periodic bubbles. Such bubbles have diameters approximately equal to their separation. If these distances are $\approx 2h_+$, the diameters must be approximately the same. The diameters of the bubbles in Figs. 5 and 6 are noticeably smaller. This means that they have not reached the steady state and are in the acceleration mode. The stationary phase can be approached from the single-mode harmonic perturbation as well as from a "solitary" bubble. In the former case, the radius of curvature at the bubble top decreases with increasing t from a large value R_{ini} and asymptotically tends to the steady-state, or the limiting value R_{stat}, $R_{\mathrm{ini}} > R_{\mathrm{stat}}$. In the latter case, on the contrary, the initial radius R_{ini} is small, $R_{\mathrm{ini}} < R_{\mathrm{stat}}$. The bubble radius increases with time, tending to R_{stat}. The initial stages of the bubble expansion in the second case resemble the expansion of a bubble in a wedge (2D) or in a cone (3D) [2] or the expansion of a solitary bubble [33] (a bubble in a vertical strip (2D) or in a constant cross-section tube (3D) corresponds to the periodic case [11]). Formula (7) in this case gives an upper estimate of the elevation speed of a bubble being accelerated.

In order to judge the relative importance of the local and integrated gravitational mechanisms of dispersion of a heavy liquid more precisely, we must analyze the fields of acceleration $a = v_t + (v\nabla)v$ and pressure. If the local action dominates, first, the acceleration a_l of the light liquid in the upper part of the mixing zone is positive and small ($\approx 2\alpha_+ \mathrm{At} g \approx 0.1(1-\mu)g$) and, second, the pressure in a bubble varies insignificantly (for moderate values of μ). Otherwise, first, the acceleration a_l at the top of a bubble is negative (the z axis is directed upwards) and its magnitude is larger than $0.1(1-\mu)g$ (it is of the order of g) and, second, the pressure in a bubble increases in the upward direction. In the entire flow, this is the only region where the pressure gradient may reverse its sign. If the elevation of a bubble is determined by the momentum of light jets (in this case, formula (33) is valid), the pressure drop along the boundary of the bubble in the direction from top to bottom must be faster than the hydrostatic pressure drop in the heavy liquid in the upward direction.

It should also be noted that positive accelerations of the heavy liquid are quite small. These accelerations are associated with the elevation of a bubble. In the periodic case, using the parabolic approximation [11], we can easily calculate the field of accelerations a_h of the heavy substance from the analytic formulas for the velocity potential. At the linear stage of the evolution of a single-mode harmonic perturbation, acceleration a_h is small and decreases monotonically on the bubble axis in the upward direction from the top. As we go over to the nonlinear mode, the dependence of $a_h(\Delta z, t)$ on Δz becomes nonmonotonic; in this case, coordinate Δz is measured from the top, $a_h(\Delta z, t) = a_h(x = 0, \Delta z, t)$ (2D) or $a_h(x = 0, y = 0, \Delta z, t)$ (3D), and the center $x = 0$ or $x = y = 0$ is at the top. The point where acceleration a_h attains its maximum value appears above the top. At the stationary stage, this maximum lies at a distance $k^{-1}\ln 2 \approx 0.1\lambda$ above the top, where λ is the period of the flow. This refers to the 2D geometry as well as to hexagonal and square lattices in the 3D geometry, which are most interesting [11] in connection with the 3D turbulence. The maximum value of a_h is $(1-\mu)g/12$ in the 2D and $(1-\mu)g/4$ in the 3D case.

24. SPECTRAL WIDTH OF THE BUMP AND DETERMINATION OF CLUSTER "SUPERSTRUCTURE"

The fact that the integrated effect is more important indicates that the value $F \approx 0.6$ given by the local estimate (7) (2D) is slightly lower than the actual value of F. The above analysis shows that coefficient F is effectively larger than the value given by (7) by 20 to 30%. Accordingly, the Froude number F^2 is actually larger approximately by a factor of 1.5. According to formula (31), the product of this number by the horizontal–vertical factor determines coefficient α_+. It turns out that in order to ensure the values of $\alpha_+ = 4$ to 5%, the ratios $R_{HV} = 1.7$–2 are required. Let us see whether this is in accord with the spectral data.

We consider transverse structures in the mixing zone. The information on these structures is contained in the spectra in the horizontal coordinate invariant to displacements. Typical examples of such spectra are presented in Figs. 9 and 12. Let us analyze these spectra. The central part of the bump in the self-similar spectrum in Fig. 12 corresponds to $\hat{n} = 180$ to 250. Figure 12 presents the results of calculations obtained with $\mu = 1/10$. For larger values of μ, the bump is displaced to the right in accordance with formula (27). The right (ultraviolet) wing of the bump is formed by high-frequency harmonics, while the left (infrared) long-wave wing corresponds to subharmonics. The value of parameter $R_{HV} = 1$ is associated with the self-similar wave number $\hat{n}_+ = 2\pi/\mathrm{At}\alpha_+$,

$\hat{n}_+(\alpha_+ = 5\%, \mu = 1/10) = 150$. It follows hence that if we are guided by the peak of the bump, the central part of the mixing zone ($\hat{z} \approx 0$) is characterized by rather small transverse scales, $\lambda_\perp = (0.6–0.8)h_+$, $\hat{n}_{\max} = (1.2–1.5)\hat{n}_+$. This is in accord with the number of spikes of the functions of x in Figs. 1c, 2, and 8 and with the correlation length in Fig. 4 (cf. the average separation between the spikes and the length of the horizontal segments in Figs. 1c and 2).

It is important to note that there is a spread in the scales of the spectra towards the long-wave region. Scales $\hat{n}_{\max} = (1.2–1.5)\hat{n}_+$ cannot ensure the values of $\alpha_+ = 4$ to 5%. Using the elevation speed of the bubbles with such a diameter and evaluating the self-similar velocity of the upper mixing front, we only obtain $\alpha_+ \approx 1\%$. The expansion dynamics is determined by the subharmonic wing in the harmonic distribution (see Fig. 12). Relation (27) leads to

$$\hat{n}_{\text{theor}}(\alpha_+ = 5\%, F = 0.6) \approx 50 \qquad (34)$$

for $\mu = 0.1$. The subharmonic wing extends to values of the order of (34) (see Fig. 12). Using the correction of F from Section 23, we obtain $\hat{n}_{\text{theor}} \approx 75$.

A well-developed structure with $n_{\max} \approx 1.2n_+$ and $\lambda_\perp \approx (0.7–1)h_+$ can be easily singled out in the central zone $\hat{z} \approx 0$ in the simulation data. A detailed analysis reveals a coarser "superstructure" responsible for the observed rate of expansion of the mixing zone. It corresponds to $\lambda_{\text{dyn}} \approx 2h_+$ and $n_{\text{dyn}} \approx n_+/2$. In the lower part of the mixing zone, the presence of the large-scale structure is manifested in the form of the envelope or a long-wave modulation. This can clearly be seen in Fig. 1c. Alternating black and white columns (the width of a pair is $\approx(0.5–0.7)h_+^{5\%}$ are grouped into clusters comprising two or three pairs of columns. Such clusters form the above-mentioned blocks of heavy jets. The clusters are separated by large bubbles. They correspond to high-intensity white columns in Fig. 1c at $x \approx 0.4$ and 3.2. Large bubbles are associated with fast light jets manifested in the form of large-amplitude positive spikes in Fig. 2b, which are narrow in the x direction. Summing up, we can state that the characteristic value of $R_{HV} = \bar{\lambda}_{\text{dyn}}/h_+$, which is estimated from the superstructure, is of the order of 1.7–2. It should be emphasized that the scale of $\bar{\lambda}_{\text{dyn}}$ determines not a simple variability with this step, but a long-wave modulation with a smaller period. For this reason, the spectral bump is broad instead of being concentrated in the vicinity of the wave number $\bar{k}_{\text{dyn}}$.

25. 3D GEOMETRY

It would be interesting in the future to generalize the above analysis to the 3D geometry. In this connection, we will briefly consider the peculiarities of the 3D case. Turbulence is uniform relative to displacements in the transverse x, y plane. For this reason, a theoretical analysis is based on the solutions periodic in coordinates x and y [11] and on the spectral models closely related to these solutions.[4] It is important to single out the horizontal–vertical characteristics and the main (or coarse) cell in the mixing layer.

Formula (7) connecting period λ and the limiting velocity w_b remains valid. Coefficient F assumes a different value. We denote it by F_{3D}. The periods $\lambda = 2\pi/k$ of the lattices are determined in [11]. The solutions with the symmetry of the quadratic ($4m$) or hexagonal ($6m$) lattices are approximately defined by the pair $\mathbf{k}_i$, $i = 1, 2$, or triple, $i = 1, 2, 3$, of wave vectors directed at 90° ($4m$) or 60° ($6m$) relative to one another. Each vector $\mathbf{k}_i$ is associated with its own plane wave. We will be interested in the spectra of quantities $f(x, y, z, t)$ in an arbitrary vertical plane. Such a plane is parallel to the z axis.[5] The characteristic wave number is slightly smaller (on the average, by 10–20%) than the wave number associated with the period of the solution due to the projection of wave vectors onto this plane (the plane forms an arbitrary angle with vector $\mathbf{k}_i$).[6] We will refer to this effect as the contraction of k and increase in R_{HV} due to projection.

The self-similar relations (1) and (5) can be extended to the 3D case. Equating velocities (7) and $\dot{h}_+$ as was done in Section 22, we arrive at formula (31) in which quantities α_+^{3D}, F_{3D}, and R_{HV}^{3D} appear. It follows from natural and numerical experiments that three dimensions weakly affect the rate of initial dispersion of the heavy substance, $\alpha_+^{2D} \approx \alpha_+^{3D}$ [35–40].[7] On the other hand, it is very important that the 3D Froude number F_{3D}^2 on the $6m$ and $4m$ lattices is thrice as large as in the 2D case [11]. These lattices are most interesting in connection with the problem of 3D turbulence. According to formula (31), this leads to the important conclusion that the relative horizontal scale in the 3D geometry is much smaller ($R_{HV}^{3D} \approx 1$) than in the 2D case. A small correction due to projection does not change this conclusion. Consequently, the main cell in the 3D case is more elongated along the vertical than in the 2D case.

[4] An interesting spectral model proposed by Wilson *et al.* [34] is worth noting.

[5] The so-called "laser knife" plane in the rated 3D region in the form of a parallelepiped [11].

[6] For instance, in the case of a quadratic lattice in a plane parallel to vector $\mathbf{k}_i$, the characteristic wave number is equal to k, while the characteristic wave number in the diagonal plane is smaller by a factor of $\sqrt{2}$, or by 30%.

[7] The effect of the flow divergence or convergence on coefficient α under the compression or dispersion of cylindrical and spherical envelopes is considered in [41].

It would be interesting to compare the roles of the local and integrated gravitational effects and to consider the collimation in fast narrow jets (see Section 23). The integrated effect is smaller in the given case probably due to the elongated shape of the cell. The 3D situation is topologically much richer than the 2D case [11, 42]. This also refers even to the simple periodic case. An interesting topological peculiarity of the 3D case is the formation of flattened jets around spherical bubbles and the intersection of these flat jets in the form of three-prong stars [11]. Such intersections are typical of a dense packing of bubbles. In the periodic case, the bubbles are densely packed in blocks or conglomerates (see Section 23). Large bubbles between the blocks are torn out and are involved in the accelerated motion. The effect of neighbors on these bubbles becomes weaker. For this reason, they expand not only in the vertical direction, but also along the horizontal. The packing in the 3D case is denser and more uniform probably due to the smallness of the horizontal–vertical ratio R_{HV}^{3D}. These effects require additional investigations.

