Journal of Experimental and Theoretical Physics, Vol. 91, No. 4, 2000, pp. 653–657.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 4, 2000, pp. 755–760.

NUCLEI, PARTICLES, AND THEIR INTERACTION

Comments on the Morita Equivalence¶

K. Saraĭkin

Landau Institute for Theoretical Physics, Moscow, 117334 Russia

Institute of Theoretical and Experimental Physics, Moscow, 117259 Russia
e-mail: saraikin@itp.ac.ru

Received May 25, 2000

Abstract—It is known that the noncommutative Yang–Mills (YM) theory with periodical boundary conditions on a torus at a rational noncommutativity parameter value is Morita equivalent to the ordinary YM theory with twisted boundary conditions on a dual torus. We give a simple derivation of this fact. We describe the one-to-one correspondence between these two theories and the corresponding gauge invariant observables. In particular, we show that under the Morita map, the Polyakov loops in the ordinary YM theory are converted to the open noncommutative Wilson loops discovered by Ishibashi, Iso, Kawai, and Kitazawa. *© 2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Noncommutative geometry deals with functions on a deformation of ordinary space where the coordinates do not commute:[1]

$$[\hat{x}_\mu, \hat{x}_\nu] = 2\pi i\theta_{\mu\nu}, \quad \mu, \nu = 1, \ldots, d. \tag{1.1}$$

The antisymmetric tensor $\theta_{\mu\nu}$ is called the noncommutativity parameter. The deformed flat ($\theta_{\mu\nu}$ = const) and compact space is called the noncommutative (quantum) torus $\mathbf{T}^d_\theta$. Recently, the noncommutative geometry and, especially, the noncommutative torus were seen to play an important role in the *M*-theory compactifications [1] and in string theory (see [2] and references therein). The noncommutative geometry is also very useful in compactifications of instanton moduli spaces [3]. The way to deal with the curved quantum spaces is provided by the Kontsevich deformation quantization.

A very intriguing subject from noncommutative geometry is the so-called Morita equivalence [4]. Roughly speaking, it states that certain bundles on different noncommutative tori are dual to each other. From the physical standpoint, this results in the equivalence between certain noncommutative and ordinary gauge theories. In what follows, we try to clarify this statement using a set of simple examples.

¶ This article was submitted by the author in English.

[1] In what follows, we use the same notation […] for the ordinary and the star-commutator. To avoid confusion, we supply all noncommutative quantities with the hats.

2. NOTATION

The algebra $\mathcal{A}_\theta$ of smooth functions on the noncommutative torus is defined using the Moyal star product:

$$f * g(\hat{\mathbf{x}}) = \exp\left(i\pi\theta_{\mu\nu}\frac{\partial}{\partial\xi_\mu}\frac{\partial}{\partial\eta_\nu}\right)f(\xi)g(\eta)|_{\xi=\eta=\hat{\mathbf{x}}}. \tag{2.1}$$

The main property of this product is its associativity. In applications, it is useful to decompose functions on the noncommutative torus into the Fourier components[2] as

$$f(\hat{\mathbf{x}}) = \sum_{\mathbf{k}\in\mathbb{Z}^d} f_{\mathbf{k}} e^{i\mathbf{k}\cdot\hat{\mathbf{x}}}. \tag{2.2}$$

This corresponds to the Weil or symmetric ordering of coordinates. The exponentials

$$\hat{U}_{\mathbf{k}} = e^{i\mathbf{k}\cdot\hat{\mathbf{x}}}$$

can serve as basis elements for the algebra $\mathcal{A}_\theta$.

A very intriguing phenomenon occurs when the θ-tensor components become rational. We first consider the two-torus $\mathbf{T}^2$,

$$[\hat{x}_\mu, \hat{x}_\nu] = 2\pi i\theta\epsilon_{\mu\nu}, \quad \mu, \nu = 1, 2, \tag{2.3}$$

[2] Without losing generality, we can consider a torus of size 2π.

1063-7761/00/9104-0653$20.00 © 2000 MAIK "Nauka/Interperiodica"

with the rational noncommutativity parameter $\theta = M/N$, where M and N are coprime integers. Then,

$$[\hat{U}_{\mathbf{n}}, \hat{U}_{\mathbf{n}'}] = 2i\sin\left(\pi M\frac{n_2n_1' - n_1n_2'}{N}\right)\hat{U}_{\mathbf{n}+\mathbf{n}'} = 2i\sin(\mathbf{n}\times\mathbf{n}')\hat{U}_{\mathbf{n}+\mathbf{n}'}, \quad (2.4)$$

where by definition $\mathbf{n}\times\mathbf{n}' \equiv -\pi\theta_{\mu\nu}n_\mu n_4'$. We note that the elements $\hat{U}_{N\mathbf{k}}$ generate the center of $\mathcal{A}_\theta$; that is, we have

$$[e^{iN\mathbf{k}\cdot\hat{\mathbf{x}}}, f(\hat{\mathbf{x}})] = 0 \quad (2.5)$$

for any $f(\hat{\mathbf{x}})$. This means that the exponentials $\{\hat{U}_{\mathbf{k}}, \mathbf{k} = 0|_{\mathrm{mod}\,N}\}$ entering decomposition (2.2) can be treated as if they were ordinary exponentials defined on ordinary (commutative) space. The other $N^2 - 1$ exponentials obtained from the set $\{\hat{U}_{\mathbf{k}}, \mathbf{k} \neq 0|_{\mathrm{mod}\,N}\}$ after factorization over the commutative part generate a closed algebra under the star-commutator. This algebra is isomorphic to $SU(N)$, as we will see momentarily. Therefore, at the rational value of the noncommutativity parameter, one can identify the algebra of functions on the noncommutative torus with the algebra of matrix-valued functions on the commutative torus.

We conclude this section by giving an explicit matrix representation for the algebra of the noncommutative exponentials (see also [5]). This representation has been well known for many years [6, 7]. We introduce the clock and shift generators

$$Q = \begin{pmatrix} 1 & & & & \\ & \omega & & & \\ & & \omega^2 & & \\ & & & \ddots & \\ & & & & \omega^{N-1} \end{pmatrix}, \quad P = \begin{pmatrix} 0 & 1 & & & 0 \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ 1 & & & & 0 \end{pmatrix}, \quad (2.6)$$

where $\omega = e^{2\pi i\theta}$. The matrices P and Q are unitary, traceless, and satisfy the relations

$$P^N = Q^N = \mathbf{1}, \quad PQ = \omega QP. \quad (2.7)$$

Moreover,

$$\mathrm{Tr}(P^nQ^m) = \begin{cases} N \text{ if } n = 0|_{\mathrm{mod}\,N} \text{ and } m = 0|_{\mathrm{mod}\,N}, \\ 0 \text{ if } n \neq 0|_{\mathrm{mod}\,N} \text{ or } m \neq 0|_{\mathrm{mod}\,N}. \end{cases} \quad (2.8)$$

It is straightforward to verify that the generators defined as

$$J_{\mathbf{n}} = \omega^{n_1n_2/2}Q^{n_1}P^{n_2}, \quad \mathbf{n} = (n_1, n_2) \quad (2.9)$$

satisfy commutation relations (2.4):

$$[J_{\mathbf{n}}, J_{\mathbf{n}'}] = 2i\sin(\mathbf{n}\times\mathbf{n}')J_{\mathbf{n}+\mathbf{n}'}. \quad (2.10)$$

This identity can be rewritten as the Lie algebra commutation relations

$$[J_{\mathbf{n}}, J_{\mathbf{m}}] = f_{\mathbf{nm}}^{\mathbf{k}}J_{\mathbf{k}} \quad (2.11)$$

with the structure constants

$$f_{\mathbf{nm}}^{\mathbf{k}} = 2i\delta_{\mathbf{n}+\mathbf{m},\mathbf{k}}\sin(\mathbf{n}\times\mathbf{m}). \quad (2.12)$$

The set of unitary unimodular $N\times N$ matrices (2.9) is sufficient to span the $SU(N)$ algebra.

3. THE MORITA EQUIVALENCE

3.1. The Dual Torus $U(1)|_{\theta = M/N} \longrightarrow U(N)$

To define the Morita map, we use an additional decomposition of function (2.2) on the noncommutative two-torus

$$\hat{f} = \sum_{k\in\mathbb{Z}^2} e^{iN\mathbf{k}\cdot\hat{\mathbf{x}}}\sum_{n_1,n_2=0}^{N-1} f_{\mathbf{kn}}e^{in_1\hat{x}_1 + in_2\hat{x}_2}. \quad (3.1)$$

We then define the corresponding $U(N)$-valued function on the ordinary two-torus as

$$f = \sum_{k\in\mathbb{Z}^2} e^{iN\mathbf{k}\cdot\mathbf{x}}\sum_{n_1,n_2=0}^{N-1} f_{\mathbf{kn}}e^{i\mathbf{n}\cdot\mathbf{x}}J_{\mathbf{n}}. \quad (3.2)$$

Due to the relation

$$J_{\mathbf{n}}J_{\mathbf{n}'} = e^{i\mathbf{n}\times\mathbf{n}'}J_{\mathbf{n}\times\mathbf{n}'}, \quad (3.3)$$

the Morita map (3.1), (3.2) takes the star-product to the matrix product. Obviously, a general $U(N)$-valued function cannot be represented in form (3.2). It turns out that this particular form corresponds to the functions with nontrivial boundary conditions. This means that under shifts of their arguments, these functions transform as

$$f\left(x_1 + 2\pi\frac{M}{N}, x_2\right) = \Omega_1 f(x_1, x_2)\Omega_1^\dagger, \quad f\left(x_1, x_2 + 2\pi\frac{M}{N}\right) = \Omega_2 f(x_1, x_2)\Omega_2^\dagger, \quad (3.4)$$

where

$$\Omega_1 = (P)^M, \quad \Omega_2 = (Q^\dagger)^M. \quad (3.5)$$

This can be considered as a constant gauge transformation. The size $2\pi M/N$ of the dual torus can be fixed by the requirement for the Morita map to be single-valued.[3] To illustrate this, we consider a torus of the size $2\pi(M/N)n$ (where $n \in \mathbf{N}$; there are no other possibilities if the functions of type (3.2) are required to be gauge-

[3] I am indebted to K. Selivanov for this comment.

conjugate by a constant matrix when translated along the torus lattice). In this case, there are functions that obviously cannot be represented in form (3.2). These functions are not conjugated when translated along the vectors $(2\pi M/N, 0)$ and $(0, 2\pi M/N)$.

Therefore, having a set of Fourier coefficients $f_{\mathbf{kn}}$, we can construct a function on the noncommutative torus of the size l and a matrix-valued function with twisted boundary conditions (3.4) on the commutative torus of the size $(M/N)l$ as follows:

$$\begin{cases} e^{i\mathbf{n}\hat{\mathbf{x}}} \longleftrightarrow e^{i\mathbf{n}\cdot\mathbf{x}} J_{\mathbf{n}}, \quad n_1, n_2 < N, \\ e^{iN\mathbf{k}\cdot\hat{\mathbf{x}}} \longleftrightarrow e^{iN\mathbf{k}\cdot\mathbf{x}} \mathbf{1}. \end{cases} \tag{3.6}$$

3.2. $\mathbf{T}^d$. $U(1)|_\theta \longrightarrow U(N_1) \times \ldots \times U(N_r)$

The generalization to the d-dimensional case goes by simple modifications of the formulas from the previous subsection. It is always possible to rotate $\theta_{\mu\nu}$ into the canonical skew-diagonal form

$$\theta_{\mu\nu} = \begin{pmatrix} 0 & \theta_1 & & & & \\ -\theta_1 & 0 & & & & \\ & & \ddots & & & \\ & & & 0 & \theta_r & \\ & & & -\theta_r & 0 & \\ & & & & & \mathbf{0}_{d-2r} \end{pmatrix}, \tag{3.7}$$

where r is the rank of $\theta_{\mu\nu}$. The algebra of a higher dimensional noncommutative torus is thereby embedded into a d-fold tensor product of r noncommutative two-torus algebras and the ordinary $(d-2r)$-torus commutative algebra. This immediately leads to other examples of the Morita equivalence, where some of these noncommutative dual tori are mapped to the commutative ones via (3.6). If

$$\theta_i = \frac{M_i}{N_i},$$

the Morita map results in the ordinary Yang–Mills (YM) theory with the gauge group $U(N_1) \times \ldots \times U(N_r)$.