It should be noted that the scaling of spectral self-similar variables in the 3D geometry changes as compared to the 2D case. This is due to the fact that the Fourier series of the physical variables $f(x, y, z_f, t_f)$ are expressed not through a single sum over harmonics n (see Section 4), but through a double sum over indices n and m of the form

$$f(x, y, \ldots) = \sum\sum (a_{nm}c_{nx}c_{my} + b_{nm}c_{nx}s_{my} + d_{nm}s_{nx}c_{my} + e_{nm}s_{nx}s_{my}),$$

where $c_{nx} = \cos nx$, $s_{nx} = \sin nx$ (see, for example, [11]). Accordingly, the spectral bump contains not $\sim\bar{n}$, but $\sim\bar{n}^2$ harmonics. At the self-similarity stage, the average wave number decreases (coarsening) according to a power law, $\bar{n} \propto 1/t^2$. Let us consider the velocity scaling. Quantity $w(x, y, \ldots)/t$ is invariant at the self-similar stage (see Section 14, formula (10)). It is created by the resultant action of $\sim\bar{n}^2$ harmonics with a random phase. Consequently, $w(x, y, \ldots)/t \sim (w_{nm}/t)\sqrt{\bar{n}^2}$ (cf. Section 16). Therefore, the invariant spectral amplitudes are given by

$$\tilde{H}_{\hat{n}\hat{m}} = \frac{H_{nm}(t)}{t^4}, \quad \tilde{\rho}_{\hat{n}\hat{m}}(\hat{z}) = \frac{\rho_{nm}(z, t)}{\rho_h t^2},$$

$$\tilde{w}_{\hat{n}\hat{m}}(\hat{z}) = \frac{w_{nm}(z, t)}{t^3}, \quad \tilde{p}_{\hat{n}\hat{m}}(\hat{z}) = \frac{p_{nm}(z, t)}{\rho_h t^4}.$$

These expressions are analogues of formulas (26) in the 3D case.

26. SIMPLE AND CLUSTER-TYPE RANDOM PERIODIC FUNCTIONS

The numerical simulation on dense meshes involves expensive and time-consuming computations. A number of spectral and statistical properties of turbulent fields can be studied using simple examples of random periodic functions [3, 4]. In particular, we are speaking of convergence problems or comparative analysis of simple and cluster cells. Such functions can be obtained using various methods. We will consider here the following example. Let us take a sequence of random points x'_j on the x' axis, such that

$$x'_{j+1} = x'_j + 1 + \beta(2\xi_j - 1), \quad x'_1 = 0, \qquad (35)$$

where $j = 1, 2, \ldots, 2N + 1$ and ξ_j is a random quantity distributed uniformly on interval [0, 1], $\langle\xi_j\rangle = 1/2$ (for each new value of j, the quantity ξ_j is cast independently). We will refer to (35) as a random periodic chain. Parameter β is the measure of randomness. For $\beta = 0$, we arrive at a periodic chain with step 1 of length $2N$. If $\beta \gg N$, we obtain the distribution of points on the x' axis, which fills an interval of length $\sim 2\beta$. The separation between adjacent pairs of points is varied over a wide range. The limiting cases $\beta \ll 1$ and $\beta \gg 1$ are of no interest. We consider the values of $\beta \sim 1$ since the turbulent functions $f(x)$ fluctuate with an average step $\bar{\lambda}$, deviating from it on each step by a quantity of the order of $\Delta\lambda \sim \bar{\lambda}$.

We construct a model function $f(x)$. Let us contract chain (35) to interval [0, 2π] on the x axis, $x_j = (x'_j/x'_{2N+1})2\pi$. We consider triples x_{2n-1}, x_{2n}, x_{2n+1}. Index $n = 1, 2, \ldots, N$ runs through pairs of ascending and descending columns. We present function $f(x)$ on subsegment $x_{2n-1} < x < x_{2n+1}$ in the form

$$f(x) = (x - x_{2n-1})(x - x_{2n})(x - x_{2n+1}).$$

If we are dealing with the vertical velocity $w(x)$, it is positive between the first pair of points (ascending column) and negative between the second pair (descending column). From functions $f(x)$ on subsegments, we "assemble" function $f(x)$ on the entire interval [0, 2π]. It is a continuous function. We continue this function to the right and to the left with period 2π beyond interval [0, 2π]. Figure 13a shows an example of such a function with $\beta = 0.6$ and $N = 100$. A part of interval [0, $L_\perp$] of width $L_\perp/10$ is given, in which 10 pairs of columns ($N = 100$), $L_\perp = 2\pi$ must be located on the average.

The correlator $K(x)$ of function $f(x)$ (Fig. 13a) calculated by formula (2) and spectrum $|f_n|$ (the Fourier expansion of the periodic function) are presented in Figs. 13b and 1c (cf. Fig. 4 and Figs. 3, 9, 12). The central peak of the correlator at $x = 0$ is an averaged column (e.g., positive), while the adjacent minimum represents a negative column paired with the positive one. Finally, the second maximum (relative to the central peak) rep-

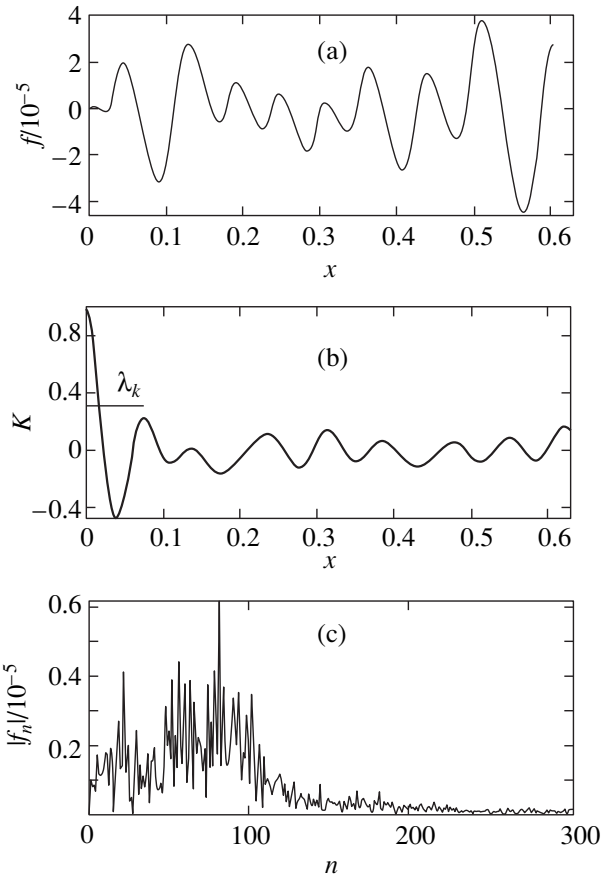


Fig. 13. (a) Example of a random periodic function specified by sequence (35), (b) its correlator and (c) spectrum.

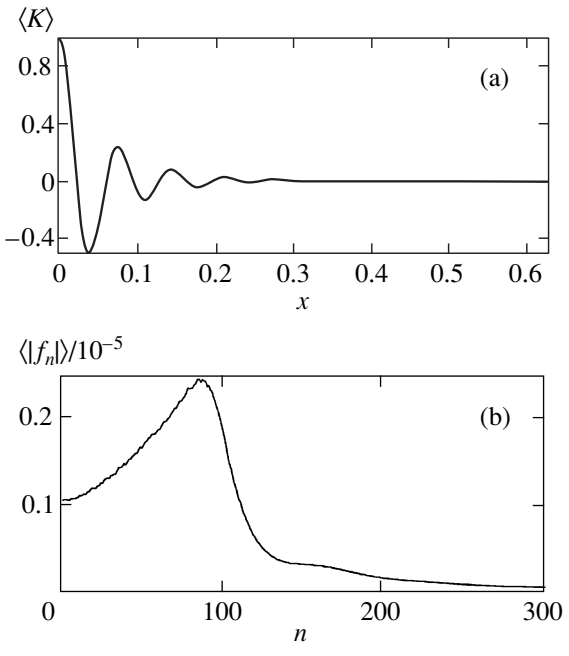


Fig. 14. Removal of insignificant phase fluctuations. (a) Correlator and (b) the spectrum of the random periodic function averaged over 4000 realizations.

resents the next positive column (relative to the first pair). In Fig. 13b, the correlation length λ_K defined in Section 4 is 20% larger than the average pair width $\bar{\lambda} = L_\perp/N$.

Figure 13 presents the realization of (35) defined by a specific random set of ξ_j, $j = 1, \ldots, 2N + 1$. Using the smoothing procedures proposed in Sections 17 and 18, which suppress the effect of a random phase, we present the results of averaging over a large number of realizations. These results are determined only by parameter β from (35). The curve in Fig. 14a describing the averaged correlator $\langle K(x) \rangle$ shows that the averaging suppresses random fluctuations (cf. Fig. 13b). The periodic component associated with step 1 in (35) is gradually "absorbed" by the random component (the term with factor β). For this reason, the peaks corresponding to consecutive positive columns decrease monotonically.[8] In the periodic case (a crystal), the correlator is periodic (see Section 4). Figure 14b shows the averaged spectrum $\langle |f_n| \rangle$. The averaging flattens random "spikes" in the spectrum.

Let us describe the spectrum $|f_n|$ presented in Fig. 13c. The typical size λ_F determined from the peak value n_F of Fourier's amplitudes ($\lambda_F = 2\pi/n_F$), as well as the correlation length λ_K, is slightly larger than the mean value $\bar{\lambda}$ (Fig. 14b). The pairs of columns $f(x)$ have no internal structure on a scale $\sim\bar{\lambda}$ (see Fig. 13a). For this reason, the short-scale wing of the spectral bump decreases rapidly for $n > N$. For $n \gg N$, we have $|f_n| \sim 1/n^2$ since in this case function $f_*(x)$ is "stitched" with the discontinuity in the derivative at the points of joining of the subsegments.

The long-wave (left in Figs. 13c and 14b) wing is important since its width determines the dynamics of mixing (see Sections 22–24). Stochastic processes broaden the spectral bump in the left direction. Let us illustrate this statement. We consider the function $f_*(x)$ composed of arcs of parabolas with random heights. Let us determine f_*. We divide interval $[0, 2\pi]$ into N_* equal subsegments by the periodic sequence of points $x_*^j = (j-1)2\pi/N_*$, $j = 1, \ldots, N_* + 1$. On each subsegment, we have $f_*(x) = (2\xi_j - 1)(x - x_*^j)(x - x_*^{j+1})$. Spectrum f_n^* decreases in the short-scale direction in proportion to n^{-2}. The central part of the spectrum corresponds to harmonics $n \approx N_*$. Let us suppose that $N_* \gg 1$.

[8] This resembles the structural function characterizing the ordering of neighboring molecules in a liquid.

We consider the long-wave asymptotic form of $n \ll N_*$, for which

$$a_n = \pi^{-1}\int_0^{2\pi} dx\, f_*(x)\cos nx.$$

We fix n and divide this integral by the sum over individual waves:

$$a_n = \sum_{i=1}^{n} I_i, \quad I_i = \int dx\, f_* \cos nx,$$

over x from $(2\pi/n)(i-1)$ to $(2\pi/n)i$. The integration domain of I_i contains N_*/n parabolic arcs. For this reason, $I_i \propto \sqrt{N_*/n}$ with a random sign. Let us determine the sum $a_n = \sum I_i$. It has n random terms I. Consequently, $a_n \sim \sqrt{n}\, I$. Substituting $I \propto \sqrt{N_*/n}$ into this expression, we obtain $a_n \propto \sqrt{N_*}$. It can be seen that the dependence on n in spectrum $|f_n^*|$ asymptotically vanishes, and the left wing assumes the form of a plateau. This explains the slow decrease in $|f_n|$ in the interval $n \in [0, N]$ in Fig. 13c and the finite values of $|f_n|$ for $n = 0$ and $n \sim 1$ in Fig. 14b.

The possible asymptotic forms of $n \ll N$ are bounded by two limiting cases. In the first case, we have

$$|f_n| \sim \sqrt{N}\,\bar{f}\bar{\lambda}, \quad \bar{f} = \sqrt{\langle (f - \langle f\rangle_x)^2\rangle_x}, \tag{36}$$

where $\langle\ldots\rangle_x$ denotes the averaging over x. This is the plateau. In the second case, we have

$$|f_n| \sim \sqrt{n}\,\bar{f}\bar{\lambda}. \tag{37}$$

The realization of a certain version depends on the behavior of the cumulative integral

$$F(x) = (x\bar{f})^{-1}\int_0^x f(x')dx'. \tag{38}$$

For a random periodic function of the general form, we have $|F(x)| \sim \sqrt{x}$. In this case, the long-wave asymptotic form has the form of plateau (36). If, however, $|F(x)| \sim 1$, the asymptotic form (37) is realized. For example, for the vertical velocity w, integral (38), which has the meaning of the volume flux, is bounded due to the conservation of the volumes of an incompressible liquid during transpositions on finite segments in x.

It can be seen that the spectral trace of a random periodic function, which is apparently localized on scales $\sim\bar{\lambda}$, extends far to the long-wave region due to stochastic processes. In order to estimate the dynamic role of this effect, we must compare the long-wave amplitude with the limiting saturated or majorizing amplitudes. For example, in the case of velocity, the spectral amplitude of saturation is given by

$$\left(|w_n^{\mathrm{sat}}| \sim \frac{F\sqrt{1-\mu}}{n}\right), \tag{39}$$

(see Section 12). It is determined by nonlinear and random processes. We write the ratio of velocities $|wn| \sim \sqrt{n}\,(F\sqrt{1-\mu}/\sqrt{N})2\pi/N$ (37) and (39), where N is the number of pairs of columns or the number of bubbles on segment $L_\perp = 2\pi$: $|w_n|/|w_n^{\mathrm{sat}}| \sim (n/N)^{3/2}$. The results of numerical simulation carried out by us show that the perturbations whose amplitudes are of the order of or smaller than a few percent of the saturation amplitude are insignificant. This leads to a rough estimate of the largest significant scales $n_{lw} \sim N/5$ generated by straggling.

The case of a simple random periodic function was considered by us earlier. Such a function is characterized by a single cell of scale $\bar{\lambda}$ since the neighboring cells are of the same order of magnitude (dispersion $\Delta\lambda \sim \bar{\lambda}$). The average cell has no substructure (see Fig. 13a). In the case of turbulent mixing, the cells, first, are nonequivalent (clusters) and, second, have an internal substructure. It would be interesting to analyze the changes in the correlator and in the spectrum associated with these circumstances.

Let us consider the example in which a function is calculated on subinterval $[x_{n-1}, x_{n+1}]$ by the formula

$$\begin{aligned} f(x) = {} & [(x-x_{2n-1})(x-x_{2n})(x-x_{2n+1})]^2 \\ & \times\left[x-\left(x_{2n-1}+\frac{x_{2n}-x_{2n-1}}{3}\right)\right] \\ & \times\left[x-\left(x_{2n-1}+2\frac{x_{2n}-x_{2n-1}}{3}\right)\right] \\ & \times\left[x-\left(x_{2n}+\frac{x_{2n+1}-x_{2n}}{3}\right)\right] \\ & \times\left[x-\left(x_{2n}+2\frac{x_{2n+1}-x_{2n}}{3}\right)\right], \end{aligned} \tag{40}$$

where sequence x_j is determined by (35) and index $n = 1, \ldots, N$ labels cells or pairs of columns. The function $f(x)$ in Fig. 13a is of the order of $\bar{\lambda}^3$, where $\bar{\lambda} = 2\pi/N \ll 1$ ($N = 100$ in our examples). Function (40) is of the order of $\bar{\lambda}^{10} \sim f_1^{10/3}$. Raising small quantities to a power significantly increases their straggling. Most cells turn out to be "weak" (characterized by small values of f); they occasionally alternate with "strong" cells in which the value of f is considerably larger than in weak cells (see Fig. 15a). Such a behavior resembles Figs. 2a and 2b.

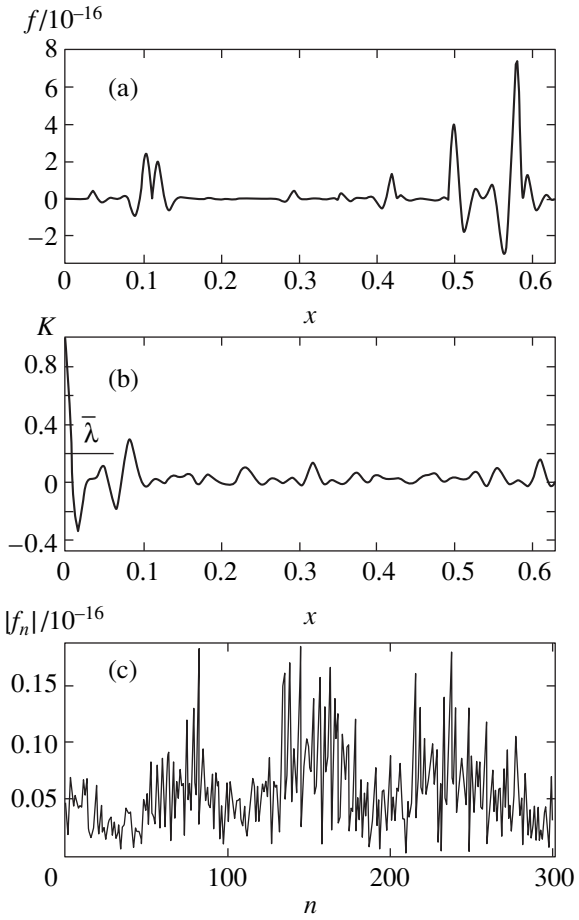


Fig. 15. A cluster-type random periodic function. Example (40), the set of ξ_j is the same as in Fig. 13 (β= 0.6). The number of cells N = 100. The cells differ significantly: 4–5 "weak" cells correspond to a "strong" cell; (b) correlator $K(x)$, (c) spectrum $|f_n|$.

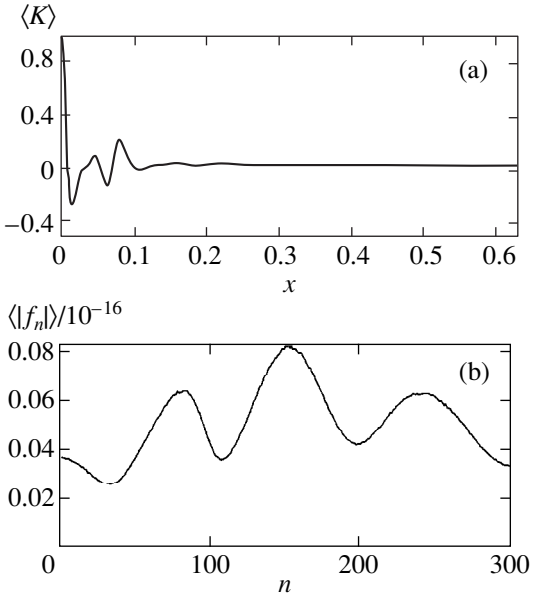


Fig. 16. (a) Correlator $\langle K(x) \rangle$ and (b) spectrum $\langle |f_n| \rangle$ averaged over 10^4 realizations in the cluster case. The division by three in function (40), which is manifested weakly on the fluctuating curve in Fig. 15c, is clearly seen on the smoothed spectrum in Fig. 16b.

It is interesting to note that the cluster structure is manifested insignificantly in the correlator (Fig. 15b). The group of first significant spikes in it has a width approximately equal to $\bar{\lambda}$, which is much smaller than the average separation between "strong" cells (cf. Figs. 15a and 15b). This group is associated with the internal structure of the cell. Because of the substructure, the spectrum in Fig. 15c is strongly broadened towards the short-scale region to the right of the characteristic wave number $\bar{n} = 2\pi/\bar{\lambda} = N = 100$. Clusters are manifested in a certain enhancement of the long-wave wing. Analyzing the partial realization (Fig. 15) and the averaged data (Fig. 16) on the whole, we conclude that the spectrum of function (40) is quite broad and is formed by the scales differing considerably.

27. SMALLNESS OF α_+ AND STIMULATION OF MIXING BY LONG-WAVE PERTURBATIONS

In the previous sections, we analyzed mixing under ideal conditions, when viscosity $\nu = +0$, the surface tension and diffusion coefficients are $\sigma = +0$ and $D = +0$, respectively, and the stabilization scale $\lambda_{micr} = +0$. Under these conditions, the boundary-value problem is incorrect in the Hadamard sense (see [43, 4] and [44], Section 8)[9], since the increment is unbounded, $\gamma \propto \sqrt{k} \longrightarrow \infty$ for $k \longrightarrow \infty$. In this case, turbulent mixing emerges "at once" (starting from $t = +0$) and is self-similar (see formulas (1), (5)) straightaway. It is natural to refer to such a mixing as spontaneous since for $t > 0$, there is no information concerning the initial perturbation. Each realization is unique.

The question arises: what can be "seen" against the background of spontaneous mixing? In order to answer this question, it was assumes implicitly that perturbations are infinitesimal for $t = 0$. It is important that the numerical value of the coefficient α_+ of spontaneous mixing is quite small (see Sections 10 and 22) and the initial integrity of the heavy substance is destroyed relatively slowly. If the breakdown of continuity occurred with the Archimedean (i.e., limiting) acceleration (in this case, $G^+_{ss} g \approx (1 - \mu)g$, see Section 10), it would be impossible to change it (i.e., to accelerate still further) using small finite-amplitude perturbations. It is due to the smallness of coefficient α_+ that a finite (in wavelengths and amplitudes) perturbation can manifest itself against the background of spontaneous mixing.

[9] In this publication, the conditions under which a smooth one-mode solution with period λ can exist are considered. It is required, first, that ratio λ_{micr}/λ be small (ideal hydrodynamics) and, second, that noise be suppressed exponentially down to scale λ_{micr}. The latter condition is satisfied if ratio λ_{micr}/λ is finite.

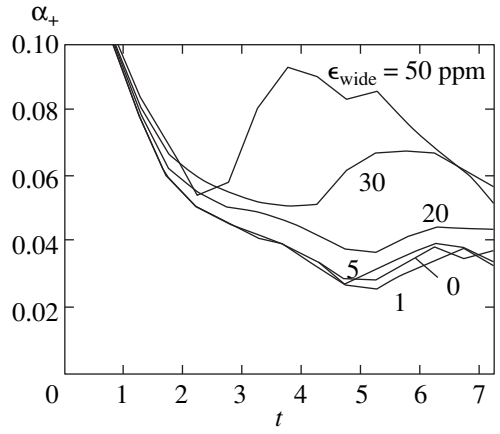


Fig. 17. Dependence of the evolution of coefficient α_+ on the amplitude ϵ_{wide} of wide-range noise indicated on the curves in pro mille.

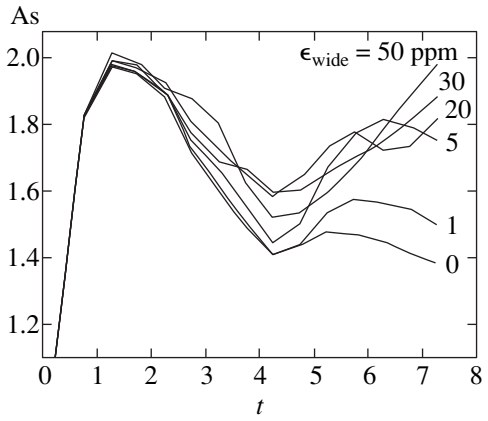


Fig. 18. Dependence of the evolution of the asymmetry coefficient As on the amplitude ϵ_{wide} of wide-range noise indicated on the curves in pro mille.