3.3. $\mathbf{T}^d$. $U(1)|_\theta \longrightarrow U(N)$

The algebra of noncommutative exponentials can also be realized using a set of $SU(N)$-valued matrices Ω_μ, $\mu = 1, \ldots, d$, obeying the relations

$$\Omega_\mu \Omega_\nu = e^{2\pi i \theta_{\mu\nu}} \Omega_\nu \Omega_\mu. \tag{3.8}$$

An explicit construction of these matrices can be found in [8]. We define the generators $J_{\mathbf{n}}$ as

$$J_{\mathbf{n}} = \exp\left(\sum_{\nu<\mu} \theta_{\nu\mu} n_\nu n_\mu\right) \Omega_1^{n_1} \ldots \Omega_d^{n_d}. \tag{3.9}$$

Then,

$$[J_{\mathbf{n}}, J_{\mathbf{m}}] = 2i \sin(\mathbf{n} \times \mathbf{m}) J_{\mathbf{n}+\mathbf{m}}, \tag{3.10}$$

which coincides with the algebra of the noncommutative exponentials. In this case, therefore, the Morita map takes the form

$$\begin{aligned} \hat{f} &= \sum_{k \in \mathbb{Z}^d} e^{iN\mathbf{k}\cdot\hat{\mathbf{x}}} \sum_{\mathbf{n} < N^{\otimes d}} f_{\mathbf{kn}} e^{i\mathbf{n}\cdot\hat{\mathbf{x}}} \\ \longleftrightarrow f &= \sum_{k \in \mathbb{Z}^d} e^{iN\mathbf{k}\cdot\mathbf{x}} \sum_{\mathbf{n} < N^{\otimes d}} f_{\mathbf{kn}} e^{i\mathbf{n}\cdot\mathbf{x}} J_{\mathbf{n}}. \end{aligned} \tag{3.11}$$

4. THE NONCOMMUTATIVE YANG–MILLS THEORY VS. THE ORDINARY YANG–MILLS THEORY

We now turn to physical applications of the Morita map. One can define a noncommutative version of the YM theory with the action

$$S_{YM} = \frac{1}{4\pi g_{YM}^2} \int d\mathbf{x} \mathrm{Tr}(F_{\mu\nu} F^{\mu\nu}) \tag{4.1}$$

by simply replacing the matrix product by the Moyal star-product in all formulas and supplementing all quantities with the hats. Therefore, the noncommutative $U(1)$ YM action is

$$\hat{S} = \frac{1}{4\pi g_{NCYM}^2} \int d\hat{\mathbf{x}} \hat{F}_{\mu\nu} * \hat{F}^{\mu\nu}, \tag{4.2}$$

where

$$\hat{F}_{\mu\nu} = \partial_\mu \hat{A}_\nu - \partial_\nu \hat{A}_\mu - i[\hat{A}_\mu, \hat{A}_\nu]_*.$$

For simplicity, we only consider the dual torus in this section. The generalization to the higher dimensional case is straightforward.

The Morita map takes noncommutative $U(1)$ gauge fields to the $U(N)$ gauge fields with nontrivial boundary conditions. In general, functions on the torus can be gauge-conjugate when shifted by the period of the torus,

$$\begin{aligned} A_\lambda(\mathbf{x} + \mathbf{l}_\mu) &= \Omega_\mu(\mathbf{x}) A_\lambda(\mathbf{x}) \Omega_\mu^{-1}(\mathbf{x}) \\ &+ i\Omega_\mu(\mathbf{x}) \partial_\lambda \Omega_\mu^{-1}(\mathbf{x}), \end{aligned} \tag{4.3}$$

where $\Omega_\mu(\mathbf{x})$ are the elements of the $U(N)$ group that are known as the twist matrices. They must satisfy the consistency conditions

$$\Omega_\mu(\mathbf{x}+\mathbf{l}_\nu)\Omega_\nu(\mathbf{x}) = \exp\left(2\pi i\frac{M}{N}\epsilon_{\mu\nu}\right)\Omega_\nu(\mathbf{x}+\mathbf{l}_\mu)\Omega_\mu(\mathbf{x}). \quad (4.4)$$

The integer M entering this formula is the so-called *'t* Hooft flux. Only three types of possible boundary conditions (solutions of Eqs. (4.4)) are known:

(1) twist eaters: $\Omega_\mu = \text{const}$;

(2) Abelian twists;

(3) non-Abelian twists.

For more details, see the recent review [9].

The map (3.6) precisely corresponds to the first case. It is not well understood how to realize the Morita map corresponding to other boundary conditions. Roughly speaking, when working in the Fourier basis (2.2), one can only multiply functions with numbers after shifts and cannot add quantities of the form $\Omega_\mu(\mathbf{x})\partial_\lambda\Omega_\mu^{-1}(\mathbf{x})$. To do this, one needs another basis for the functions on the noncommutative torus (creation/annihilation operators, noncommutative theta-functions?).

Under the Morita map defined in the previous section, actions go into actions, equations of motions go into equations of motions, and solutions (e.g., instantons) go into solutions even at the quantum level. These properties of the Morita map can be encoded in the identity

$$\int d\hat{\mathbf{x}}\hat{A}_\mu * \hat{A}_\nu * \ldots * \hat{A}_\lambda = \frac{1}{N}\int d\mathbf{x}\,\mathrm{Tr}(A_\mu A_\nu \ldots A_\lambda), \quad (4.5)$$

which is easy to prove using the definition

$$\int d\hat{\mathbf{x}} e^{i\mathbf{k}\cdot\hat{\mathbf{x}}} = \delta_{\mathbf{k},0} \quad (4.6)$$

and property (2.8) of the clock and shift generators. In fact, one can insert an arbitrary number of derivatives into the integrals in (4.5) and thus obtain equivalent gauge-invariant quantities in the noncommutative and ordinary gauge theories. Using identity (4.5), we can relate the correlators as

$$\int \mathcal{D}A^\mu_{\mathbf{k},\mathbf{n}}\exp\left(\hat{S}\left[\theta=\frac{M}{N}\right]\right)\hat{\mathbb{O}}_1\ldots\hat{\mathbb{O}}_l = \int \mathcal{D}A^\mu_{\mathbf{k},\mathbf{n}}\exp(S_{YM})\Big|_{fxd\ bndry\ conds,\ flux=M}\mathbb{O}_1\ldots\mathbb{O}_l, \quad (4.7)$$

where $g^2_{NCYM} = Ng^2_{YM}$ and

$$\hat{\mathbb{O}} = \int d\hat{\mathbf{x}}(\hat{F}_{\mu\nu})^{*n}, \qquad \mathbb{O} = \frac{1}{N}\int d\mathbf{x}\,\mathrm{Tr}(F_{\mu\nu})^n. \quad (4.8)$$

Other important gauge-invariant quantities of the YM theory are the Wilson loops

$$W[C] = \mathrm{Tr}P\exp\left(i\oint_C A_\mu(\mathbf{x})(dx_\mu)\right) \quad (4.9)$$

corresponding to a closed path C. On the torus, there are paths from different homotopy classes, which can be classified by winding numbers w_μ around the μth direction. The corresponding Wilson loops are called the Polyakov loops. The simplest Polyakov loop corresponds to the straight line along the μth direction,

$$W_P[\mathbf{x},\mu] = \mathrm{Tr}\left[P\exp\left(i\int_{\mathbf{x}}^{\mathbf{x}+\mathbf{l}_\mu}A_\mu(\mathbf{x})dx_\mu\right)\Omega_\mu e^{ix_\mu}\right], \quad (4.10)$$

where the insertion of twist matrix (3.5) is necessary to guarantee gauge invariance.

Wilson lines were constructed in the noncommutative YM theory by Ishibashi, Iso, Kawai, and Kitazawa [10] (see also [11, 12]). This construction goes as follows. An oriented curve C in the auxiliary commutative two-dimensional space parametrized by the functions $\xi(\sigma)$ with $0 \le \sigma \le 1$ is introduced. One fixes the starting point $\xi_\mu(0) = 0$ and the endpoint $\xi_\mu(1) = v_\mu$ and then assigns to this curve a noncommutative analog of the parallel transport operator

$$\mathcal{U}[\hat{\mathbf{x}},C] = 1 + \sum_{n=1}^{\infty} i^n\int_0^1 d\sigma_1\int_{\sigma_1}^1 d\sigma_2\ldots\int_{\sigma_{n-1}}^1 d\sigma_n\frac{d\xi_{\mu_1}(\sigma_1)}{d\sigma_1}\ldots\frac{d\xi_{\mu_n}(\sigma_n)}{d\sigma_n} \times \hat{A}_{\mu_1}(\hat{\mathbf{x}}+\xi(\sigma_1)) * \ldots * A_{\mu_n}(\hat{\mathbf{x}}+\xi(\sigma_n)). \quad (4.11)$$

The series in (4.11) is a noncommutative analog of the P-exponential. The star-gauge invariant quantity is then

$$\hat{\mathbb{O}}[C] = \int d\hat{\mathbf{x}}\mathcal{U}[\hat{\mathbf{x}},C] * S[\hat{\mathbf{x}},C], \quad (4.12)$$

where $S[\hat{\mathbf{x}}, C] = 1$ if the path C is closed and

$$S[\hat{\mathbf{x}},C] = \exp(i(\theta^{-1})_{\mu\nu}v_\nu\hat{x}_\mu) \quad (4.13)$$

if the path is open. Gauge invariance requires that the endpoint coordinates must be equal to

$$v_\mu = 2\pi r_\mu\frac{M}{N}, \quad r_\mu = 0,\ldots,N-1.$$

In the simplest case, where C_μ is the straight line along the μth direction and $v_\mu = 2\pi M/N$, the function $S[\hat{\mathbf{x}}, C_\mu]$

goes to the twist function $\Omega_\mu e^{ix_\mu}$ under the Morita map (3.6). Therefore, identity (4.5) allows us to obtain the following relation between the Polyakov loops in the ordinary YM theory and open noncommutative Wilson loops:

$$\frac{1}{N}\int d\mathbf{x} W_P[\mathbf{x}, \mu] = \hat{\mathbb{C}}[C_\mu]. \quad (4.14)$$

5. CONCLUSIONS

In this paper, we have made some comments on the Morita equivalence between noncommutative and ordinary gauge theories. We present a simple prescription, whereby gauge fields and correlators of the gauge-invariant observables in the $U(1)$ noncommutative YM theory on a torus at a rational θ-parameter value can be identified with those in the ordinary $U(N)$ or $U(N_1) \times \ldots \times U(N_r)$ YM theory with nontrivial boundary conditions on the dual torus. The size of the dual torus is determined by the requirement for the Morita map to be single-valued. We also show that under the Morita map, the Polyakov loops in the ordinary YM theory are converted to the open noncommutative Wilson loops.[4]

An open question concerns the generalization of the Morita equivalence to boundary conditions of the non-twist-eater type. Another interesting direction is to link three different descriptions of the Morita equivalence: the field theory approach using the Fourier components, the string theory approach using *T*-duality and the brane language [13, 14], and the mathematical approach via twisted bundles over the noncommutative torus [4, 15].

ACKNOWLEDGMENTS

I am grateful to D. Belov and N. Nekrasov for important comments reviving the interest in the subject. I thank S. Gorsky for numerous discussions and correspondence, and A. Morozov for his interest in this work and encouragement. I am also grateful to I. Polyubin, A. Rosly and K. Selivanov for useful comments on the manuscript. I acknowledge Yu. Makeenko for helpful discussion. I would like to thank I. Vashkevich for technical support. The work was partly supported by the Russian President's grant no. 00-15-99296 and by Russian Foundation for Basic Research, grant no. 98-02-16575.

[4] This fact was first mentioned in [12].

REFERENCES

1. A. Connes, M. R. Douglas, and A. Schwarz, J. High Energy Phys. **9802**, 003 (1998); hep-th/9711162.
2. N. Seiberg and E. Witten, J. High Energy Phys. **9909**, 032 (1999); hep-th/9908142.
3. N. Nekrasov and A. Schwarz, Commun. Math. Phys. **198**, 689 (1998); hep-th/9802068.
4. A. Schwarz, Nucl. Phys. B **534**, 720 (1998); hep-th/9805034.
5. G. Landi, F. Lizzi, and R. J. Szabo, hep-th/9912130.
6. G. 't Hooft, Nucl. Phys. B **138**, 1 (1978); A. Belavin, Pis'ma Zh. Éksp. Teor. Fiz. **32**, 182 (1980) [JETP Lett. **32**, 169 (1980)].
7. D. B. Fairlie and C. K. Zachos, Phys. Lett. B **224**, 101 (1989); D. B. Fairlie, P. Fletcher, and C. K. Zachos, J. Math. Phys. **31**, 1088 (1990).
8. P. van Baal and B. van Geemen, J. Math. Phys. **27**, 455 (1986); D. R. Lebedev and M. I. Polikarpov, Nucl. Phys. B **269**, 285 (1986).
9. A. Gonzalez–Arroyo, hep-th/9807108.
10. N. Ishibashi, S. Iso, H. Kawai, and Y. Kitazawa, hep-th/9910004.
11. J. Ambjorn, Y. M. Makeenko, J. Nishimura, and R. J. Szabo, hep-th/0002158; J. Ambjorn, Y. M. Makeenko, J. Nishimura, and R. J. Szabo, JHEP **9911**, 029 (1999); hep-th/9911041.
12. J. Ambjorn, Y. M. Makeenko, J. Nishimura, and R. J. Szabo, hep-th/0004147.
13. B. Pioline and A. Schwarz, J. High Energy Phys. **9908**, 021 (1999); hep-th/9908019.
14. R. Cai and N. Ohta, J. High Energy Phys. **0003**, 009 (2000); hep-th/0001213.
15. C. Hofman and E. Verlinde, Nucl. Phys. B **547**, 157 (1999); hep-th/9810219.