Let us consider two cases. In one case, perturbations are periodic (one wavelength λ), while in the other, perturbations have a logarithmically wide range (a set of strongly differing wavelengths λ, e.g., the mixture of waves of the millimeter, centimeter, decimeter, etc. ranges). It should be emphasized that perturbations are finite. This means that there exists a threshold. Its existence has a clear meaning since we are speaking of the competition of a perturbation with spontaneous mixing, viz., a flow with finite velocities and displacements. The perturbation can be either noticeable (superthreshold amplitude) or unnoticeable (subthreshold amplitude) against the background of the spontaneous flow. The threshold amplitudes in the case of periodic perturbations were considered in [45, 46]. In the superthreshold mode, a periodic perturbation is "coated" with short-scale modes associated with spontaneous turbulence "emerging" from the short-wave region (see [4], Section 18, and [47, 48]). The period λ of a periodic perturbation determines the characteristic time $t_\lambda = (\mathrm{At}gk)^{-1/2}$. For $t \gg t_\lambda$, the trace of this perturbation is healed and the mixing returns to the spontaneous mode. Let us consider the wide-range case.[10] A perturbation uniform in the wave numbers k is most interesting for our analysis. It does not introduce a preferred scale and, hence, preserves the quadratic self-similarity (1) and (5). It should be emphasized that this indicates the conservation of the self-similarity exponent or index. The self-similar functions $\hat{f}(\hat{x}, \hat{z})$ (10) and $\tilde{f}_{\hat{n}}$ (25), (26) of self-similar variables for a considerable amplitude of the perturbation exceeding the threshold vary depending on the dimensionless parameter determining the excess over the threshold value. Superthreshold wide-range perturbations cannot be healed.

Let us present uniform wide-range velocity perturbations convenient for numerical simulation. We write the initial perturbations in the form of an expansion in harmonics (3). Let us suppose that

$$a_n^0 = \frac{F\sqrt{1-\mu}}{n}(2\xi_n - 1)\epsilon_{wide},$$
$$b_n^0 = \frac{F\sqrt{1-\mu}}{n}(2\xi_n' - 1)\epsilon_{wide}, \quad (41)$$

where ξ_n is the random number generator (35), and ξ_n and ξ_n' are independent. It can be seen that the initial field (3), (41) is formed by three components. First, it is the saturation velocity (39), second, the random operator, and third, the dimensionless parameter ϵ_{wide}. This field has no preferred scale. It is defined by only one parameter, viz., amplitude ϵ_{wide} of random wide-range noise.

Figures 17–19 summarize the main results of calculations of numerous versions, made using the method of large particles (see Section 5), N_x = 200 to 400, μ = 1/10. The initial perturbation was presented in the form of the sum of short-range (see Section 6) and wide-range (41) perturbations. Coefficient α_+ was determined in accordance with the rule $\alpha_+ = [h_+(t_{i+1}) - h_+(t_i)]/\mathrm{At}(t_{i+1}^2 - t_i^2)$ over segments $t_{i+1} - t_i$ = 0.5 to 2. The results of calculations indicate the existence of a threshold $(\epsilon_{wide})_{thr}$ of approximately 10 pro mille. This is a small amplitude amounting approximately to one percent of the saturation velocity (39) for which the

[10] A theory of stimulation was proposed in [22, 23, 3, 4]. In these publications, first, the expansion of noise in a dense series of random harmonics (41) was proposed, which makes it possible to define the noise amplitude on a unified basis. (In [3, 4], noise was determined not only by the amplitude, but also by the choice of a certain lacunar expansion. It was possible to compare only noises with different amplitudes but with the same lacunar expansion.) Second, the results of systematic numerical calculations were considered, the threshold was determined, and the superthreshold dynamics was studied.

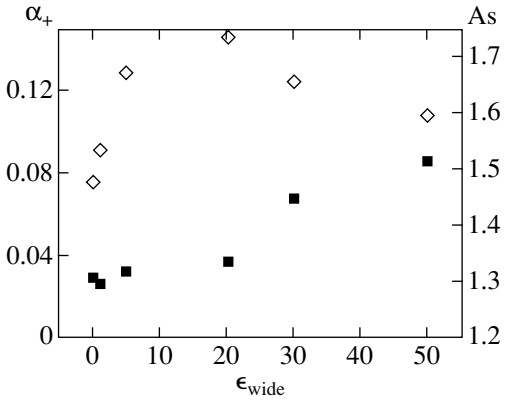


Fig. 19. Dependence of coefficients $\alpha_+(t = 5)$ (■) and $As(t = 5)$ (◇) on the amplitude ϵ_{wide} in pro mille. For weak noise, the flow does not differ from a spontaneous flow. The effect of noise becomes significant for $(\epsilon_{wide})_{thr} \approx 10^{-2}$.

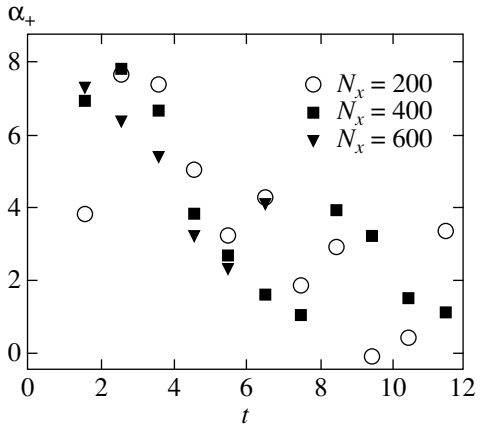


Fig. 20. Coefficient α_+ in the absence of stimulation. The noticeable decrease in the value of α_+ for $t \approx 7$ is apparently associated with an early effect of the lateral boundary conditions. The net-characteristic method, $\mu = 1/10$, N_x is the number of nodes along the horizontal.

nonlinearity becomes significant. Consequently, near-threshold perturbations are linear. The smallness of the threshold is associated with the smallness of coefficient α_+.

For amplitudes smaller than the threshold values, mixing occurs in the spontaneous mode and the effect of noise can be disregarded. For superthreshold amplitudes, the flow is rearranged. The spectral bump is displaced towards longer wavelengths. This reflects the enhancement of subharmonics. Ratio R_{HV} increases. The enhancement of the long-wave wing is associated with the intensification of mixing which is manifested in an increase in coefficient α_+. In a spontaneous process, the origin of subharmonics is associated with the central bump in the spectrum, while in the case of stimulations, a fraction of subharmonics is "extracted" from wide-range noises which are present from the very outset.

The threshold amplitude is significant. The threshold value is noticeable. If we consider ideal conditions ($\lambda_{micr} = +0$, …), the expensive decrease in the noise amplitude due to technological improvements (such as polishing) below the threshold amplitude is not expedient (limiting cleansing).

28. EFFECT OF LATERAL BOUNDARY CONDITIONS AND CHANGE OF SELF-SIMILAR ASYMPTOTIC FORM

We have studied spontaneous and stimulated mixing. The case when exchange processes occur under free conditions (in the sense that external geometrical boundaries are far away, $\bar{\lambda} \ll L_\perp, h_{up}, h_{down}$; see Section 5) has been considered. Scale $\bar{\lambda}$ increases with time (see Figs. 1a and 1b) in accordance with formula (29) with coefficient α_H and, hence, approaches the external scales. The horizontal (h_{up}, h_{down}, top, and bottom) and the lateral ($L_\perp$) boundaries slow down the exchange. The effect of horizontal boundaries was considered in [5, 40].[11] For the "box" geometry chosen in our calculations (see Section 5), lateral boundaries are the first to appear (restriction of the flow by a "bottleneck", see Sections 4, 7, and 20).

Figure 20 illustrates the behavior of coefficient α_+ in the spontaneous case. This case also includes the curves with a low noise in Fig. 17, $\epsilon_{wide} = 0$, 1, 5 pro mille. Noticeable fluctuations and the decrease in α_+ for $t \approx 7$ (calculations are made with $\mu = 1/10$) are probably associated with the beginning of the decelerating effect of the bottleneck, although the values of $h_+(t = 7)/L_\perp = 1/5$ to $1/3$ appear as still quite small. In order to estimate the effect of lateral boundaries, we must take the ratio $\lambda_{dyn}/L_\perp$, where λ_{dyn} is the most developed scale in the amplitude spectrum. In the 2D case, the ratio $R_{HV} = \lambda_{dyn}/h_+$ is quite large (see Sections 23 and 24) and,

[11] The mixing of a layer having the initial thickness h_0 at the upper or lower boundary was considered. In small time intervals, the condition $t \ll t_{trans} = \sqrt{h_0/g}/\sqrt{At\alpha_\pm}$ is satisfied, $\bar{\lambda} \approx h_\pm \propto t^2$, where the plus or minus sign is used depending on the boundary at which the layer is located. For $t \sim t_{trans}$, the rate of expansion of the layer tends to the asymptotic value $\dot{h} \sim \sqrt{gh\Delta\rho/\rho_{out}} \approx \sqrt{gh(\rho_{in}h_0/h)/\rho_{out}}$, where $h(t)$ is the running thickness of the layer, ρ_{in} and ρ_{out} are the densities of the layer (initial density) and of the medium in which it is located. In connection with what has been said above, we can mention a close self-similar solution of the problem on nonstationary convection near the boundary of a medium with neutral buoyancy [49] (the solution is sought in the turbulent diffusion approximation). This solution is important for geophysics (temperature wave generated when the mechanism of cooling at the sea surface is "engaged").

hence, the effect of the lateral boundaries is manifested at an early stage.

In boxes strongly elongated along the vertical (see Section 5 and [7]), the change in the self-similar mode follows the transient process at $t_{trans} = \sqrt{L_\perp/g}/\sqrt{\mathrm{At}\alpha_+}$. Let us consider the asymptotic form $h = h_+ + h_- \gg L_\perp$ using the methods of the theory of the turbulent mixing length [5, 6]. We write the following expression for the turbulent velocity:

$$w_{turb} = \sqrt{g l_{turb}\left(l_{turb}\frac{\partial \ln\bar{\rho}}{\partial z}\right)}, \quad (42)$$

where l_{turb} is the size of the largest vortices determining the turbulent transfer, the combination in the parentheses plays the role of the local Atwood number, and $\bar{\rho}$ is the density averaged over the horizontal. In the asymptotic form under investigation, l_{turb} in (42) is determined by the transverse dimension of the box or the 2D, 3D tube:

$$l_{turb} = L_\perp. \quad (43)$$

For this reason, the turbulent diffusion coefficient assumes the form

$$D_{turb} = \frac{1}{3} l_{turb} w_{turb} = \frac{1}{3}\sqrt{g} l_{turb}^2 \sqrt{\bar{\rho}_z'/\bar{\rho}}.$$

Accordingly, the diffusion equation can be written in the form

$$\frac{\partial\bar{\rho}}{\partial t} = \frac{1}{3} g^{1/2} l_{turb}^2 \frac{\partial}{\partial z}\left[\frac{1}{\bar{\rho}^{1/2}}\left(\frac{\partial\bar{\rho}}{\partial z}\right)^{3/2}\right]. \quad (44)$$

In the problem on the breakdown of the initial density jump, there are no parameters with dimensions $[l]$ or $[t]$. Consequently, the solution of Eq. (44) must be self-similar:

$$\bar{\rho}(z,t) = \bar{\rho}(\xi), \quad \xi = \mathrm{const}\, z/t^\delta. \quad (45)$$

Substituting relations (45) into Eq. (44), we obtain

$$-\xi\bar{\rho}_\xi' = \frac{\mathrm{const}^{5/2} g^{1/2}}{3\delta}\frac{l_{turb}^2}{t^{5\delta/2-1}}\frac{d}{d\xi}\left[\frac{(\bar{\rho}_\xi')^{3/2}}{\bar{\rho}^{1/2}}\right]. \quad (46)$$

This equation becomes an ordinary second-order differential equation if factor $l_{turb}^2/t^{5\delta/2-1}$ is constant. In the case of free turbulence considered in this way in [5], we have $l_{turb} \propto h \propto t^2$, whence $\delta = 2$ (see Sections 10, 11, 14). In the case under investigation of a long tube, scale l_{turb} (43) is fixed. Consequently, the law of expansion of the mixed zone in such 2D, 3D tubes is determined by a power dependence of the form

$$h = L_\perp(t/t_{trans})^{2/5} = C(\alpha_+\mathrm{At}g)^{1/5} L_\perp^{4/5} t^{2/5}, \quad (47)$$

where $C \sim 1$ is a dimensionless constant.

The vertical self-similar profile of the average density can be found by integrating Eq. (46). It can be seen that this equation is the same for free and restricted turbulence. An analysis of this equation is given in [5]. The profile has the shape of a smoothed step.[12] Thus, the quadratic dependence (1), (5) for $t \sim t_{trans}$ is replaced by law (47) with attenuating acceleration. We can formally write

$$h_+(t) = \alpha_+(t)\mathrm{At}g t^2 = C_1(\alpha_+^{free}\mathrm{At}g)^{1/5} L_\perp^{4/5} t^{2/5},$$

where α_+^{free} is the value in the free case. This gives

$$\alpha_+(t) = C_1(\alpha_+^{free})^{1/5}(L_\perp/\mathrm{At}g)^{4/5} t^{-1.6} = C_2\alpha_+^{free}(t/t_{trans})^{-1.6}.$$

Coefficient $\alpha_+(t)$ is a decreasing function. This is a manifestation of the drag associated with lateral boundaries.

29. CONCLUSION

We have investigated the Rayleigh–Taylor turbulence. The rate of vertical expansion of the mixing layer is determined by the transverse structure of the layer. The layer is formed by a horizontal row of basic cells or clusters undergoing continuous changes. The time evolution is manifested in the enhancement of long-wave fluctuations leading to the generation of large-scale fluctuations relative to current mass flows. Above and under the main row, fragments of the previous generations of the cascade are moving. We analyzed the substructure of a basic cell. In the 2D case, the cell is formed by several pairs of ascending and descending columns (cluster) rather than by a single pair.

We have analyzed the factors affecting the value of coefficient α_+. The initial long-wave noise increases α_+ by stimulating the mixing. The lateral boundaries of the rated region produce the opposite effect. They restrict the horizontal size of clusters, hamper the mixing, and reduce the value of α_+. We proposed the important formula (41) describing wide-range noise. The threshold amplitude of noise affecting the value of α_+ is determined. The problem of the asymptotic form of the flow over large time intervals is solved for extended rated regions. The onset of the rearrangement of the flow with a free asymptotic form, $h \propto t^2$, to a restricted form, $h \propto t^{2/5}$, reduces coefficient α_+.

The information on the structures is contained in the spectrum. It is important to process numerical data appropriately, first, to obtain the universal (self-similar) spectra and, second, to separate the random component associates with stochastic phases from the systematic component. These problems were solved by us here.

[12] It should be noted that for $\mu = +0$ (mixing with a "vacuum"), we have $h_+ \propto t$, $h_- \propto t^2$. In this special case, law (47) is inapplicable.

First, we proposed a self-similar reduction of the spectral amplitudes (26) to the dimensionless form, which takes into account the randomness of the phases of adjacent harmonics. Second, we determined the smoothing procedure for spectral data in order to single out the systematic component. The effectiveness of this procedure is demonstrated on the model of random periodic functions.

ACKNOWLEDGMENTS

The authors are grateful to S.I. Anisimov and O.M. Belotserkovskiĭ for fruitful discussions. This work was supported financially by the Russian Foundation for Basic Research (projects nos. 99-02-16666 and 00-01-00250).

REFERENCES

1. D. H. Sharp, Physica D (Amsterdam) **12**, 3 (1984).
2. H.-J. Kull, Phys. Rep. **206**, 197 (1991).
3. N. A. Inogamov, Astrophys. Space Phys. Rev. **10** (2), 1 (1999).
4. N. A. Inogamov, A. Yu. Dem'yanov, and É. E. Son, *Hydrodynamics of Mixing* (Mosk. Fiz.-Tekh. Inst., Moscow, 1999).
5. S. Z. Belen'kiĭ and E. S. Fradkin, Tr. Fiz. Inst. Akad. Nauk SSSR **29**, 207 (1965).
6. V. E. Neuvazhaev, Prikl. Mekh. Tekh. Fiz., No. 6, 82 (1976).
7. http://esd.gsfc.nasa.gov/ESS/annual.

   reports/ess95contents/app.inhouse.fryxell.html.
8. O. M. Belotserkovskiĭ, *Numerical Simulation in Mechanics of Continuous Media* (Fizmatlit, Moscow, 1994).
9. O. M. Belotserkovskiĭ, V. A. Gushchin, and V. N. Kon'shin, Zh. Vychisl. Mat. Mat. Fiz. **27**, 594 (1987).
10. O. M. Belotserkovskiĭ and A. M. Oparin, *Numerical Experiment in Turbulence: from Order to Chaos* (Nauka, Moscow, 2000).
11. N. A. Inogamov and A. M. Oparin, Zh. Éksp. Teor. Fiz. **116**, 908 (1999) [JETP **89**, 481 (1999)].
12. *Proceedings of the 6th International Workshop on the Physics of Compressible Turbulent Mixing*, Ed. by G. Jourdan and L. Houas (Institut Universitaire des Systemes Thermiques Industriels, Marseille, 1997).
13. S. W. Haan, Phys. Rev. A **39** (11), 5812 (1989).
14. S. W. Haan, Phys. Fluids B **3** (8), 2349 (1991).
15. R. D. Richtmyer, Commun. Pure Appl. Math. **13** (2), 297 (1960).
16. E. E. Meshkov, Izv. Akad. Nauk SSSR, Mekh. Zhidk. Gaza, No. 5, 151 (1969).
17. S. G. Zaĭtsev, E. V. Lazareva, V. V. Chernukha, and V. M. Belyaev, Dokl. Akad. Nauk SSSR **283** (1), 94 (1985) [Sov. Phys. Dokl. **30**, 579 (1985)].
18. J. F. Haas and B. Sturtevant, J. Fluid Mech. **181**, 41 (1987).
19. V. B. Rozanov, I. G. Lebo, S. G. Zaĭtsev, *et al.*, Preprint No. 56, FIAN (Lebedev Institute of Physics, Academy of Sciences of USSR, Moscow, 1990).
20. U. Alon, J. Hecht, D. Ofer, and D. Shvarts, Phys. Rev. Lett. **74**, 534 (1995).
21. *Mathematical Encyclopedia*, Ed. by I. M. Vinogradov (Sovetskaya Éntsiklopediya, Moscow, 1979).
22. S. I. Anisimov, Ya. B. Zel'dovich, N. A. Inogamov, and M. F. Ivanov, in *Shock Waves, Explosions and Detonation*, Ed. by M. Summerfield (AIAA, Washington, 1983), Vol. 87, p. 218.
23. N. A. Inogamov, Pis'ma Zh. Tekh. Fiz. **4** (12), 743 (1978) [Sov. Tech. Phys. Lett. **4**, 299 (1978)].
24. A. M. Prokhorov, S. I. Anisimov, and P. P. Pashinin, Usp. Fiz. Nauk **119** (3), 401 (1976) [Sov. Phys. Usp. **19**, 547 (1976)].
25. D. T. Dumitrescu, Z. Angew. Math. Mech. **23** (3), 139 (1943).
26. R. M. Davies and G. I. Taylor, Proc. R. Soc. London, Ser. A **200**, 375 (1950).
27. P. R. Garabedian, Proc. R. Soc. London, Ser. A **241** (1226), 423 (1957).
28. G. Birkhoff and D. Carter, J. Math. Mech. **6** (6), 769 (1957).
29. J.-M. Vanden-Broeck, Phys. Fluids **27** (5), 1090 (1984).
30. S. I. Anisimov, A. V. Chekhlov, A. Yu. Dem'yanov, and N. A. Inogamov, Russ. J. Comput. Mech. **1**, 5 (1993).
31. G. Birkhoff and E. H. Zarantonello, *Jets, Wakes and Cavities* (Academic, New York, 1957; Mir, Moscow, 1964).
32. M. A. Lavrent'ev and B. V. Shabat, *Problems of Hydrodynamics and Their Mathematical Models* (Nauka, Moscow, 1970).
33. S. F. Garanin, Vopr. At. Nauki Tekh., Ser.: Teor. Prikl. Fiz., No. 3/1, 12 (1995). (S. F. Garanin, in *Proceedings of the 5th International Workshop Compressible Turbulent Mixing*, Ed. by D. Youngs, J. Glimm, and B. Boston (World Scientific, Singapore, 1995).)
34. P. Wilson, M. Andrews, and F. Harlow, Phys. Fluids **11** (8), 2425 (1999).
35. M. B. Schneider, G. Dimonte, and B. Remington, Phys. Rev. Lett. **80** (16), 3507 (1998).
36. D. L. Youngs, Phys. Fluids A **3** (5), 1312 (1991).
37. K. I. Read, Physica D (Amsterdam) **12** (1–3), 45 (1984).
38. Yu. V. Yanilkin, Vopr. At. Nauki Tekh., Ser.: Mat. Model. Fiz. Protsessov, No. 4, 88 (1999).

39. J. Glimm, J. W. Grove, X.-L. Li, *et al.*, SIAM J. Sci. Comput. (USA) **19**, 703 (1998).

40. N. N. Anuchina, N. S. Es'kov, A. V. Polionov, *et al.*, in *Proceedings of the 6th International Workshop on the Physics of Compressible Turbulent Mixing*, Ed. by G. Jourdan and L. Houas (Institut Universitaireles Systemes Thermiques Industriels, Marseille, 1997).

41. M. D. Kamchibekov, E. E. Meshkov, N. V. Nevmerzhitskiĭ, and E. A. Sotskov, Preprint No. 46-96, RFYaTs, VNIIÉF (Sarov, 1996).

42. X. He, R. Zhang, S. Chen, and G. D. Doolen, Phys. Fluids **11** (5), 1143 (1999).

43. G. Birkhoff, in *Proceedings of Symposia in Applied Mathematics*, Vol. XIII: *Hydrodynamic Instability*, Ed. by G. Birkhoff, R. Bellman, and C. C. Lin (American Mathematical Society, 1962; Mir, Moscow, 1964).

44. N. A. Inogamov, Zh. Éksp. Teor. Fiz. **107** (5), 1596 (1995) [JETP **80**, 890 (1995)].

45. D. L. Youngs, Physica D (Amsterdam) **12** (1–3), 32 (1984).

46. P. F. Linden, J. M. Redondo, and D. L. Youngs, J. Fluid Mech. **265**, 97 (1994).

47. D. Ofer, D. Shvarts, Z. Zinamon, and S. A. Orszag, Phys. Fluids B **4**, 3549 (1992).

48. D. Ofer, U. Alon, D. Shvarts, *et al.*, Phys. Plasmas **3** (8), 3073 (1996).

49. S. I. Voropayev and Y. D. Afanasyev, Phys. Fluids A **5** (10), 2461 (1993).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 92, No. 4, 2001, pp. 744–751.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 119, No. 4, 2001, pp. 853–861.

MISCELLANEOUS

Suppression of Dynamic Chaos in Hamiltonian Systems

V. V. Vecheslavov

Budker Institute of Nuclear Physics, Siberian Division, Russian Academy of Sciences,
pr. Akademika Lavrent'eva 11, Novosibirsk, 630090 Russia
e-mail: vecheslavov@inp.nsk.ru

Received June 26, 2000

Abstract—The paper describes the results of a recent numerical study on the canonical mapping with a sawtooth force. The dynamic effects of the formation of invariant resonance structures of various orders, whose presence prevents the development of global chaos and restricts momentum diffusion in the phase space, are discussed. The dynamic situation near an integer resonance separatrix in the neighborhood of the critical state is studied, and the conditions responsible for the stability of this separatrix in the critical state are determined. Along with the mapping, the related continuous Hamiltonian system is considered. For this system, the separatrix mapping and the Mel'nikov–Arnold integral are introduced, whose analysis facilitates understanding the reasons responsible for the unusual dynamics. This dynamics is shown to be preserved under substantial saw shape changes. Relevant new problems and open questions are formulated. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

According to the modern theory of nonlinear Hamiltonian systems, the arising of dynamic chaos is related to the destruction of resonance separatrices and the formation of chaotic layers in their place [1–3]. The separatrix of a solitary nonlinear resonance is known to be two spatially coinciding branches corresponding to forward and backward motions along the time axis. Both branches are continuous trajectories with an infinite period of motion, which issue out of the unstable equilibrium point (saddle) and then asymptotically approach it. In the presence of other nonlinear resonances in the system, they split and, starting as previously from the saddle toward each other, never return to the saddle point again. Free ends of split separatrix branches form an infinite number of loops of an infinitely increasing length, which fill the narrow region near the unperturbed separatrix and produce a chaotic layer [1–3]. The overlapping of the chaotic layers of all system resonances is exactly the arising of global chaos.