Journal of Experimental and Theoretical Physics, Vol. 91, No. 4, 2000, pp. 658–660.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 4, 2000, pp. 761–763.

ATOMS, SPECTRA, RADIATION

Limitations Due to the Spontaneous Emission of Low-Frequency Photons on the Applicability of the Newtonian Description of Particle Motion in a Gas

M. E. Gertsenshteĭn[a] and Yu. A. Kravtsov[b, *]

[a]*Institute of Nuclear Physics, M. V. Lomonosov Moscow State University, Moscow, 119899 Russia*
[b]*Space Research Institute, Russian Academy of Sciences, Moscow, 117810 Russia;*
Space Research Center, Polish Academy of Sciences, 00716 Warsaw, Poland
**e-mail: kravtsov@asp.iki.rsi.ru*

Received March 28, 2000

Abstract—It is shown that the spontaneous emission of low-frequency photons in collision of atoms and molecules in a gas is very rapidly, in several free-transit intervals, radically changes the trajectories and the order of the collisions of the particles as compared with the predictions of the equations of classical mechanics. *© 2000 MAIK "Nauka/Interperiodica".*

The model of absolute elastic interaction of molecules in a gas is so deeply ingrained in classical theory that a discussion of the question of the limits of the predictability of this model could seem inappropriate. At the same time, the classical theory by no means reflects all properties of real atoms and molecules, consisting of charged particles, and cannot claim to give a universal description of the behavior of a system of atoms and molecules. Specifically, the classical model of collisions completely neglects the quantum-mechanical spreading of the wave packets associated with the particles [1–4], it ignores the capability of particles to become polarized in an external electromagnetic field (radical changes in the trajectories of particles under the action of a thermoelectromagnetic field were described in [5]), and finally it does not admit radiation of electromagnetic waves by the colliding particles. The latter limitation was pointed out in [6–8], where the question of the radiation of low-frequency photons in a gas at moderate temperatures was discussed.

It is obvious that the effects enumerated above (quantum-mechanical spreading of wave packets, action of a thermal electromagnetic field, and quantum emission of low-frequency photons) will ultimately destroy the regularities dictated by classical mechanics. This makes it important to estimate the time during which the system still follows the classical description.

The present paper is devoted to an analysis of the influence of the emission of low-frequency photons on the trajectory of colliding gas molecules ("photon mechanism" of the perturbation of trajectories). We shall estimate the time interval in which the Newtonian predictions concerning the particle trajectories and the sequence of particle collisions become meaningless, and we shall show that the duration of this interval is only several free-path intervals. In addition, we shall note the direct relation between the quantum effect discussed here and the problem of substantiating statistical physics.

At moderate temperatures, $T \sim 300$ K, the kinetic energy of gas molecules $kT \sim (1/40)$ eV is small compared with the energy of photons corresponding to optical transitions. However, this energy is sufficient for emission of low-frequency (IR and microwave) photons with energy $\hbar\omega \le kT$ (such low-frequency photons correspond to frequencies $\nu = \omega/2\pi < 6 \times 10^{12}$ Hz and wavelengths $\lambda > 50$ μm).

The emission of low-frequency photons can arise as a result of two types of weakly inelastic processes. In the first place, during collisions the electronic shells of atoms and molecules become deformed, as a result of which the particles acquire electric dipole moments which vary with characteristic collision time $\Delta t_c \approx a/v_T$, where a is the diameter of the molecules and v_T is the thermal velocity of the molecules. The collision durations Δt_c correspond to characteristic radiation frequency $\nu_c = \omega_c/2\pi \approx 1/\Delta t_c \approx v_T/a$. For air at normal conditions (T = 300 K, pressure 1 bar, molecular mass about 30) the collision duration Δt_c is estimated to be 0.5×10^{-12} s, and the characteristic radiation frequency is estimated to be $\nu_c \approx 2 \times 10^{12}$ s^{-1}. The corresponding wavelength $\lambda_c \approx 0.15$ mm lies in the submillimeter range of the electromagnetic spectrum. In the second case during collisions of molecules quantum transitions can be excited in the vibrational or rotational spectra of the molecules with frequencies less than ν_c. A large number of such transitions lies in the microwave range (frequencies 10^9–3×10^{10} s^{-1}, wavelength 1–30 cm).

The low-frequency photons arising as a result of weakly inelastic collisions of gas particles carry off

1063-7761/00/9104-0658$20.00

momentum $p_{ph} = \hbar\omega/c$ and thereby perturb the trajectories of the receding particles. The initial angular perturbation $\Delta\theta_0$ of the trajectories accompanying the emission of a single photon is proportional to the Planck constant $\hbar$ and can be estimated as

$$\Delta\theta_0 \approx \frac{p_{ph}}{p_T} \approx \frac{\hbar\omega}{cmv_T} \approx \frac{3}{2}\frac{\hbar\omega}{kT}\frac{v_T}{c}, \quad (1)$$

where $p_T = mv_T$ is the typical momentum of the thermal motion of the molecules, and kT is the average translational kinetic energy of the molecules.

As a result of the exponential instability of the particle trajectories in a gas (a description of this instability, responsible for the appearance of molecular chaos, can be found in, for example, [9]) the angular deviation $\Delta\theta$ increases in each collision by a factor of l/a, where a is the particle radius and l is the average mean-free path length. After M collisions the angular deviation $\Delta\theta$ of a relatively unperturbed trajectory is

$$\Delta\theta_M \approx \Delta\theta_0\left(\frac{l}{a}\right)^M. \quad (2)$$

In air at normal conditions the ratio l/a is 10^4, so that the formula (2) describes the mechanism of the giant enhancement of fluctuations in a gas: after M collisions the angular deviation of the molecules increases by a factor of 10^{4M}!

The situation where the angular deviation $\Delta\theta$ of the trajectory reaches the angular diameter a/l of a molecule is critical: after the critical angle $\Delta\theta_c \approx a/l$ is reached, the order of the collisions between the molecules changes radically compared with the scenario prescribed by Newton's equations and the classical model of absolutely elastic collisions. For $\Delta\theta < \Delta\theta_c$ the perturbations of the molecular trajectories are small, whereas when the critical value $\Delta\theta_c$ is reached the predictions based on Newton's equations and the model of absolutely elastic collisions become meaningless, since for $\Delta\theta > \Delta\theta_c$ the form of the particle trajectories changes radically and no longer resembles the trajectory of the unperturbed motion.

The critical number M_c of collisions corresponding to attainment of the critical perturbation $\Delta\theta_c \approx a/l$ is determined from the condition $\Delta\theta_M \approx \Delta\theta_c \approx a/l$:

$$M_c = -\frac{\log\Delta\theta_0}{\log l/a} - 1. \quad (3)$$

When submillimeter, millimeter, and centimeter range photons are emitted, the critical number of collisions is surprisingly small: $M_c \leq 2$. It is helpful to note that the effect of 300-degree thermal electromagnetic field gives comparable values: $M_c \approx 2.5$ [5].

The estimates presented above mean that the "photon mechanism" for perturbation of the classical trajectories, just as the action of an external thermal electromagnetic field, limits the applicability of the Newtonian description of the motion of molecules only to several free-path intervals of particles after the emission of a photon.

The surprising thing is not so much that Newton's equations are of limited applicability, so much as the time interval during which the equations are valid is extremely short. Here it is relevant to note that under the conditions of the mechanism of gigantic amplification of fluctuations which operates in the gas any, even the weakest, factor falling outside the limits of the Newton paradigm will be manifested in a microscopically short time. Thus, the inclusion of an additional factor which is 1000 orders of magnitude (i.e., a factor of 10^{1000}!) weaker than the Newtonian forces will be manifested in only $\log(10^{1000})/\log(l/a) = 1000 : 4 = 250$ collisions [5]. This estimate shows that the Newtonian description of the trajectories of molecules of a gas on any long observation times must be treated with great skepticism.

The subject of the present paper has a direct bearing on the discussion of the problem of substantiating statistical physics, the importance of which was recently underscored by Ginzburg [10]. Zaslavskiĭ [9] formulated the basic question of the discussion as follows: "Whence does chaos appear in a system of colliding particles, making it possible to use various probabilistic methods to describe the system?" In our view, the spontaneous emission of low-frequency photons can easily play the role of the factor which introduces random perturbations into the Newtonian equations of motion. Under the conditions of giant amplification of fluctuations, these perturbations produce the conditions for a transition from a dynamical to a probabilistic description.

The photon mechanism for perturbation of particle trajectories is unlikely to be strongly reflected in the statistical characteristics of a gas, such as the velocity distribution of the particles, the correlation functions of the positions of the molecules, and so on. However, the photon mechanism is of fundamental importance for solving the question of how rapidly the dynamical description of the motion of molecules on the basis of Newton's equations gives way to the probabilistic description on the basis of Boltzmann's kinetic equation. The estimates presented above show that the transition from the reversible Newton equations to the irreversible kinetic equation can occur in very short times of the order of several free-path intervals.

Thus, the quantum mechanism, giving a transition from the Newtonian description to the coarse probabilistic description, is incorporated in the nature of the phenomena itself.

ACKNOWLEDGMENTS

In conclusion, we express our deep appreciation to O.A. Chichigina for diverse assistance in the prepara-

tion of this paper and to Yu.A. Kukharenko and G.É. Norman for their interest in this work.

The work of one of us (Yu. A. K.) was supported by the Federal Target Program “Integratsiya” (project no. A-0030/90).

REFERENCES

1. A. S. Kaklyugin and G. É. Norman, in *Thermodynamics of Irreversible Processes* (Nauka, Moscow, 1987), p. 5.
2. A. S. Kaklyugin and G. E. Norman, J. Mosc. Phys. Soc. **5**, 167 (1995).
3. A. S. Kaklyugin and G. E. Norman, J. Mosc. Phys. Soc. **8**, 283 (1998).
4. A. S. Kaklyugin and G. É. Norman, Ross. Khim. Zh. **44** (3), (2000) (in press).
5. Yu. A. Kravtsov, Zh. Éksp. Teor. Fiz. **96**, 1661 (1989) [Sov. Phys. JETP **69**, 940 (1989)].
6. M. E. Gertsenshteĭn, I. A. Boloshin, and M. P. Suvorov, Izv. Vyssh. Uchebn. Zaved., Fiz., No. 2, 119 (1996).
7. M. E. Gertsenshteĭn, Nauka Tekhnol. Ross., No. 5, 16 (1997).
8. M. E. Gertsenshteĭn and Yu. A. Kravtsov, Nauka Tekhnol. Ross., No. 6, 9 (1998).
9. G. M. Zaslavsky, *Chaos in Dynamical Systems* (Nauka, Moscow, 1984; Harwood, Chur, 1985).
10. V. L. Ginzburg, Usp. Fiz. Nauk **169**, 419 (1999).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 4, 2000, pp. 661–672.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 4, 2000, pp. 764–776.

ATOMS, SPECTRA, RADIATION

On Excitation and Loss of an Electron by Incident Ions in Relativistic Collisions with Atoms

A. B. Voĭtkiv[a, b, *], N. Grün[b], and W. Scheid[b]

[a] *Arifov Institute of Electronics, Academy of Sciences of the Republic of Uzbekistan, Tashkent, 700143 Uzbekistan*

[b] *Institute for Theoretical Physics (Theorie II) at the Justus-Liebig-University Giessen, Heinrich-Buff-Ring 16, D-35392 Giessen, Germany*

**e-mail: Alexander.Voitkiv@theo.physik.uni-giessen.de*

Received July 23, 1999

Abstract—The excitation and loss of an electron by ions in relativistic collisions with atoms are studied in first-order perturbation theory. General expressions are obtained for the cross sections for the excitation and loss of an electron. In the limit of nonrelativistic collision velocities these expressions pass into the well-known nonrelativistic results. It is shown that, in contradistinction to the nonrelativistic collisions, in ultrarelativistic collisions the screening of the nucleus of the target atom by the atomic electrons is very important for excitation and loss of an electron by ions even for collisions of heavy ions with light atoms. Our computational results for the cross section for electron loss are in good agreement with existing experimental data. *© 2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The excitation and loss of electrons by incident particles in nonrelativistic collisions with atomic targets have been studied quite intensively for the last several decades (see, e.g., [1–3] and the literature cited there). This work has resulted in a fairly good understanding of these processes, especially the processes that can be described on the basis of first-order perturbation theory in the interaction of the colliding composite particles. In addition, in the last few years certain approaches which fall outside the scope of first-order perturbation theory have been used to investigate these processes (see, e.g., [4] and the literature cited there).