The assertion that resonance separatrix splitting and the formation of a chaotic layer in its place in a typical (that is, nonintegrable) Hamiltonian system occurs under almost every perturbation can be considered central to modern views on the problem. It is also believed that the destruction of separatrices occurs, first of all, because they have zero frequencies and the interaction of nonlinear resonances in their vicinity is always substantial [1–3]. In this "usual" dynamics, the transition form a chaotic layer to a chaotic sea caused by an increase in the perturbation parameter has been studied rather in detail and is explained by the destruction of invariant curves with irrational rotation numbers and the formation of so-called Cantor tori [4].

The conditions of global chaos formation are different for dynamic systems with analytic and smooth phase dependences of perturbation. The principal question is that of the degree of smoothness (the number l of continuous derivatives of force) at which global chaos arises starting with a certain threshold rather than an arbitrarily small system parameter value. Moser and Russman [5] showed that such a threshold always exists if $l > 3$. A proof of the reverse assertion that, at $l \le 3$, there is no threshold and no restrictions on momentum diffusion in the phase space is lacking.

Recently, Ovsyannikov [6] has proved that, for a quite definite countable set of critical system parameter K values, there exist undestroyed integer resonance separatrices in a two-dimensional canonical mapping of the form

$$\bar{p} = p + Kf(x), \quad \bar{x} = x + \bar{p} \pmod 1 \qquad (1)$$

with periodic $f(x + 1) = f(x)$ and antisymmetric $f(-x) = -f(x)$ sawtooth forces (degree of smoothness $l = 0$)

$$f(x) = \begin{cases} 4x & \text{at } x \le 0.25, \\ 4(0.5 - x) & \text{for } |0.5 - x| \le 0.25, \\ 4(x - 1) & \text{for } |x - 1| \le 0.25. \end{cases} \qquad (2)$$

This important result is still unpublished; for the reader's convenience, communication [6] is cited here in appendix form.

The Ovsyannikov theorem prompted us to undertake thorough numerical and theoretical studies of system (1), (2) and its modifications [7, 8]. We found that not only integer but also fractional resonances of arbitrary orders have critical numbers of their own at which their separatrices also do not split. A brief review of our numerical results is given in Section 2. Unfortunately, the important and interesting work by Bullet [9], in

which the same model was studied, was not known to the authors of [6–8]. For this reason, Section 2 also includes the most important and strictly proven results from [9]. The differences between the approach developed by Bullet and that used in our works are discussed.

Section 3 considers the dynamic situation in the vicinity of an integer resonance separatrix near critical conditions. Certain circumstances ensuring the surprising stability of this separatrix in the critical state are specified.

In addition to mappings (1), (2), we consider a related continuous system with the Hamiltonian

$$\begin{aligned} H(x, p, t) &= H_0(x, p) + \epsilon U(x, t), \\ h_0 &= p^2/2 + \omega_0^2 V(x), \\ U(x, t) &= \cos(2\pi x - \tau - \tau_0) + \cos(2\pi x + \tau + \tau_0), \\ \tau &= \Omega t, \end{aligned} \tag{3}$$

where $V(x) = -\int f(x)dx$ is the sawtooth potential, $\epsilon \ll 1$, and the perturbation contains two symmetrical harmonics. The study of this system performed in Section 4 facilitates understanding some details of the behavior of mapping (1), (2).

In Section 5, a general asymmetric saw, whose particular case is symmetrical saw (2), is considered. We show that the unusual dynamics is preserved in the whole range of possible skewness parameter variations and only disappears at limit points (where the two-tooth saw transforms into one-tooth). In the Conclusion, we summarize the results and discuss open questions.

2. CRITICAL NUMBERS AND CRITICAL STATES OF SYSTEM (1), (2)

A periodic orbit and the corresponding resonance are usually denoted by the rotation number $\nu = P/Q$, where Q is the number of mapping iterations to P orbit periods [2]. Resonances with $Q = 1$ are called integer, and the other resonances, fractional. Unperturbed resonance separatrices have $\nu = 0$, and they are precisely the objects of our study.

One of the important features of the dynamics under consideration is the presence, for resonance Q of an arbitrary order, of a presumably countable set of "critical" parameter $K_{Q,m}$ ($m = 1, 2, \ldots$) values of its own. At these values, the resonance separatrix not only does not split (in spite of the presence of strong local chaos, see Fig. 1) but it is even an impermeable barrier to the other trajectories. It extends through the whole phase and isolates the interior of its resonance (a region with a finite phase volume) from the remaining part of the phase space. This invariant manifold is known as an invariant resonance structure [7]. Like the presence of invariant curves, the presence of such structures completely forbids global chaos and imposes restrictions on momentum diffusion.

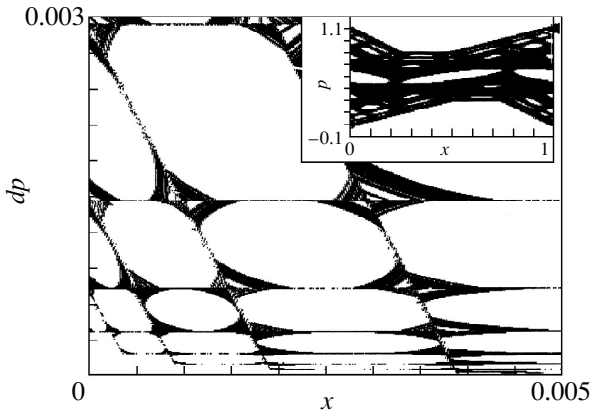


Fig. 1. The system parameter equals the second critical number $K = K_{1,2} = 1/8$. A strongly enlarged (on the momentum scale) small "gap" region between a chaotic trajectory with the initial coordinates $x = 0$, $p = 0.37$ and the lower integer resonance separatrix (see text). Here, $dp = p - p_s$, where $p(x)$ is the trajectory momentum and $p_s(x)$ is the momentum on the separatrix. In the inset: the region occupied by this trajectory. The number of mapping iterations 10^9.

Critical conditions were found by the technique developed by us earlier and based on measuring separatrix intersection angles α_Q at the central homoclinic point (the details and results of such measurements for the standard mapping are described in [10]).

The numerical $\alpha_Q(K)$ dependences obtained for all system (1), (2) resonances studied in this work are alternating and oscillating [7], which qualitatively distinguishes them from a similar dependence for the standard mapping, which has been studied fairly thoroughly (this dependence is alternating and strictly monotonic [10]). This difference is related to substantially different dynamic behaviors of these two systems.

The $K_{Q,m}$ ($m = 1, 2, \ldots$) critical numbers of a resonance of order Q satisfy the condition $\alpha_Q(K_{Q,m}) = 0$ and can be of two types: at odd m, the angle smoothly changes while passing zero, and, at even m, it passes zero in a jump.

An analysis shows that even critical numbers of system (1), (2) integer resonances exactly coincide with the elements of the countable set from the Ovsyannikov theorem and are determined by solutions to transcendent equation (1.4) at integer coefficient k values (see Appendix). We found that odd critical numbers not included in the Ovsyannikov theorem were also determined by solutions to the same equation, however, with half-integral coefficient values. For an arbitrary integer resonance critical number we therefore have

$$K_{1,m} = \sin^2(\beta_m/2), \quad m = 1, 2\ldots, \tag{4}$$

Critical numbers of fractional resonances in the "gap"

Q	13	15	17	19	21	23
$K_{1,2} - K_Q$	6.71×10^{-8}	4.19×10^{-9}	2.62×10^{-10}	1.63×10^{-11}	1.02×10^{-12}	6.39×10^{-14}

where β_m is the smallest positive root of the equation

$$\sqrt{2}\sin\frac{m\beta}{2} = \cos\frac{\beta}{2}. \quad (5)$$

These equations, in particular, yield exact values of the first two critical numbers, $K_{1,1} = 1/3$ and $K_{1,2} = 1/8$.

Mapping (1), (2) has also been studied in detail in [9], where the existence of invariant curves with rational and irrational rotation numbers in a wide range of parameter K variations is rigorously proved. The conservation of separatrices and the formation of invariant resonance structures of various orders is not specially discussed in [9], but there is a reference to the invariant properties of integer resonance separatrices with $\nu = 0$ for the first critical numbers $K_{1,1} = 1/3$ and $K_{1,2} = 1/8$. It appears that precisely in [9], the fact of the "unusual" dynamic behavior of system (1), (2) has been established for the first time.

The existence of a countable set of parameter K values at which resonance separatrices remain intact in spite of the perturbing action of many other resonances has important dynamic consequences.

The region occupied by one chaotic trajectory at the system parameter value equal to the first even critical number $K = K_{1,2} = 1/8$ is shown in the inset in Fig. 1. The upper and lower boundaries of this region are very close (see Section 3) to the integer resonance separatrices calculated by the exact Ovsyannikov equations [Eqs. (1.5)]. Here, everything is at variance with "usual" dynamics concepts.

Infinitely many "nontypical" Hamiltonian systems with all resonance separatrices unsplit are known to exist. These are so-called completely integrable systems with no chaos in their dynamic behavior [2, 3]. The striking feature of the situation shown in Fig. 1 is the coexistence of undestroyed separatrices of two neighboring integer resonances and the region of utter chaos, in which all the other invariant resonance structures and invariant curves with very stable irrational rotation numbers are destroyed and the chaotic layers of all fractional resonances overlap each other.

One of the most important questions is that of the dynamic situation near such an intact separatrix. In the next section, we give a partial answer to it.

The situation with the first odd critical number $K = K_{1,1} = 1/3$ is even more intriguing. For all fractional resonances determined by us numerically (accurate to 25 decimal places), we obtained $K_{Q,1} = K_{1,1} = 1/3$. It follows that, at $K = 1/3$, resonances of all orders are invariant resonance structures, and chaotic trajectories never intersect the separatrix of "their own" resonance within which they originate.

3. DYNAMICS NEAR AN INTEGER RESONANCE SEPARATRIX AT $K \approx K_{1,2}$

In this section, we discuss one of the main problems, that of the dynamic situation near an integer resonance separatrix when the system approaches the critical state, in the critical sate (Fig. 1), and at the exit from it (see Fig. 1).

At a small positive deviation of system parameter K from the $K_{1,2} = 1/8$ critical value, the separatrix begins to transmit other trajectories, but the mean time (the number of iterations) T_C of resonance passage depends on the detuning $\Delta K = K - K_{1,2} > 0$. An empirical equation for the mean integer resonance passage time was obtained in [7]. This equation has the form

$$\langle T_C \rangle = 135\left(\frac{K_{1,2}}{K - K_{1,2}}\right)^{1.193}(1 \pm 0.09), \quad K > K_{1,2}. \quad (6)$$

Clearly, the passage time infinitely increases as the system approaches the critical state.

In the critical state (Fig. 1), a chaotic trajectory, as mentioned, fairly closely approaches separatrices. To quantitatively estimate this closeness, we determined the smallest (in momentum) distance between this trajectory and integer resonance separatrices calculated by the exact Ovsyannikov formulas. It was found that there is a gap between the trajectory and the lower separatrix (we call it main) with a minimum width $dp_{\min} \approx 3 \times 10^{-6}$. A strongly enlarged small part of the gap is shown in Fig. 1. As follows from this figure, in the immediate vicinity of the separatrix, there is a region filled by fractional resonances $1/Q$ of comparatively high orders, which are arranged in layers. The larger the Q value, the closer its region to the separatrix (the lowest resonance discernible in the figure is 1/23). As each such resonance has full extension over the phase, getting into the lower layer requires traversing the upper. For some of the gap resonances, critical numbers closest to $K_{1,2} = 1/8$ were found; these values are listed in the table. The Q resonance rotation number $1/Q$ and critical number K_Q tend to the main separatrix rotation number (zero) and critical number $K_{1,2} = 1/8$ as the order of the resonance increases. For arbitrary Q, $K_Q < K_{1,2}$.

A chaotic trajectory that tends to approximate the main separatrix should traverse these resonances, and traversing each resonance takes a certain time. Direct measurements of the time of trajectory penetration

from the chaotic sea to the chaotic region within a high-order fractional resonance are technically impossible. For this reason, we measured another value, the mean time of escape from a resonance of the chaotic trajectory started within it, T_E [7]:

$$\langle T_E \rangle = 80e^{0.705Q}(1 \pm 0.07). \quad (7)$$

According to this formula, gap resonance critical numbers approach the $K_{1,2} = 1/8$ value from below as Q increases. As a consequence, intrinsic separatrices of these resonances become increasingly less "transparent," and the time of escape of a chaotic trajectory from them grows infinitely. Of course, there can be no global chaos in the critical state of an integer resonance.

The situation that arises when the system gets out of the critical state (that is, when K deviates from $K_{1,2}$ toward lower values) has been found to be much more complex.

Global chaos certainly exists at $K_{1,2} = 0.125 > K > K_{1,3} = 0.06191\ldots$ and far from critical numbers. For instance, in the $0.1090 \le K \le 0.1094$ interval, global diffusion is observed whose rate sharply decreases toward interval end points (the initial conditions $x = 0$, $p = 0.37$; the number of iterations 10^9). However, if the parameter value smoothly decreases starting with $K = K_{1,2} = 0.125$, then, as follows from the table, gap resonances successively (one after another) reach their critical states and prevent the trajectory from approaching the main separatrix at a distance smaller than some $dp_{\min}$. For instance, at $K = 0.12499999581\ldots$ (critical number for $Q = 15$), the minimum distance between the trajectory and the separatrix is $dp_{\min} \approx 4.6 \times 10^{-5}$, and at $K = 0.12499993294\ldots$ (critical number for $Q = 13$), $dp_{\min} \approx 1.8 \times 10^{-4}$; the number of mapping iterations was 10^{11} for both trajectories.

These results can be summarized in a somewhat pictorial manner. It is as though high-order resonances bar the passage to the main separatrix and resist the approach of chaotic trajectories to it. In the supercritical state, $K > K_{1,2}$, a trajectory can traverse the whole resonance, and the closer K to $K_{1,2}$, the longer the time this takes. At $K = K_{1,2}$, a trajectory can approximate the main separatrix arbitrarily closely but never reaches it. At the exit from the critical state, the system successively passes the critical states of gap resonances as K smoothly decreases, and the development of global chaos is possible in principle only within intervals between neighboring critical numbers.

4. SEPARATRIX MAPPING AND THE MEL'NIKOV–ARNOLD INTEGRAL FOR CONTINUOUS SYSTEM (3)

The empirical results described above require an explanation. First and foremost, we would like to understand the nature of the mechanism responsible for the oscillating and alternating behavior of the main $\alpha_1(K)$ dependence obtained in Section 2. It is a matter of general experience that mappings are more convenient to study by numerical methods, whereas continuous systems more easily lend themselves to analytic treatment [2]. For this reason, we will deal with Hamiltonian (3) in this section and then apply the results to mapping (1), (2).

The separatrix ($x_s(\psi)$, $\psi = 2\omega_0 t$) of an unperturbed ($\epsilon = 0$) system (3) can be written in the form

$$x_s(\psi) = \begin{cases} ae^{\psi} \text{ at } -\infty < \psi < -\pi/4, \\ \frac{1}{2}\left(1 + \frac{1}{\sqrt{2}}\sin\psi\right) \text{ at } -\pi/4 < \psi < \pi/4, \\ 1 - ae^{-\psi} \text{ at } \pi/4 < \psi < \infty, \end{cases} \quad (8)$$

where $a = e^{\pi/4}/4$.

The Hamiltonian value on the separatrix is $H_{0,s} = \omega_0^2/4$, and the period of motion close to the separatrix is calculated by the formula

$$T_s(w) \approx \frac{1}{2\omega_0}\ln\left(\frac{2e^{\pi/2}}{|w|}\right). \quad (9)$$

Here and throughout, $w = H_0/H_{0,s} - 1$ is the relative deviation from the separatrix in energy. Note that equation (9) is comparatively accurate, and the values calculated by (9) deviate from those found numerically by less than 5% in the interval $-0.47 \le w \le 1.06$.

The separatrix mapping for system (3) is written in the form [1]

$$\begin{aligned} \bar{w} &= w + W\sin\tau, \\ \bar{\tau} &= \tau + \lambda\ln\left(\frac{2e^{\pi/2}}{|\bar{w}|}\right) \pmod{2\pi}. \end{aligned} \quad (10)$$

Here, $\lambda = \Omega/(2\omega_0)$ is the adiabatic parameter ($2\omega_0$ is the frequency of oscillations near the stable equilibrium point $x = 0.5$, $p = 0$). Amplitude W in (10) can be found numerically (we then write $W = W_E$) or theoretically ($W = W_T$); the corresponding dependences will now be compared.

The theoretical W_T amplitude value is given by the formula

$$\begin{aligned} W_T(\lambda) &= \pm\max|\bar{w} - w| \\ &= \pm\frac{\max(\Delta H_0)}{H_{0,s}} = \frac{16\pi\epsilon}{\omega_0^2}W_{MA}(\lambda), \end{aligned} \quad (11)$$

where the sign coinciding with the sign of ΔH_0 and W_{MA} is the Mel'nikov–Arnold integral.

Following the technique described in [1], we will seek changes in unperturbed H_0 energy during a half-period of vibrations or a period of rotations,

$$\Delta H_0 = \int_{-\infty}^{\infty}\{H_0, H\}dt = -\epsilon\int_{-\infty}^{\infty}p(t)\frac{\partial U}{\partial x}dt.$$

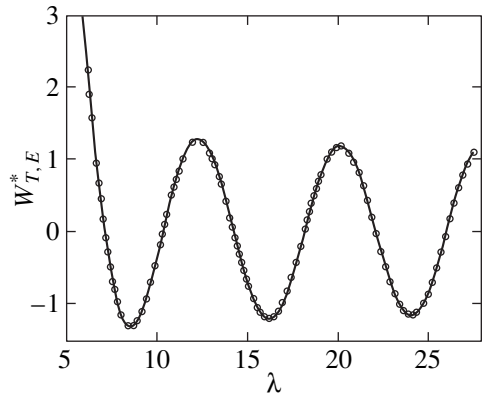


Fig. 2. Solid line: normalized theoretical amplitude W_T^* of separatrix mapping (10); solid circles: normalized W_E^* amplitude obtained by numerically constructing this mapping over system motion periods.

Here, {…} is the Poisson bracket. Assuming that the system moves fairly closely to the separatrix and calculating the latter relation at $x = x_s$, $p = p_s = \dot{x}_s = 2\omega_0 dx_s/d\psi$, we obtain

$$\Delta H_0 = -4\pi\epsilon\sin\tau_0 \times \int_{-\infty}^{\infty}\left[\frac{p_s(\psi)}{2\omega_0}\right]\sin[2\pi x_s(\psi)]\sin(\lambda\psi)d\psi.$$

Note in passing that the p_s factor in the integrand removes the well-known problem of special normalization of this integral to suppress its oscillating part, because this factor tends to zero at both integration limits (see Section 4.4 in review [1] for details).

The Mel'nikov–Arnold integral can eventually be written as

$$W_{MA}(\lambda) = \frac{\max\Delta H_0}{4\pi\epsilon} = \frac{1}{\sqrt{2}}\int_0^{\pi/4}\sin\left(\frac{\pi}{\sqrt{2}}\sin\psi\right)\cos\psi\sin(\lambda\psi)d\psi + 2a\int_{\pi/4}^{\infty}e^{-\psi}\sin(2\pi a e^{-\psi})\sin(\lambda\psi)d\psi. \quad (12)$$

The first term in (12) describes the contribution of the "elliptical" separatrix portion ($0.25 \le x_s \le 0.75$), and the second term is the contribution of the remaining separatrix part. Both terms are oscillating and alternating, and their oscillations are almost in antiphase to each other. The resulting $W_{MA}(\lambda)$ function is also alternating and oscillating (see Fig. 2). Interestingly, (12) admits a very simple asymptotic ($\lambda \gg 1$) representation

$$W_{MA}(\lambda) \approx -1.17\lambda^{-3.034}\cos\left(\frac{\pi}{4}\lambda\right). \quad (13)$$

The root-mean-square error of calculations by (13) is about 4% in the $10 \le \lambda \le 150$ interval and about 1.3% at $50 \le \lambda \le 150$.

Theoretical W_T amplitudes calculated by (11) and (12) should be compared with numerically calculated W_E values, which were determined as follows (see [11] for details). The central intersection point between the main resonance branches of system (3) lay on the $x = 0.5$ line, and its ordinate p_h was very accurately determined precisely along this line. The initial conditions for orbits were randomly selected on the same line in a narrow neighborhood of p_h, assuredly within the chaotic layer. The trajectory made the prescribed number of periods of motion, N (the period of motion, T, is the time interval between successive moments of passing the stable phase $x = 0.5$). For each T, the stable energy was calculated by the formula

$$w = 2\exp\left(\frac{\pi}{2} - 2\omega_0 T\right). \quad (14)$$

Separatrix mapping (10) (δw_k, $\tau_k = \Omega t_k (\mathrm{mod}2\pi)$, $k = 1, 2, \ldots, N-1$) can be constructed and its amplitude W_E found by determining energy changes $\delta w = \bar{w} - w$ for a pair of neighboring periods and assigning to time moment t_k common to these periods.

Theoretical W_T values calculated by (12) and numerical W_E amplitudes of separatrix mapping (10) are compared in Fig. 2. For convenience, not W_T and W_E themselves but their normalized values

$$W_T^* = AW_T, \quad W_E^* = AW_E, \quad A = \frac{\omega_0^2\lambda^{3.034}}{16\pi\epsilon} \quad (15)$$

are compared [exponent λ in the correcting factor is from adjustment equation (13)]. Because of the excellent coincidence of the two curves, we can confidently use formula (12) for the Mel'nikov–Arnold integral and assume that $W = W_T$ in (10).

The value most important for practical applications is the half-width of the chaotic layer ($w_{\max}$) rather than the separatrix mapping amplitude (W). For all amplitude values shown by solid circles in Fig. 2, we iterated the separatrix mapping to find $w_{\max}$. We found that the relation between the size of the layer ($w_{\max}$) and the amplitude (W) suggested by Chirikov (Eq. (6.8) in [1]),

$$w_{\max} \approx \lambda|W|, \quad (16)$$

was fairly accurate in our case, even in the neighborhoods of zero W amplitudes. For instance, at $\lambda > 10$, we have $0.94 \le w_{\max}/\lambda|W| \le 1.3$.

The results obtained for continuous system (3) can be extended to mapping (1), (2). The $\lambda_{1,m}$ ($m = 1, 2, \ldots$) values at which $W_{MA}(\lambda_{1,m}) = 0$ will be called critical (the first index shows that the values refer to the principal resonance). Clearly, the transition of continuous system (3) with parameter λ to mapping (1), (2) with parameter K and vice versa should be performed with the use of the relation

$$K = \omega_0^2 = (\pi/\lambda)^2. \quad (17)$$

This equality allows the critical $\lambda_{1,m}$ values of system (3) to be recalculated to the critical mapping parameter values. Let the values found in this way be denoted by K^*; that is,

$$K^*_{1,m} = \left(\frac{\pi}{\lambda_{1,m}}\right)^2, \quad m = 1, 2, \ldots \quad (18)$$

These values should be compared with $K_{1,m}$ calculated by (4), (5). The ratio between K and K^* is well described by the formula

$$\frac{K_{1,m}}{K^*_{1,m}} \approx 1 + 0.676m^{-0.875}, \quad m = 1, 2, \ldots$$

In the asymptotic limit ($m \gg 1$), sets $K_{1,m}$ and $K^*_{1,m}$ coincide, which justifies extending the results obtained in this section to mapping (1), (2).