In first-order perturbation theory the excitation (loss) of an electron by an incident particle in a collision with an atom is conventionally divided into so-called screening and antiscreening processes (the terms "elastic" (for an atom) and "coherent" mode are used for the first process and "inelastic" and "incoherent" mode are used for the second process; see, e.g., [1]). In the first case the transition of an electron between the states of the incident particle occurs as a result of the interaction of this electron with the nucleus of the target atom, screened by "passive" atomic electrons which remain in their initial atomic state. In the second case the process leading to the excitation (loss) of an electron by the incident particle proceeds by "direct" interaction of this electron with the atomic electrons, which as a result of this undergo transitions into excited states (including the continuum) of the atom.

In sharp contrast to the state of the theory of nonrelativistic collisions, the theory of relativistic collisions of two composite atomic particles, each of which carries an electron(s), has not been formulated even in first-order perturbation theory (see, e.g., the discussion in the monograph [2, pp. 133–135]). Only a few attempts have been made to give a description of the process leading to the loss of an electron by an ion in relativistic collisions with atoms. One was made in [5, 6], where, essentially, only the contribution to the cross section for the loss of an electron from an elastic mode was estimated. The method used in these works is based on the first-order perturbation theory for ionization and excitation of *K*-shell electrons in relativistic collisions with structureless point charges [7, 8]. In order to take into account the fundamental difference between collisions with a point charge and a neutral atom the authors of [5, 6] used the well-known results for losses of electrons by ions in nonrelativistic collisions with neutral atoms, introducing in the process certain intuitive assumptions in order to adapt this nonrelativistic case to relativistic collisions. A more complete set of results obtained for the cross sections for electron losses by this method is presented in [6], where the cross sections for losses of an electron by incident ions were calculated for various projectile-ion–target-atom pairs right up to very high collision energies corresponding to $\gamma \leq 1000$, where γ is the Lorentz factor.

The capture and loss of an electron by Pb ions, passing through various solid-state targets (from Be to Au) with collision energy $E = 160$ GeV/au, were studied

1063-7761/00/9104-0661$20.00

experimentally in [9]. The results of this experiment for the electron loss cross sections differ substantially from the theoretical predictions [6].

In [10] a simple semiquantitative analysis for the contribution of screening to the total electron loss cross section was given on the basis of first-order perturbation theory. In the approach employed in [10] this part of the total loss cross section is divided into contributions from "close" and "distant" collisions. The contribution from "close" collisions to the cross section was calculated in the approximation of a "binary" collision between the electron of the incident ion and the nucleus of the neutral target atom. The contribution of distant collisions to the cross section was estimated by the method of equivalent photons. Since the contribution of antiscreening to the electron loss cross section in principle cannot be studied in this manner, this part of the loss cross section was estimated in [10] using the relation between the screening and antiscreening cross sections that follows from the so-called free collisions approximation first introduced by Bohr (see, e.g., [1, 2]). The estimates obtained in [10] for the loss cross sections are in reasonable agreement with experiment [9].

The present work is an attempt to give on the basis of the first Born approximation a more accurate and general description of the processes leading to the excitation and loss of electrons by ions in relativistic collisions with atoms. Our analysis is limited primarily to ions with only one electron. The criteria for the applicability of first-order perturbation theory for analysis of electron transitions between the levels of the ion because of the interaction with the atom can be formulated as follows:

(1) $Z_I < Z_A$, where Z_I is the ion charge and Z_A is the nuclear charge of the atom and

(2) $Z_A < v$, where v is the collision velocity. This condition (2) is applicable for any ratio of Z_I and Z_A, and it holds for relativistic collisions for any possible values of Z_A.

This paper is organized as follows. A general analysis of the excitation and loss of electrons by ions in relativistic collisions with atoms is given in Section 2. The nonrelativistic-atom approximation is introduced in Section 3, and the contributions to the cross section for the excitation and loss of an electron from elastic and inelastic collision modes are calculated on the basis of this approximation. The numerical results are presented in Section 4, where they are compared with existing experimental data [9, 11] and computational results [6, 10].

A four-dimensional metric, determined by the metric tensor $g_{\mu\nu}(\mu, \nu = 0, 3)$ with the components $g_{00} = -g_{11} = -g_{22} = -g_{33} = 1$ and $g_{\mu\nu} = 0$, $\mu \neq \nu$, is used in this paper. Unless stated otherwise, atomic units are employed.

2. GENERAL ANALYSIS

Since collisions in which the nuclei of the colliding atomic particles (ion and atom) penetrate into one another make a negligibly small contribution to the cross section of atomic processes, such collisions and, correspondingly, nonelectromagnetic forces can be neglected in this problem. Then, in first-order perturbation theory the S matrix element of the transition has the form (see, e.g., [12])

$$S_{fi} = -\frac{i}{c}\int d^4x J^I_\mu(x) A^\mu_A(x), \tag{1}$$

where $J^I_\mu(x)$ is a four-dimensional electromagnetic transition current for an ion in a space-time point x, $A^\mu_A(x)$ is the four-dimensional potential of the electromagnetic field produced by an atom at the same point x, and c is the velocity of light in vacuum. In Eq. (1) and below summation over repeated Greek indices is assumed. The potential is calculated from the Maxwell equation

$$\Box A^\mu_A(x) = -\frac{4\pi}{c} J^I_A(x), \tag{2}$$

where $J^\mu_A(x)$ is the four-dimensional transition current of the atom.

Since the nuclear and atomic energy scales are strongly different, the contribution of Coulomb collisions between ion and atom accompanied by excitation and breakup of nuclei to the process of electron excitation and loss can be neglected. Consequently, the nuclei of the colliding particles can be treated as structureless point charges. In addition, simple estimates show that in a coordinate system where the nucleus of one of the colliding particles is initially at rest, the typical velocity of the nucleus after a collision is much smaller than the characteristic atomic velocity $v_0 \sim 1$ au. On this basis, a convenient method for calculating the matrix element (1) is as follows. First, we calculate the current $J^I_\mu(x)$ in the coordinate system K_I, where the ion is initially at rest. Second, we determine the current $J^\mu_A(x_A)$ in the coordinate system K_A, where the atom was initially at rest, and we find in this system the value of the four-potential $A^\mu_A(x_A)$. Then, using the transformation properties of the potential, we transform this potential into the coordinate system K_I and calculate there the transition matrix elements and the corresponding cross sections.

Assuming that the ion carries only one electron, the transition ionic current in the coordinate system K_I can be written as

$$J^I_\mu(x) = \int d^3R_I \int d^3r \bar{\Psi}_f(\mathbf{R}_I, \mathbf{r}, t) j^I_\mu \Psi_i(\mathbf{R}_I, \mathbf{r}, t), \quad (3)$$

where j^I_μ is the current density operator. Since this operator is local, it can be written (in the first-quantization formalism) as

$$j^I_\mu = Z_I v^I_\mu \delta^3(\mathbf{x} - \mathbf{R}_I) - c\gamma_\mu \delta^3(\mathbf{x} - \mathbf{R}_I - \mathbf{r}), \quad (4)$$

where Z_I is the atomic number of the ion, $\mathbf{R}_I$ is the coordinate of the nucleus of the ion, $\mathbf{r}$ is the coordinate of the electron on the ion with respect to the nucleus of this ion, γ_μ are the Dirac γ matrices for the electron, and $\delta^{(3)}$ is a three-dimensional δ function. Since in the coordinate system K_I the three-dimensional velocity of the ion is zero in its initial state and negligibly small in the final state, the four-component quantity v^I_μ appearing in Eq. (4) can be written as $v^I_\mu = (c, 0, 0, 0)$. In Eq. (4) we neglected the nuclear spin in the nuclear part of the current operator. The justification for this step is that the difference between the masses of the electrons and nuclei is extremely large (see, e.g., the discussion of the method of equivalent photons [13]).

In Eq. (3) the wave function of the initial and final states of the ion has the form

$$\Psi_{i,f}(\mathbf{R}_I, \mathbf{r}, t) = \frac{1}{\sqrt{V_I}} \exp(i\mathbf{p}_{i,f} \cdot \mathbf{R}_I - i\varepsilon_{i,f} t) \psi_{0,n}(\mathbf{r}). \quad (5)$$

The indices i and f in Eq. (5) denote, respectively, the initial and final states of the ion, $\mathbf{p}_{i,f}$ is the total three-dimensional momentum ($\mathbf{p}_i = 0$), $\varepsilon_{i,f}$ is the total energy (including the rest energy) of the ion, $\psi_{0,n}$ are the relative Dirac bispinors describing the initial and final internal states of the ion, and V_I is the normalization volume for the plane wave describing the free motion of the ion before and after the collision. In what follows we shall be interested only in collisions in which the internal state of the ion changes, $n \neq 0$. For electronic excitation the final state ψ_n is a discrete state of the ion. In the opposite case the state ψ_n is a state in the continuous spectrum of the ion, normalized appropriately, and describes the loss of an electron by the ion.

The ansatz (5) represents the standard form of the wave function for a free atomic system moving with nonrelativistic velocity[1] (see, e.g., [14]), where we have neglected the difference, which is unimportant for the problem at hand, between the coordinate of the ion nucleus and the coordinate of the center of mass of the ion.

Substituting the expressions (4) and (5) into Eq. (3) and performing an elementary integration over the coordinate $\mathbf{R}_I$, we obtain (for $n \neq 0$)

$$J^I_\mu(x) = \frac{cF^I_\mu(n0; \mathbf{p}_i - \mathbf{p}_f)}{V_I} \times \exp[i(\mathbf{p}_i - \mathbf{p}_f) \cdot \mathbf{x} - i(\varepsilon_i - \varepsilon_f)t]. \quad (6)$$

We shall call the four-component quantity

$$F^I_\mu(n0; \mathbf{p}_i - \mathbf{p}_f) = -\int d^3r \bar{\psi}(\mathbf{r}) \times \exp[i(\mathbf{p}_i - \mathbf{p}_f) \cdot \mathbf{r}] \gamma_\mu \psi_0(\mathbf{r}) \quad (7)$$

with the components

$$F^I_0(n0; \mathbf{p}_i - \mathbf{p}_f) = -\int d^3r \psi^\dagger_n(\mathbf{r}) \times \exp[-i(\mathbf{p}_i - \mathbf{p}_f) \cdot \mathbf{r}] \psi_0(\mathbf{r}), \quad \mu = 0,$$
$$F^I_l(n0; \mathbf{p}_i - \mathbf{p}_f) = -\int d^3r \psi^\dagger_n(\mathbf{r}) \times \exp[-i(\mathbf{p}_i - \mathbf{p}_f) \cdot \mathbf{r}] \alpha_l \psi_0(\mathbf{r}), \quad \mu = l = 1, 2, 3, \quad (8)$$

the (inelastic) form factor of the ion. In Eq. (7) α_l are the Dirac α matrices. It is easy to see from Eq. (7) that although the quantity F^I_μ itself does not possess the transformation properties of a relativistic 4-vector, the quantity F^I_μ / V_I is such a vector.

We shall now calculate the potential produced by the atom. We shall find the atomic current $J'^\mu_A(x_A)$ in the coordinate system K_A and obtain in the system the potential $A'^\mu_A(x_A)$, where $x_A = (ct_A, \mathbf{x}_A)$ is the space-time coordinate in the system K_A. Using a method which is completely analogous to the method used above to obtain the ion current, we obtain for the atomic 4-current of the transition

$$J'^\mu_A(x_A) = c\frac{F^\mu_A(m0; \mathbf{P}'_i - \mathbf{P}'_f)}{V'_A} \times \exp[i(\mathbf{P}'_i - \mathbf{P}'_f) \cdot \mathbf{x}_A - i(E'_i - E'_f)t_A], \quad (9)$$

where $\mathbf{P}'_{i,f}$ are the three-dimensional momenta and $E'_{i,f}$ are the total energies (including the rest energy) of the atom, respectively, in the initial and final states, and V'_A is the normalization volume for the atom in the coordinate system K_A. The components of the atomic form factor F^μ_A are determined as follows:

[1] Since the plane wave $\exp(kx)$ has a Lorentz-covariant form, the expression (5) can also be used to describe a free atomic system moving with relativistic velocity, if the spin of the nucleus of the atomic system is not important for the collision process.

$$F_A^0(m0;\mathbf{Q}) = Z_A\delta_{m0} - \int\prod_{i=1}^{N_A} d^3\xi_i u_m^\dagger(\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_{N_A}) u_0(\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_{N_A}) \times\sum_{i=1}^{N_A}\exp(-i\mathbf{Q}\cdot\boldsymbol{\xi}_i),$$
$$F_A^l(m0;\mathbf{Q}) = -\int\prod_{i=1}^{N_A} d^3\xi_i u_m^\dagger(\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_{N_A}) \times\sum_{i=1}^{N_A}\alpha_{l(i)}\exp(-i\mathbf{Q}\cdot\boldsymbol{\xi}_i) u_0(\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_{N_A}),\ l=1,2,3. \quad (10)$$

Here Z_A is the atomic number, N_A is the number of electrons in the atom (for a neutral atom $N_A = Z_A$), $\alpha_{l(i)}$ are the Dirac α matrices for the ith atomic electron, $u_{0,m}$ are the wave functions describing the initial and final internal states of the atom, and $\boldsymbol{\xi}_i$ is the coordinate of the ith atomic electron with respect to the atomic nucleus. Since the motion of the internal electrons in heavy atoms is relativistic, it is assumed in the general case that $u_{0,m}$ are the relativistic wave functions.