The goal formulated in the beginning of this section can be considered partially achieved. Namely, the Mel'nikov–Arnold integral is responsible for the behavior of the W_T separatrix mapping amplitude, which periodically passes zero, and for the absence of separatrix splittings at these points. Our discussion also explains why energy changes along the elliptical and hyperbolic separatrix regions have different signs.

5. MODEL GENERALIZATION

The dynamics described above is, in our view, far from trivial. The natural questions arise: is not function (2), which is responsible for this dynamics, something unique, and will not all the unusual dynamic effects disappear when the construction is slightly "stirred?" To answer these, we will use this short section to consider an asymmetric saw. The model will be constructed by the introduction of skewness parameter s into (2):

$$f(x, s) = \begin{cases} \dfrac{4x}{1+4s} & \text{at } 0 \le x \le 0.25 + s, \\ \dfrac{2-4x}{1-4s} & \text{at } 0.25 + s \le x \le 0.75 - s, \\ \dfrac{4(x-1)}{1+4s} & \text{at } 0.75 - s \le x \le 1. \end{cases} \quad (19)$$

The s parameter can vary in the interval $-0.25 \le s \le 0.25$; at $s = 0$, we return to symmetrical model (2).

At the end points of this interval, the two-tooth saw degenerates into a one-tooth saw, but the dynamic behaviors of the system at these limit points are qualitatively different. If $s = -0.25$, motion along the separatrix is simple harmonic oscillations, which we will not consider. In the other limit ($s = 0.25$), separatrices split at arbitrary K values, and chaos becomes global. The corresponding dynamics has been intensely studied, and there is nothing unusual in it (e.g., see [12]). The main goal of this section is to understand the most important details of this transition. Let us use the equivalence of the continuous and discrete descriptions established above and trace the evolution of the Mel'nikov–Arnold integral accompanying parameter s variations.

At $s \neq 0$, the elliptical and hyperbolic saw period parts become unequal, and motion in these parts is described by different frequencies (ω_1 and ω_2, respectively):

$$\omega_1 = \omega_0/\sqrt{1-4s}, \quad \omega_2 = \omega_0/2\sqrt{1+4s}. \quad (20)$$

Here, time t can conveniently be used as an independent variable. The duration of motion along the separatrix from the stable equilibrium point ($x = 0.5$) to the break point ($x = 0.75 - s$), T_{sm}, is given by the formula

$$T_{sm} = \frac{1}{2\omega_1}\arcsin\sqrt{\frac{1-4s}{2}}. \quad (21)$$

Separatrix $x_s(t)$ of unperturbed system (3) should be written as

$$x_s(t, s) = \begin{cases} a_s e^{2\omega_2 t} & \text{at } -\infty < t < -T_{sm}, \\ \dfrac{1}{2}\left(1 + \sqrt{\dfrac{1-4s}{2}}\sin 2\omega_1 t\right) & \text{at } -T_{sm} < t < -T_{sm}, \\ 1 - a_s e^{-2\omega_2 t} & \text{at } T_{sm} < t < \infty, \end{cases} \quad (22)$$

where $a_s = (1 + 4s)e^{2\omega_2 T_{sm}}/4$.

The Mel'nikov–Arnol'd integral then takes the form

$$W_{MA}(\lambda, s) = 2\omega_1\sqrt{\frac{1-4s}{2}} \times \int_0^{T_{sm}} \sin\left(\pi\sqrt{\frac{1-4s}{2}}\sin 2\omega_1 t\right)\cos(2\omega_1 t)\sin(\Omega t)\,dt \quad (23)$$

$$+ 4\omega_2 a_s \int_{T_{sm}}^{\infty} e^{-2\omega_2 t}\sin(2\pi a_s e^{-2\omega_2 t})\sin(\Omega t)\,dt.$$

Using this formula, we were able to study the whole interval of parameter s variations. The following picture emerged. Everywhere within this interval, the Mel'nikov–Arnold integral remained alternating and oscillating, and, therefore, there remained a set of critical numbers and the

corresponding invariant resonance structures. The half-period of oscillations Λ_0 (the distance along λ between neighboring zeros of the integral), however, infinitely increased as $s \longrightarrow 0.25$. At $s = 0.25$, the first zero of the Mel'nikov–Arnold integral went into infinity and the "usual" dynamics was restored [12].

This conclusion fully agrees with the results of [9], where an asymmetric saw model similar to (1), (19) was considered, and the existence of invariant curves with various rotation numbers was proved for it. Considered among others was the integer resonance separatrix with $\nu = 0$.

6. CONCLUSION

In our view, mapping (1), (2) and continuous system (3) with an unusual dynamics deserve further study. Several of the most important problems should be mentioned.

In [9] and this work, the following very complex dynamic picture of mapping (1), (2) has been revealed. Each resonance of the countable set of system resonances has its own (presumably countable) set of critical parameter K values, at which the resonance becomes an invariant structure (with rotation number $\nu = 0$) impermeable to other trajectories. This set should be augmented by all those K values (they can also be called critical numbers) for which the existence of invariant curves with rational and irrational $\nu > 0$ was proved in [9]. Of obvious interest would be a study of global diffusion for system (1), (2) at $K \le 1/3$. Such a study would, we believe, allow us to gain insight into how the set of all critical numbers is organized and what in essence it is.

In Section 3, we described the dynamics of the approach to and exit from the critical state for an integer resonance. Whether or not this scenario is typical and whether or not it is valid for fractional resonances is an open question.

A simple theory explaining the nature of alternating $\alpha_1(K)$ dependences and separatrix stability for an integer resonance was suggested in [8]. A similar analysis of fractional resonances would be useful.

Possibly, answers to these and other questions would allow us to essentially refine some modern theoretical concepts concerning nonlinear Hamiltonian systems.

ACKNOWLEDGMENTS

The author is deeply indebted to B.V. Chirikov for numerous discussions and advice and to D.L. Shpelyanskii for information about Bullet's paper [9]. This work was financially supported by the Russian Foundation for Basic Research (project no. 97-01-00865).

APPENDIX

Ovsyannikov Theorem on Sawtooth Mapping Separatrices [6]

Consider the equation for the $x(t)$ function defined on the entire real line $\mathcal{R}(-\infty < t < +\infty)$:

$$x(t+h) + x(t-h) - 2x(t) = h^2 f(x(t)), \qquad \text{(A.1)}$$

where $h > 0$ is a fixed constant and $f(x)$ is an odd 4-periodic function defined for $0 \le x \le 2$ by the formula

$$f(x) = \begin{cases} -x, & 0 \le x \le 1, \\ x - 2, & 1 \le x \le 2. \end{cases} \qquad \text{(A.2)}$$

Solution $x(t)$ of (A.1) with function (A.2) is called a separatrix if $x(t)$ is monotonic and continuous on the entire real line $\mathcal{R}$, $x(0) = 0$, and $x(t) \longrightarrow 2$ as $t \longrightarrow +\infty$ and $x(t) \longrightarrow -2$ as $t \longrightarrow -\infty$.

Theorem. There exists a sequence $\{h_k\}$, $k = 1, 2, \ldots$, such that $h_k \longrightarrow 0$ as $k \longrightarrow \infty$, and for each $h = h_k$, there exists the separatrix $x = x^k(t)$ as a solution to (A.1) with function (A.2). As $k \longrightarrow \infty$, the sequence $\{x^k(t)\}$ converges uniformly on $\mathcal{R}$ to a solution that is the separatrix for the limiting equation $x''(t) = f(x)$,

$$x(t) = \begin{cases} \sqrt{2}\sin t, & 0 \le t \le \pi/4, \\ 2 - e^{\pi/4 - t}, & \pi/4 \le t \le \infty. \end{cases} \qquad \text{(A.3)}$$

Construction. Put $h_k = 2\sin(\alpha_k/2)$, where α_k is the smallest positive root of the equation

$$\sqrt{2}\sin(k\alpha) = \cos(\alpha/2). \qquad \text{(A.4)}$$

For $nh_k \le t \le (n+1)h_k$, the separatrix is then given by the equations

$$x^k(t) = X_n\left(n + 1 - \frac{t}{h_k}\right) + X_{n+1}\left(\frac{t}{h_k} - n\right), \qquad \text{(A.5)}$$
$$n = 0, 1, 2\ldots$$

with the coefficients

$$X_n = \frac{\sin n\alpha_k}{\sin k\alpha_k}, \quad n \le k,$$

$$X_n = 2 - \left(\frac{\sqrt{h_k^2 + 4} - h_k}{2}\right)^{2(n-k)}, \quad n \ge k.$$

Note that, because the period and position of force (A.2) are different from those in the main text [Eq. (2)], the transition from the Ovsyannikov parameter h in (A.1) to the K parameter of mapping (1) and vice versa requires the use of the relation

$$K = h^2/4.$$

REFERENCES

1. B. V. Chirikov, Phys. Rep. **52**, 263 (1979).
2. A. Lichtenberg and M. Lieberman, *Regular and Chaotic Dynamics* (Springer-Verlag, Berlin, 1992).
3. G. M. Zaslavskiĭ and R. Z. Sagdeev, *Introduction to the Nonlinear Physics* (Nauka, Moscow, 1988).
4. R. S. MacKay, J. D. Meiss, and I. C. Percival, Physica D (Amsterdam) **13**, 55 (1984).
5. J. Moser, Nachr. Akad. Wiss. Goett. II, Math.-Phys. Kl., No. 1, 1 (1962); *Stable and Random Motion in Dynamical Systems* (Princeton Univ. Press, Princeton, 1973), Annals of Mathematics Studies, No. 77.
6. L. V. Ovsyannikov, private communications (1999).
7. V. V. Vecheslavov, Preprint No. 2000-27, IYaF (Budker Inst. of Nuclear Physics, Novosibirsk, 2000); nlin.CD/0005048 (2000).
8. V. V. Vecheslavov and B. V. Chirikov, Preprint No. 2000-68, IYaF (Budker Inst. of Nuclear Physics, Novosibirsk, 2000).
9. S. Bullet, Commun. Math. Phys. **107**, 241 (1986).
10. V. V. Vecheslavov and B. V. Chirikov, Zh. Éksp. Teor. Fiz. **114**, 1516 (1998) [JETP **87**, 823 (1998)]; V. V. Vecheslavov, Zh. Éksp. Teor. Fiz. **116**, 336 (1999) [JETP **89**, 182 (1999)].
11. V. V. Vecheslavov, Zh. Éksp. Teor. Fiz. **109**, 2208 (1996) [JETP **82**, 1190 (1996)].
12. I. Dana, N. W. Murray, and I. C. Persival, Phys. Rev. Lett. **62**, 233 (1989).

Translated by V. Sipachev

Journal of Experimental and Theoretical Physics, Vol. 92, No. 4, 2001, p. 752.

ERRATA

Erratum: "Coulomb Effects in Spatially Separated Electron and Hole Layers in Coupled Quantum Wells" [*JETP* 92 (2), 260 (2001)]

L. V. Butov, A. Imamoglu, K. L. Campman, and A. C. Gossard

1. The correct reference numbering in Fig. 4 is [13, 15] instead of [14, 16]; [18] instead of [19]; and [14–16] instead of [15–17];
2. Page 264, right column, 5th line from top: 1.5 meV instead of 13 meV;
3. Fig. 6, left inset: 20 W/cm^2 instead of 200 W/cm^2;
4. Fig. 6, right inset: ×50 should be added for the lower spectrum.

1063-7761/01/9204-0752$21.00