The equations (9) and (10) were obtained under the same approximations as Eqs. (6) and (7). The only substantial difference between the form factors (7) and (10) is that for an atom we take account of the possibility that it remains in the initial state u_0 after a collision. The zeroth components of the expressions (7) and (10) have the familiar form of the form factors that appear in the description of the excitation and loss of an electron by atoms in nonrelativistic collisions (see, e.g., [1]). The three other components of these form factors have no analogs in the nonrelativistic theory. We note that F_A^μ/V_A' is a 4-vector.

To solve Eq. (2) it is convenient to use the four-dimensional Fourier transform:

$$A_A'^\mu(x_A) = \frac{1}{(2\pi)^2}\int d^4k B_A^\mu(k)\exp(-ikx_A),$$
$$J_A'^\mu(x_A) = \frac{c}{V_A'}\int d^4k\exp(-ikx_A) \times\delta^{(4)}(k+P_f-P_i)F_A^\mu(m0;\mathbf{k}). \quad (11)$$

In Eq. (11) kx_A is the scalar product of two four-dimensional vectors k and x_A, and $\mathbf{k}$ is the "spatial" part of k. Substituting the expressions (11) into the equation for the Fourier transform $B_A^\mu(k)$,

$$\Box' A_A'^\mu(x_A) = -\frac{4\pi}{c}J_A'^\mu(x_A),$$

we find

$$B_A^\mu(k) = 4\pi\frac{(2\pi)^2\delta^{(4)}(k+P_f-P_i)}{k^2-i0}\frac{F_A^\mu(m0;\mathbf{k})}{V_A'}. \quad (12)$$

Correspondingly, we obtain for the 4-potential

$$A_A'^\mu(x_A) = 4\pi \times\frac{\exp[-i(P_i'-P_f')x_A]}{(P_i'-P_f')^2-i0}\frac{F_A^\mu(m0;\mathbf{P}_i'-\mathbf{P}_f')}{V_A'}. \quad (13)$$

In Eq. (13) the term $-i0$ gives the rule for handling the singularity in the denominator (see, e.g., [12]). This singularity appears if

$$(P_i'-P_f')^2 = (\mathbf{P}_i'-\mathbf{P}_f')^2-(E_i'-E_f')^2/c^2 = 0.$$

Let $a_{\mu\nu}$ be the matrix of the Lorentz transformation from the coordinate system K_A into the coordinate system K_I. Then we have for the potential produced by an atom in the coordinate system K_I

$$A_A^\mu(x) = a_\nu^\mu A_A'^\nu(a^{-1}x) = 4\pi\frac{\exp[-i(P_i-P_f)x]}{(P_i-P_f)^2-i0}a_\nu^\mu\frac{F_A^\nu(m0;\mathbf{Q})}{\gamma V_A}. \quad (14)$$

In Eq. (14) $P_{i(f)}$ is the initial (final) 4-momentum of an atom in the system K_I, $V_A = V_A'/\gamma$ is the normalization volume of the atom in the system, $\gamma = 1/\sqrt{1-v^2/c^2}$ is the Lorentz factor, and $\mathbf{v}$ is the velocity of the incident atom in the coordinate system K_I (collision velocity).

The momentum transfer $\mathbf{Q} = \mathbf{P}_i' - \mathbf{P}_f'$ appearing in Eq. (14) can be rewritten as

$$\mathbf{Q} = \left(\mathbf{P}_{i\perp}-\mathbf{P}_{f\perp}, \frac{1}{\gamma}(\mathbf{P}_{i\|}-\mathbf{P}_{f\|})+\frac{v}{c^2}(E_f'-E_i')\right) \approx\left(\mathbf{P}_{i\perp}-\mathbf{P}_{f\perp}, \frac{1}{\gamma}(\mathbf{P}_{i\|}-\mathbf{P}_{f\|})+\frac{v}{c^2}(\epsilon_m-\epsilon_0)\right), \quad (15)$$

where $\mathbf{P}_\perp$ is the transverse (perpendicular to the collision velocity) and $\mathbf{P}_\|$ is the longitudinal (parallel to the collision velocity) components of the three-dimensional momentum $\mathbf{P}$ of an atom in the system K_I, and ϵ_0 and ϵ_m are the energies of the electrons of the atom in the initial (0) and final (m) (internal) states of the atom, given in the coordinate system K_A. We note that the transverse and longitudinal components of the momentum transfer $\mathbf{P}_i - \mathbf{P}_f$ appear differently in the atomic form factor in Eq. (14); this leads to the appearance of important features in the compensation of the field of the atomic nucleus by the fields of the atomic electrons in ultrarelativistic collisions.

Substituting the right-hand sides of Eqs. (6) and (14) into Eq. (1) and integrating over d^4x we obtain

$$S_{fi} = i\frac{4\pi}{V_I V_A}(2\pi)^4\delta^{(4)}(p_i + P_i - p_f - P_f)G_{fi}, \quad (16)$$

where

$$G_{fi} = \frac{F^I_\mu(n0; \mathbf{P}_f - \mathbf{P}_i)\gamma^{-1}a^\mu_\nu F^\nu_A(m0; \mathbf{Q})}{(P_i - P_f)^2 - i0}. \quad (17)$$

We recall that the form factors F^I_μ and F^ν_A, given by the expressions (7) and (10), were determined in the coordinate systems K_I and K_A, respectively.

Using the standard technique (see, e.g., [12]) for obtaining the cross section of a process on the basis of the known S matrix element of the transition, we obtain for the cross section of the process in which an ion and atom undergo the transitions $\psi_0 \longrightarrow \psi_n$ and $u_0 \longrightarrow u_m$, respectively,

$$\sigma^{0\to m}_{0\to n} = \frac{4}{v^2}\frac{E_i + \varepsilon_i - \varepsilon_f}{E_i}$$

$$\times\sum_{s_I}\sum_{s_A}\int d^2q_\perp|G_{fi}|^2 = \frac{4}{v^2}\frac{E_i + \varepsilon_i - \varepsilon_f}{E_i}$$

$$\times\sum_{s_I}\sum_{s_A}\int d^2q_\perp\left|F^I_\mu(n0; -\mathbf{q}_\perp, -q_{\min})\right. \quad (18)$$

$$\left.\times\gamma^{-1}a^\mu_\nu F^\nu_A\left(m0; \mathbf{q}_\perp, \frac{q_{\min}}{\gamma} + \frac{v}{c^2}(\epsilon_m - \epsilon_0)\right)\right|^2$$

$$\times\left[q^2_\perp + q^2_{\min} - \frac{(\varepsilon_f - \varepsilon_i)^2}{c^2}\right]^{-2},$$

where E_i and $E_f = E_i + \varepsilon_i - \varepsilon_f$ are the total initial and final energies of the atom in the coordinate system K_I. The summation in Eq. (18) means summation over spins of the electrons in the ion and atom. This summation presumes averaging over the projections of the spins of the electrons in the initial state and summation over the projections of the spins in the final state. In Eq. (18) the integration over the absolute value of the transverse component $\mathbf{q}_\perp$ of the moment transfer extends from 0 up to a maximum value $q^{\max}_\perp$, which in our case, as usual, can be set equal to infinity (see, e.g., [2]). The factor $(E_i + \varepsilon_i - \varepsilon_f)/E_i$ in Eq. (18) can be replaced by 1 to the same accuracy. The minimum value of the moment transfer $q_{\min} = |\mathbf{P}_i| - |\mathbf{P}_f|$ appearing in Eq. (18) can be determined from the law of conservation of energy for the collisions considered:

$$\varepsilon_i + \sqrt{c^2\mathbf{P}_i^2 + M^2_{Ai}c^4} = \varepsilon_f + \sqrt{c^2\mathbf{P}_f^2 + M^2_{Af}c^4}, \quad (19)$$

where M_{Ai} and M_{Af} are the rest masses of the atom in the initial u_0 and final u_m states, respectively. Because of the large difference between the electron and ion masses the recoil energy of the ion in the coordinate system K_I can be neglected, and we have for the difference of the total energies of the ion before and after a collision $\varepsilon_f - \varepsilon_i \approx \varepsilon_n - \varepsilon_0$, where $\varepsilon_{0(n)}$ is the energy of the ion electron (ion) in the initial (final) internal state $\psi_{0(n)}$ of the ion. Since (1) the change in the absolute value of the three-dimensional momentum of the atom is small compared with its initial value, $P_i + P_f \approx 2P_i$, (2) the difference in the rest masses of the atom is small compared with its initial rest mass, $M_{Ai} + M_{Af} \approx 2M_A$, and (3) the total energy of the atom in the coordinate system K_I is much greater than the difference between the final and initial energies of the ion in this system, Eq. (19) can be easily solved. As a result we have

$$q_{\min} = \frac{\varepsilon_n - \varepsilon_0}{v} + \frac{(M_{Af} - M_{Ai})c^2}{v\gamma} = \frac{\varepsilon_n - \varepsilon_0}{v} + \frac{\epsilon_m - \epsilon_0}{v\gamma}, \quad (20)$$

where $(M_{Af} - M_{Ai})c^2 = \epsilon_m - \epsilon_0$ is the difference between the final and initial energies of the atom in the coordinate system K_A.

It is also convenient to introduce the quantity

$$Q_{\min} = \frac{q_{\min}}{\gamma} + \frac{v}{c^2}(\epsilon_m - \epsilon_0) = \frac{\epsilon_m - \epsilon_0}{v} + \frac{\varepsilon_m - \varepsilon_0}{v\gamma}, \quad (21)$$

which for the atom plays the same role as the quantity $q_{\min}$ plays for the ion.

Using the expressions (20) and (21), the expression for the cross section (18) can be rewritten in a form that clearly underscores the symmetry of the parameters of the ion and atom which appear in this expression:

$$\sigma^{0\to m}_{0\to n} = \frac{4}{v^2}\sum_{S_I}\sum_{S_A}\int d^2q_\perp$$

$$\times\left|F^I_\mu\left(n0; -\mathbf{q}_\perp, -\frac{\varepsilon_n - \varepsilon_0}{v} - \frac{\epsilon_m - \epsilon_0}{v\gamma}\right)\right.$$

$$\left.\times a^\mu_\nu F^\nu_A\left(m0; \mathbf{q}_\perp, -\frac{\epsilon_m - \epsilon_0}{v} + \frac{\varepsilon_n - \varepsilon_0}{v\gamma}\right)\right|^2 \quad (22)$$

$$\left[q^2_\perp + \frac{(\varepsilon_n - \varepsilon_0 + \epsilon_m - \epsilon_0)^2}{v^2\gamma^2}\right.$$

$$\left. + 2(\gamma - 1)\frac{(\varepsilon_n - \varepsilon_0)(\epsilon_m - \epsilon_0)}{v^2\gamma^2}\right]^{-2}.$$

The expressions (18) and (22) are the first basic result of this work. If the ion and atom are in the ground states initially, then it follows from Eq. (22) that the singularity does not arise in Eqs. (13) and (14) and subsequent expressions. Since our objective is to study such collisions, we drop the term $-i0$ in Eq. (18) and subsequent expressions.

Using the explicit form of the matrix a_μ^ν (see, e.g., [2]), the relation between the form factors in Eq. (22) can be written in the symmetric form

$$F_\mu^I \gamma^{-1} a_\mu^\nu F_A^\nu = \left(F_0^I - \frac{v}{c}F_3^I\right)\left(F_A^0 - \frac{v}{c}F_A^3\right) + \frac{F_3^I F_A^3}{\gamma^2} + \frac{F_1^I F_A^1 + F_2^I F_A^2}{\gamma}. \tag{23}$$

Compared with the known form of the cross sections for excitation and loss of an electron by an ion in nonrelativistic collisions with atoms, the expressions (22) and (23) contain two types of relativistic effects. The first one is due to the magnitude of the collision velocity v and vanishes for $v/c \ll 1$. This type of effect includes the retardation of the electromagnetic field, described by the term $(\varepsilon_f - \varepsilon_i)^2/c^2 \approx (\varepsilon_n - \varepsilon_0)^2/c^2$ in the denominators in the integrands in Eqs. (18) and (22), the different dependence of the form factors of the ion and atom on the transition energies $\varepsilon_n - \varepsilon_0$ and $\epsilon_m - \epsilon_0$, as well as the relation between the zeroth and third components of the form factors in the expression (23). The second type of relativistic effect is due to the possible relativistic motion of the electrons inside an atom and ion. Consequently, it does not vanish in the limit $v/c \ll 1$. It includes the relation between the spatial components of the corresponding form factors in Eq. (23). In the limit $c \longrightarrow \infty$ both types of relativistic effects vanish and the expression (22) assumes the well-known form (see, e.g., [1, 3] and the literature cited there) of the corresponding nonrelativistic cross sections.

If after a collision the internal state of an atom is not detected, then it is necessary to sum over the entire set of possible states. Then the expression (18) gives

$$\sigma_{0\to n} = \frac{4}{v^2}\sum_{s_I}\sum_m \int d^2 q_\perp \times \left|F_\mu^I(n0; -\mathbf{q}_\perp, -q_{\min})\gamma^{-1} a_\nu^\mu F_A^\nu(m0; \mathbf{q}_\perp, Q_{\min})\right|^2 \times \left[q_\perp^2 + q_{\min}^2 - \frac{(\varepsilon_n - \varepsilon_0)^2}{c^2}\right]^{-2}. \tag{24}$$

The summation over the atomic states in Eq. (24) also assumes summation over the spins of the atomic electrons. The cross section (24) can be represented as a sum of contributions from the elastic ($m = 0$) and inelastic (all $m \neq 0$) modes. Taking account of Eq. (20), we obtain for the contribution of the elastic mode (screening cross section) to the cross section

$$\sigma_{0\to n}^s = \frac{4}{v^2}\sum_{s_I} d^2 q_\perp \times \left|F_\mu^I(n0; -\mathbf{q}_\perp, -q_{\min})\gamma^{-1} a_\nu^\mu F_A^\nu(00; \mathbf{q}_\perp, q_{\min}/\gamma)\right|^2 \times \left[q_\perp^2 + \frac{(\varepsilon_n - \varepsilon_0)^2}{v^2\gamma^2}\right]^{-2}. \tag{25}$$

Correspondingly, we have for the contribution of the inelastic mode (antiscreening cross section) to the cross section

$$\sigma_{0\to n}^a = \frac{4}{v^2}\sum_{s_I}\sum_{m\neq 0} d^2 q_\perp \times \left|F_\mu^I(n0; -\mathbf{q}_\perp, -q_{\min})\gamma^{-1} a_\nu^\mu F_A^\nu(m0; \mathbf{q}_\perp, Q_{\min})\right|^2 \times \left[q_\perp^2 + \frac{(\varepsilon_n - \varepsilon_0 + \epsilon_m - \epsilon_0)^2}{v^2\gamma^2} + 2(\gamma - 1)\frac{(\varepsilon_n - \varepsilon_0)(\epsilon_m - \epsilon_0)}{v^2\gamma^2}\right]^{-2}. \tag{26}$$

3. "NONRELATIVISTIC" ATOM APPROXIMATION

The combination (23), where the form factors of the ion and atom appear in the expression for the cross section, is quite complicated. Consequently, to obtain simpler expressions for the cross sections we introduce the following approximation: we shall neglect all spatial components of the atomic form factor. Some semiqualitative considerations to justify this are given below.

Let us consider the atomic form factor (10) in greater detail. The component $F_A^0(m0; \mathbf{Q})$ of this form factor is related with the static charge distribution inside the atom. In contradistinction, the components $F_A^l(m0; \mathbf{Q})$ are related with the current produced by the atom in the rest system of the atom. In this coordinate system such a current is related with the motion of the electrons inside the atom. The quantity $F_A^l(m0; \mathbf{Q})$ can be roughly estimated as

$$F_A^l(m0; \mathbf{Q}) \sim \frac{v_e}{c}F_A^0(m0; \mathbf{Q}),$$

where v_e is the characteristic velocity of the atomic electrons. For light and medium atoms $v_e \ll c$ for all atomic electrons, and all three components $F_A^l(m0; \mathbf{Q})$ in Eq. (10) can be neglected compared with $F_A^0(m0; \mathbf{Q})$. In heavy atoms the most strongly bound electrons can possess velocities comparable in order of magnitude to the velocity of light. However, since the relative number of such electrons is small, it can be expected that they do not appreciably increase the value of $F_A^l(m0; \mathbf{Q})$. Consequently, neglecting $F_A^l(m0; \mathbf{Q})$ seems to be a quite reasonable approximation for describing collisions with heavy atoms. In what follows, we shall employ this approximation: $F_A^l(m0; \mathbf{Q}) \approx 0$.

We shall call this approximation the nonrelativistic atom approximation (NAA). The NAA destroys symmetry in which the form factors of the atom and ion appear in the expression for the cross section. Consequently, in general it can be expected that this approximation is more suitable for the elastic mode, when the electron of the ion undergoes transitions and the electrons of the atom does not, and correspondingly the symmetry between the ion and atom is already destroyed to a certain extent. Indeed, the analysis performed above for the elastic atomic form factor shows that the NAA can be used to calculate the contribution of the elastic mode to the cross section for any possible collision energies and ion–atom pairs.

The situation becomes more complex when the NAA is used to calculate the inelastic mode. In the arguments presented above in the second paragraph, the characteristic velocity of electrons in the ground state of an atom was studied. In collisions with heavy ions the minimum momentum transfer

$$Q_{\min} = \frac{\epsilon_m - \epsilon_0}{v} + \frac{\varepsilon_m - \varepsilon_0}{v\gamma}$$

can be large compared with the typical values of the momenta of the electrons in an atom. In such collisions (in the inelastic mode) the velocities of the atomic electrons in the final state can be much greater than their typical velocities in the ground state of the atom. Since we have assumed that the atomic electrons in a collision are nonrelativistic (in the coordinate system K_A), the condition $Q_{\min} \ll m_e c^2$, where $m_e = 1$ is the rest mass of an electron, must be satisfied. Setting $\varepsilon_n - \varepsilon_0 \approx Z_I^2$, we obtain the following limit on the use of the NAA for calculating the contribution of the inelastic mode to the cross section:

$$\gamma \gg \frac{Z_I^2}{vc}.$$

This condition definitely holds for collisions with any heavy ion, for example, $\gamma > 4$.

There is also another important restriction on the use of the NAA for the inelastic mode. Our analysis of the behavior of the inelastic atomic form factor in the limit of small momentum transverse shows that

$$\frac{\varepsilon_n - \varepsilon_0}{\gamma} \gg \epsilon_m - \epsilon_0$$

for all transitions of atomic electrons, which make an appreciable contribution to the inelastic mode. This condition imposes an upper limit on the energy of collisions for which the inelastic mode can be studied on the basis of the NAA. In the present paper we are interested, first and foremost, in the excitation and loss of an electron by heavy ions. For collisions of such ions with light atoms, when the contribution of the inelastic mode to the cross section is relatively important, the condition presented above is satisfied for a very wide range of collision energies. For collisions with heavy atoms containing a large number of electrons, when the condition presented above may not be satisfied even for relatively small values of γ, it can be expected by analogy with nonrelativistic collisions that the inelastic mode does not make an appreciable contribution to the cross section for excitation and loss of an electron by an ion.

Remembering that the arguments given above for the NAA are more qualitative than quantitative, we shall employ this approximation below in our calculations.

Neglecting all spatial components, $F_A^l(m0; \mathbf{Q}) \approx 0$, we obtain for the cross section (25)

$$\sigma_{0\to n}^s = \frac{4}{v^2}\sum_{s_I} d^2 q_\perp Z_{A,\,\mathrm{eff}}^2(\mathbf{Q}_0^S) \times \left|\langle \psi_n(\mathbf{r})\left|\left(1 - \frac{v}{c}\alpha_z\right)\exp(i\mathbf{q}_0\cdot\mathbf{r})\right|\psi_0(\mathbf{r})\rangle\right|^2 \times \left[q_\perp^2 + \frac{(\varepsilon_n - \varepsilon_0)^2}{v^2\gamma^2}\right]^{-2}. \quad (27)$$

Here $\mathbf{q}_0 = (\mathbf{q}_\perp, q_{\min})$, $\mathbf{Q}_0^S = (\mathbf{q}_\perp, q_{\min}/\gamma)$, and α_z are the Dirac α matrices, and we employ the Dirac notation for the electronic state vectors. In Eq. (27)

$$Z_{A,\,\mathrm{eff}}(\mathbf{Q}) = Z_A - \langle u_0(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{N_A})| \times \sum_{j=1}^{N_A} \exp(-i\mathbf{Q}\cdot\boldsymbol{\xi}_j)|u_0(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{N_A})\rangle \quad (28)$$

is the so-called effective charge of an atom in the ground state. In contrast to collisions with a bare

atomic nucleus, this charge is a function of the momentum transfer and varies in the range $(0, Z_A)$.

Similarly, we have for the cross section (26)

$$\sigma^a_{0\to n} = \frac{4}{v^2}\sum_{s_I}\sum_{m\neq 0}\int d^2q_\perp \left|\langle u_m(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{N_A})| \times \sum_{j=1}^{N_A}\exp(-i\mathbf{Q}^a_0\cdot\boldsymbol{\xi}_j)|u_0(\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{N_A})\rangle\right|^2 \times \left|\langle\psi_n(\mathbf{r})|\left(1-\frac{v}{c}\alpha_z\right)\exp(i\mathbf{q}_0\cdot\mathbf{r})|\psi_0(\mathbf{r})\rangle\right|^2 \times\left[q_\perp^2+\frac{(\varepsilon_n-\varepsilon_0+\epsilon_m-\epsilon_0)^2}{v^2\gamma^2} + 2(\gamma-1)\frac{(\varepsilon_n-\varepsilon_0)(\epsilon_m-\epsilon_0)}{v^2\gamma^2}\right]^{-2}, \quad (29)$$

where $\mathbf{Q}^a_0 = (\mathbf{q}_\perp, Q_{\min})$.

In accordance with the expression (13) the approximation $F^l_A(m0;\mathbf{Q}) \approx 0$ used to obtain the cross sections (27) and (29) actually means that we neglected the vector potential generated by the atom in the coordinate system K_A compared with its scalar potential in the system. Then the scalar A^0 and vector $\mathbf{A}$ potentials of the atom in the coordinate system K_I are related by the simple relation $\mathbf{A} = (\mathbf{v}/c)A^0$, and Eqs. (24) and (26) reduce to Eqs. (27) and (29). The latter equations are the second basic result of this paper.

3.1. Contribution of the Elastic Mode to the Cross Section

In this section we shall examine in greater detail the contribution (27) to the total cross section for the excitation (loss) of an electron.

The effective charge (28) can be rewritten as

$$Z_{A,\,\mathrm{eff}}(\mathbf{Q}^S_0) = Z_A - \int d\boldsymbol{\xi}\rho_{el}(\boldsymbol{\xi})\exp(-i\mathbf{Q}^S_0\cdot\boldsymbol{\xi}), \quad (30)$$

where $\rho_{el}(\boldsymbol{\xi})$ is the charge of the electrons in the atom. We underscore that $\rho_{el}(\boldsymbol{\xi})$ in Eq. (30) is the density calculated in the coordinate system of the incident atom. According to [15], this density can be approximated in the form

$$\rho_{el}(\boldsymbol{\xi}) = \frac{Z_A}{4\pi\xi^2}\sum_{i=1}^{3}A_i\kappa_i^2\exp(-\kappa_i\xi). \quad (31)$$

In this expression A_i and κ_i are constants for a given atom. In [15] they are tabulated for all atomic elements. Substituting the expression (31) into Eq. (30) and performing the integration over $\boldsymbol{\xi}$, the effective charge $Z_{A,\,\mathrm{eff}}(\mathbf{Q}^s_0)$ can be rewritten as

$$Z_{A,\,\mathrm{eff}}(\mathbf{Q}^s_0) = Z_A\left(1-\sum_{i=1}^{3}\frac{A_i\kappa_i^2}{\kappa_i^2+q_\perp^2+q_{\min}^2/\gamma^2}\right) = Z_A\left(q_\perp^2+\frac{q_{\min}^2}{\gamma^2}\right)\sum_{i=1}^{3}\frac{A_i}{\kappa_i^2+q_\perp^2+q_{\min}^2/\gamma^2}. \quad (32)$$

We note that the condition $\sum_i A_i = 1$ [15] was used to obtain the last equality in Eq. (32).

Substituting the expression (32) into Eq. (27), we obtain

$$\sigma^s_{0\to n} = \frac{4Z_A^2}{v^2}\sum_{s_I}\sum_{i,j}A_iA_j\int d^2q_\perp \times\left|\langle\psi_n(\mathbf{r})|\exp(i\mathbf{q}_0\cdot\mathbf{r})\left(1-\frac{\mathbf{v}}{c}\cdot\boldsymbol{\alpha}\right)|\psi_0(\mathbf{r})\rangle\right|^2 \times\left[\left(\mathbf{q}_\perp^2+\frac{\omega_{n0}^2}{v^2\gamma^2}+\kappa_i^2\right)\left(\mathbf{q}_\perp^2+\frac{\omega_{n0}^2}{v^2\gamma^2}+\kappa_j^2\right)\right]^{-1}, \quad (33)$$

where $\omega_{n0} = \varepsilon_n - \varepsilon_0$.

If all κ_i are set equal to zero in Eq. (33) (i.e., the atom is completely devoid of electrons), then this expression is exactly the well-known form of the cross sections for the excitation and ionization in collisions with bare nuclei (see, e.g., [2]). In collisions with bare nuclei, the region of small values of $q_\perp$ ($0 \le q_\perp \lesssim \omega_{n0}/v\gamma$) makes the main contribution to these cross sections; this results in the logarithmic growth of these cross sections with increasing γ: $\sigma_{n0} \sim \ln\gamma$ (see, e.g., [2]).

For collisions with neutral atoms, Eq. (33) describes an important general feature in the screening of an atomic nucleus by the atomic electrons in relativistic collisions that does not occur in nonrelativistic collisions. It is worth underscoring the fact that this feature follows directly from Eq. (15) and, correspondingly, is not contributed by the particular model (31) chosen to describe this screening on the basis of convenience considerations. In nonrelativistic collisions, if an ion carrying the electron is a heavy ion ($Z_I \gg 1$) and the atom is a light atom (with all screening constants κ_i of the order of 1), then screening is not important, since the term $\omega_{n0}^2/v^2 \sim Z_I^4/v^2$ dominates all κ_i^2 in the denominator in the integrand in Eq. (33). However, for ultrarelativistic collisions, where $\gamma \gg 1$, the situation changes substantially. In this case, the terms κ_i^2 can be greater than $\omega_{n0}^2/v^2\gamma^2 \sim Z_I^4 v^2\gamma^2$, which results in a

large decrease of the cross section. Our calculations (see below) confirm that, in contradistinction to nonrelativistic collisions, in ultrarelativistic collisions the screening of an atomic nucleus by atomic electrons is important even for collisions of heavy-ion–light-atom pairs.

Analysis of the expression (33) shows that in collisions with neutral atoms with large values of γ the cross section $\sigma^s_{0\to n}$ approaches a constant. Thus, even for the heaviest ions the cross section $\sigma^s_{0\to n}$ becomes essentially constant for $\gamma \lesssim 100$; for lighter ions this occurs for smaller values of γ. This is at variance with the results of [6], which predicts that for any ion–atom pair $\sigma^s_{0\to n} \sim \ln\gamma$ at least for $\gamma < 1000$.

3.2. Contribution of the Inelastic Mode to the Cross Section

The equation (29) can be greatly simplified by using the so-called completeness method (see, e.g., [3] and the literature cited there). In this method the same average energy $\Delta\epsilon$ is assigned to all possible transitions of atomic electrons. It is well known that this approximation gives good results in describing fast (but nonrelativistic) collisions at velocities appreciably higher than the threshold velocity, starting with which the process of energy loss by an ion in a collision with a free electron first becomes energetically possible. By analogy to nonrelativistic collisions this method should also give reasonable results for $v \approx c$, when the kinetic energy T of the incident free electron in the rest system of the ion is appreciably higher than the binding energy of the electron in the ion:

$$T = m_e c^2(\gamma - 1) \gg |\varepsilon_0|.$$

This condition holds for any heavy ion, starting with, say, $\gamma \geq 3$–4, which is quite close to the lower limit of the region of possible values of γ, imposed by the nonrelativistic-atom approximation. The completeness method makes it possible to use the condition that the electronic states of an atom are complete,

$$\sum_m |u_m\rangle\langle u_m| = \mathbf{I}, \tag{34}$$

in order to perform the summation over the final states of the atom in Eq. (29). This gives

$$\sigma^a_{0\to n} = \frac{4}{v^2}\sum_{S_I}\int d^2q_\perp \times \left|\langle\psi_n(\mathbf{r})|\exp(i\mathbf{q}_0\cdot\mathbf{r})\left(1 - \frac{\mathbf{v}}{c}\cdot\boldsymbol{\alpha}\right)|\psi_0(\mathbf{r})\rangle\right|^2 \times \left[q_\perp^2 + \frac{(\omega_{n0} + \Delta\epsilon)^2}{v^2\gamma^2} + 2(\gamma - 1)\frac{\omega_{n0}\Delta\epsilon}{v^2\gamma^2}\right]^{-2} \times \left(\langle u_0|\sum_{i,j}\exp[-i\mathbf{Q}_0^a\cdot(\boldsymbol{\xi}_j - \boldsymbol{\xi}_i)]|u_0\rangle - \left|\langle u_0|\sum_j \exp[-i\mathbf{Q}_0^a\cdot\boldsymbol{\xi}_j]|u_0\rangle\right|^2\right), \tag{35}$$

where we have redefined $\mathbf{Q}_0^a$ as

$$\mathbf{Q}_0^a = \left(\mathbf{q}_\perp, \frac{\Delta\epsilon}{v} + \frac{\varepsilon_n - \varepsilon_0}{v\gamma}\right).$$

If the region of large (on the atomic scale) transferred momenta Q_0^a makes the main contribution to the integral on the right-hand side in Eq. (35), then only the diagonal terms ($i = j$) in the double sum in the factor in parentheses in Eq. (35) makes an appreciable contribution. The numerical value of this factor reduces in this case simply to the number of atomic electrons Z_A, and the cross section (35) describes transitions of the electron in the ion as a result of incoherent electromagnetic interaction with Z_A "free" electrons.

Just as for nonrelativistic collisions, Eq. (35) can be further simplified if the asymmetry of the ground state of the atom is ignored and the wave function of this state is written in the approximate form

$$u_0 = \prod_\lambda \phi_\lambda(\boldsymbol{\xi}), \tag{36}$$

where $\phi_\lambda(\boldsymbol{\xi})$ are single-electron orbitals. In [5] (see also [16]) it was shown that under such conditions (for neutral atoms)

$$\langle u_0|\sum_{i,j}\exp[-i\mathbf{Q}_0^a\cdot(\boldsymbol{\xi}_j - \boldsymbol{\xi}_i)]|u_0\rangle - \left|\langle u_0|\sum_j\exp[-i\mathbf{Q}_0^a\cdot\boldsymbol{\xi}_j]|u_0\rangle\right|^2 = Z_A - \sum_\lambda\left|\langle\phi_\lambda|\exp[-i\mathbf{Q}_0^a\cdot\boldsymbol{\xi}]|\phi_\lambda\rangle\right|^2. \tag{37}$$

Substituting the right-hand side of Eq. (37) into the expression (35) we obtain

$$\sigma^a_{0\to n} = \frac{4}{v^2}\sum_{S_I}\int d^2q_\perp \times \left|\langle\psi_n(\mathbf{r})|\exp(i\mathbf{q}_0\cdot\mathbf{r})\left(1 - \frac{\mathbf{v}}{c}\cdot\alpha\right)|\psi_0(\mathbf{r})\rangle\right|^2 \times \left[q_\perp^2 + \frac{(\omega_{n0} + \Delta\epsilon)^2}{v^2\gamma^2} + 2(\gamma - 1)\frac{\omega_{n0}\Delta\epsilon}{v^2\gamma^2}\right]^{-2} \times \left(Z_A - \sum_\lambda\left|\langle\phi_\lambda|\exp[-i\mathbf{Q}_0^a\cdot\boldsymbol{\xi}]|\phi_\lambda\rangle\right|^2\right). \tag{38}$$

Anholt [5] pointed out that

$$\sum_\lambda \langle\phi_\lambda|\exp[-i\mathbf{Q}\cdot\boldsymbol{\xi}]|\phi_\lambda\rangle$$

$$\le \sum_\lambda |\langle\phi_\lambda|\exp[-i\mathbf{Q}\cdot\boldsymbol{\xi}]|\phi_\lambda\rangle|^2 \qquad (39)$$

$$\le \frac{1}{Z_A}\left|\sum_\lambda \langle\phi_\lambda|\exp[-i\mathbf{Q}\cdot\boldsymbol{\xi}]|\phi_\lambda\rangle\right|^2.$$

He performed the calculation [5] using two substitutions:

$$\sum_\lambda |\langle\phi_\lambda|\exp[-i\mathbf{Q}\cdot\boldsymbol{\xi}]|\phi_\lambda\rangle|^2$$

$$\longrightarrow \sum_\lambda \langle\phi_\lambda|\exp[-i\mathbf{Q}\cdot\boldsymbol{\xi}]|\phi_\lambda\rangle,$$

$$\sum_\lambda |\langle\phi_\lambda|\exp[-i\mathbf{Q}\cdot\boldsymbol{\xi}]|\phi_\lambda\rangle|^2$$

$$\longrightarrow \frac{1}{Z_A}\left|\sum_\lambda \langle\phi_\lambda|\exp[-i\mathbf{Q}\cdot\boldsymbol{\xi}]|\phi_\lambda\rangle\right|^2,$$

and found that the difference between the results of these calculations is very small. Consequently, we shall simply set

$$Z_A - \sum_\lambda |\langle\phi_\lambda|\exp[-i\mathbf{Q}_0^a\cdot\boldsymbol{\xi}]|\phi_\lambda\rangle|^2 \qquad (40)$$

$$\longrightarrow Z_A - \sum_\lambda \langle\phi_\lambda|\exp[-i\mathbf{Q}_0^a\cdot\boldsymbol{\xi}]|\phi_\lambda\rangle = Z_{A,\,\mathrm{eff}}(\mathbf{Q}_0^a),$$

where $Z_{A,\,\mathrm{eff}}(\mathbf{Q}_0^a)$ is determined by Eq. (28). Then we obtain finally for the antiscreening cross section

$$\sigma^a_{0\to n} = \frac{4}{v^2}\sum_{s_I}\int d^2q_\perp\, Z_{A,\,\mathrm{eff}}(\mathbf{Q}_0^a)$$

$$\times\left|\langle\psi_n(\mathbf{r})|\exp(i\mathbf{q}_0\cdot\mathbf{r})\left(1-\frac{\mathbf{v}}{c}\cdot\boldsymbol{\alpha}\right)|\psi_0(\mathbf{r})\rangle\right|^2 \qquad (41)$$

$$\times\left[q_\perp^2 + \frac{(\omega_{n0}+\Delta\epsilon)^2}{v^2\gamma^2} + 2(\gamma-1)\frac{\omega_{n0}\Delta\epsilon}{v^2\gamma^2}\right]^{-2}.$$

The completeness method and the approximation (36) and (40) are widely used to study the loss of an electron by an ion in nonrelativistic collisions with an atom. It is easy to see that the use of these approximations with $v \sim c$ does not touch the relativistic features in the antiscreening cross section.

4. RESULTS AND DISCUSSION

In Table 1 the experimental data from [11] are compared with the theoretical results obtained in [6] and [10] and in the present paper. In [11] the cross sections for the loss of an electron by Au^{78+} ions were measured as these ions passed through various solid-state targets with collision energy 10.8 GeV/au, corresponding to $\gamma = 12.6$. Our numerical calculations are based on Eqs. (33) and (41), where the final states of the electron are states in the continuous spectrum of the ion and where the summation extends over all of these states. To describe the electronic states of the ion we used approximate relativistic wave functions, which were used in [8, 17] to calculate the ionization of the *K* shell in relativistic collisions with point charges. Just as in these works, we took account of the electronic transitions with no change and with a change in the electron spin. It is known (see, e.g., [7, 8, 17]) that these approximate wave functions give good results even for describing such heavy single-electron ions as lead and gold.

To calculate the cross section (41), generally speaking, it is necessary to know the parameters $\Delta\epsilon$. As far as we know, there are no sufficiently accurate but simple prescriptions for choosing these parameters for multielectronic atoms [1, 3]. In the present work we took for this parameter the average atomic excitation energy, which is used in calculating energy losses by fast charged particles (see, e.g., [18]). We note that in reality the accuracy in determining the values of the parameter $\Delta\epsilon$ is not critical for the present calculation. For electron loss by such multiply charged ions as Au^{78+} and Pb^{81+} in collisions with light targets, such as Be, C, and Al, considered here, the terms with $\Delta\epsilon$ and the expression (41) are negligible even for ultrarelativistic collisions, studied experimentally in [9] as well as below (see Table 2). For collisions of Au^{78+} and Pb^{81+} with heavier targets (Cu and so on, see Tables 1 and 2) the antiscreening cross section is only a very small correction (about 1–3%) to the screening cross section. Since the accuracy of our calculation of the screening cross section taking account of the approximations made above is estimated to be 15–20%, there is no sense in determining the parameter $\Delta\epsilon$ more accurately for the collisions considered here.

As follows from Table 1, for a collision energy of 10.8 GeV/au there is no appreciable difference between the experiment and calculations, just as between various calculations. The agreement between the computational results obtained in [6], predicting for the loss cross section a dependence of the form $\sigma \sim \ln\gamma$, with the experimental results for $\gamma \sim 10$ is due to the fact that in this experiment very heavy ions, such as Au^{78+}, were

Table 1. Experimental and theoretical cross sections for loss (in kb) of an electron by Au^{78+} ions with energy 10.8 GeV/au passing through various targets. The ions are initially in the ground state

Atom	Z_A	Experiment [11]	Theory [6]	Theory [10]	This work
C	6	0.31	0.31	0.27	0.31
Al	13	1.18	1.28	1.15	1.2
Cu	29	5.26	5.8	5.37	5.65
Ag	47	16.2	14.4	13.7	14.7
Au	79	38.2	38.8	38.0	38.5

used. In this case the energy required to detach an electron from an ion is so high that the term $\omega_{n0}^2/v^2\gamma^2 \sim Z_I^4/v^2\gamma^2$ dominates all screening constants κ_i^2 even for $\gamma \sim 10$, and the effective atomic charge $Z_{A,\,\mathrm{eff}}$ in the cross section (27) and (41) is essentially identical to the charge of the atomic nucleus.

The situation changes substantially for ultrarelativistic collisions, studied experimentally in [9], where the cross section for the loss of an electron by a Pb^{81+} ion was measured for collision energy 160 GeV/au In this case the parameter $\gamma = 170$ is very large and for the ion–atom pairs considered ultrarelativistic features of the screening of the atomic nucleus by the atomic electrons become important. The experimental data from [9] and various theoretical calculations are compared in Table 2, where we have included results for the loss cross sections, which were measured in a so-called ionization experiment and which are more accurate [19]. Compared with Table 1, we added several new columns of computational results. The data from [10], now presented in two columns, show the computational results for the cross section for losses in collisions with bare atomic nuclei and the corresponding neutral atoms. Our computational results are also given in two columns. The first one is the cross section for the loss of an electron in collisions in bare atomic nuclei, when this cross section is proportional to Z_A^2. The second column gives the computational results for collisions with neutral atoms, using Eqs. (33) and (41).

Our results for collisions with neutral atoms are much smaller than the results in [6] and are in quite good agreement with the experimental data. Compared with the calculations in [10], our results for the cross sections are much higher, especially for collisions with bare atomic nuclei, where the difference reaches about 30%.

The comparatively small difference between our results for the cross section for electron loss in collisions with neutral Be atoms and Be^{4+} ions is a consequence of the contribution of the antiscreening cross section to the total loss cross section. For collisions with light atoms, the inelastic mode makes a large contribution to the total loss cross section, reaching about 20% of the total cross section in collisions with Be atoms. Our calculation shows that for a collision energy of 160 GeV/au the compensation of the field of the atomic nucleus by the fields of the atomic electrons in screening and antiscreening processes decrease the total cross section by approximately 35% for collisions of Pb^{81+} ions with Be atoms and by approximately 50% for collisions of Pb^{81+} ions with Au atoms.

In conclusion, we note that the general, fairly good agreement between our computational data and experiment gives another, though indirect, argument justifying the use of the nonrelativistic-atom approximation.

Table 2. Experimental and theoretical cross sections for loss (in kb) of an electron by Pb^{81+} ions with energy 160 GeV/au passing through various targets. The ions are initially in the ground state

Atom	Z_A	Experiment [9]	Theory [6]	Theory [10]		This work	
			atom	nucleus	atom	nucleus	atom
Be	4	0.14	0.24	0.15	0.14	0.2	0.17
C	6	0.31	0.49	0.33	0.28	0.45	0.35
Al	13	1.3	2.0	1.6	1.1	2.14	1.4
Cu	29	6.9	9.0	7.8	5.2	10.6	6.5
Sn	50	15	25	23	15	31.5	17.6
Au	79	42	60	58	35	78.7	40.1

5. CONCLUSIONS

In this paper the excitation (loss) of an electron by an ion in relativistic collisions with atoms was studied. The analysis was based on first-order perturbation theory for the relativistic electromagnetic interaction between two composite particles—an ion and an atom. The expressions obtained for the excitation (loss) cross sections pass in the limit $c \longrightarrow \infty$ into the well-known nonrelativistic results. For (ultra)relativistic collisions the expressions obtained describe an important feature in the screening of atomic electrons of the field of the atomic nucleus that is absent in the nonrelativistic collisions. The numerical results which are based on the expressions obtained for the cross sections are in good agreement with existing experimental data.

REFERENCES

1. J. H. McGuire, *Electron Correlation Dynamics in Atomic Collisions* (Cambridge Univ. Press, Cambridge, 1997).
2. J. Eichler and W. E. Meyerhof, *Relativistic Atomic Collisions* (Academic, New York, 1995).
3. E. C. Montenegro, W. E. Meyerhof, and J. H. McGuire, Adv. At., Mol., Opt. Phys. **34**, 249 (1994).
4. A. B. Voitkiv, G. M. Sigaud, and E. C. Montenegro, Phys. Rev. A **59**, 2794 (1999).
5. R. Anholt, Phys. Rev. A **31**, 3579 (1985).
6. R. Anholt and U. Becker, Phys. Rev. A **36**, 4628 (1987).
7. D. M. Davidovic, B. L. Moiseiwitsch, and P. H. Norrington, J. Phys. B **11**, 847 (1978).
8. R. Anholt, Phys. Rev. A **19**, 1004 (1979).
9. H. F. Krause, S. R. Vane, S. Datz, *et al.*, Phys. Rev. Lett. **80**, 1190 (1998).
10. A. H. Sørensen, Phys. Rev. A **58**, 2895 (1998).
11. N. Claytor, A. Belkacem, T. Dinneen, *et al.*, Phys. Rev. A **55**, R842 (1997).
12. J. D. Bjorken and S. D. Drell, *Relativistic Quantum Mechanics* (McGraw-Hill, New York, 1964; Nauka, Moscow, 1978).
13. V. B. Berestetskii, E. M. Lifshitz, and L. P. Pitaevskii, *Relativistic Quantum Theory* (Nauka, Moscow, 1968, 1971; Pergamon, Oxford, 1971).
14. M. R. C. McDowell and J. P. Coleman, *Introduction to the Theory of Ion-Atom Collisions* (North-Holland, Amsterdam, 1970).
15. F. Salvat, J. D. Martinez, R. Mayol, and J. Parellada, Phys. Rev. A **36**, 467 (1987).
16. E. C. Montenegro and W. E. Meyerhof, Phys. Rev. A **46**, 5506 (1992).
17. B. L. Moiseiwitsch, Phys. Rep. **118**, 133 (1985).
18. U. Fano, Annu. Rev. Nucl. Sci. **13**, 1 (1963).
19. R. H. Schuch, private communication.

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 4, 2000, pp. 673–681.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 4, 2000, pp. 777–786.

ATOMS, SPECTRA, RADIATION

Momentum-Transfer Scattering Cross Section and the Aharonov–Bohm Effect on a Toroidal Solenoid

V. V. Lyuboshitz and V. L. Lyuboshitz*

Joint Institute for Nuclear Research, Dubna, Moscow oblast, 141980 Russia

**e-mail: lyubosh@sunhe.jinr.ru*

Received March 30, 2000

Abstract—The quantum-mechanical Aharonov–Bohm effect in the diffraction of charged particles by a toroidal solenoid containing a magnetic field is investigated. The total and differential elastic scattering cross sections depend on the magnetic flux inside the solenoid, even in the presence of a "black" ring-shaped screen which prevents charged particles from entering the region where the magnetic field is localized. Relations describing the momentum-transfer cross section for the elastic scattering of charged particles by a toroidal solenoid are obtained in the eikonal approximation and in a unitary model of scattering with a sharp jump in the partial amplitudes. The momentum-transfer scattering cross section is proportional to the average transfer of the longitudinal momentum of the scattered particle and can be expressed in terms of a force operator. It is shown that in the absence of a screen the momentum-transfer scattering cross section of toroidal solenoid is indeed determined only by the part of the incident beam that intersects the inner region of the toroidal solenoid, where the magnetic field intensity and, therefore, the Lorentz force are nonzero. At the same time, the momentum-transfer cross section for the scattering of charged particles by a toroidal solenoid covered by a "black" ring-shaped screen does not depend on the magnetic flux inside the solenoid and is identical to the momentum-transfer cross section for diffraction by the same screen. The contribution from scattering by an opening in the screen, which depends on the magnetic flux, is completely compensated by the contribution of the interference of the scattering amplitudes of the opening and the "black" screen. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

About 40 years ago Aharonov and Bohm showed [1–3] that a magnetic field completely screened from the penetration of charged particles influences the interference of two coherent beams of these particles passing over the magnetic flux. The first works on the Aharonov–Bohm effect suffered from a methodological drawback because the source of the magnetic field was an infinite cylindrical solenoid. The switch in [4, 5] and later in [6] from an infinite solenoid to a finite toroidal solenoid, topologically separating space into regions with and without a magnetic field removed a number of theoretical ambiguities and made the Aharonov–Bohm effect a real quantum-mechanical phenomenon that can be checked experimentally. The elegant experiments performed by Tonomura *et al.* [7–10] investigating electron diffraction by screened toroidal solenoids convincingly confirmed the seemingly puzzling, and having no a classical analog, "nonlocal" effect of a magnetic field on charged particles through the vector potential.

According to [4, 5], if one of the coherent beams of charged particles passes through the "hole" of a solenoid curled into a torus while the other beam passes outside this hole, then when a current is passed through the closed solenoid an additional phase difference

$$\alpha = \frac{e\Phi}{\hbar c}, \tag{1}$$

where e is the electric charge of a particle, Φ is the magnetic flux confined inside the toroid (outside the toroid the magnetic field is zero [4]) and equal to the circulation of the vector potential

$$\oint \mathbf{A} d\mathbf{l}$$

along trajectories passing around the magnetic flux (see Fig. 1), arises between the interfering beams. The effective cross section for Fraunhofer diffraction of charged particles by a toroidal solenoid covered by a "black," completely absorbing, screen Σ_1 with an opening Σ_2, entirely located inside the projection of the "hole" of the solenoid onto the plane of the screen, was calculated in [4, 5] on the basis of the relation (1) in a gauge-invariant eikonal approximation. It was assumed that the transverse size of the incident wave packet is large compared with the size of the screen and the solenoid.

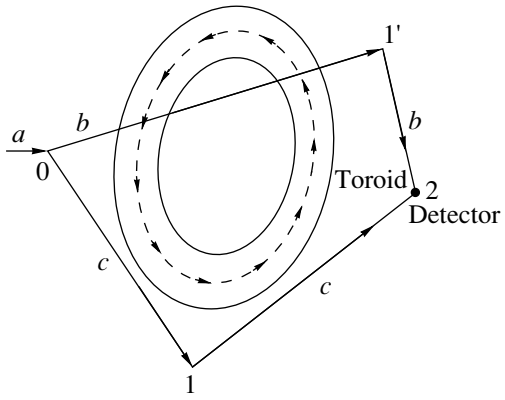

Fig. 1. Arrangement of the interference experiment. A beam *a* of particles with charge *e* is split at the point zero into two coherent beams *b* and *c*; the beam *b* passes through the "hole" of the toroidal solenoid with magnetic flux Φ (the trajectory $0 \longrightarrow 1' \longrightarrow 2$), and the beam *c* passes outside the toroid (the trajectory $0 \longrightarrow 1 \longrightarrow 2$). Both beams intersect and interfere at the point 2. The phase difference is $\delta = 2\pi(L_c - L_b)/\lambda + e\Phi/\hbar c$, where L_c is the length of the path $0 \longrightarrow 1 \longrightarrow 2$ and L_b is the length of the path $0 \longrightarrow 1' \longrightarrow 2$, λ is the de Broglie wavelength of the charged particle. The sign of the magnetic flux depends on the direction of the magnetic field **B** in the solenoid. In the configuration shown in the figure $\Phi > 0$.

It was found that under these conditions the elastic scattering amplitude possesses the structure

$$f(\mathbf{q}) = f_{\Sigma_1}(\mathbf{q}) - f_{\Sigma_2}(\mathbf{q})\exp\left(i\frac{e\Phi}{\hbar c}\right), \qquad (2)$$

where $f_{\Sigma_1}(\mathbf{q})$ and $f_{\Sigma_2}(\mathbf{q})$ are the amplitudes for the diffraction by continuous, completely absorbing screens with the shapes Σ_1 and Σ_2, respectively, and areas S_1 and S_2, and $\hbar\mathbf{q}$ is the change in the momentum on scattering. According to the optical theorem, the total elastic scattering cross section of a toroidal solenoid with a flat ring-shaped screen oriented perpendicular to the initial momentum $\hbar\mathbf{k}$ is [4, 5]

$$\sigma_{el} = \frac{4\pi}{k}\mathrm{Im}\, f(0) - \Delta S = 4S_2\sin^2\left(\frac{e\Phi}{2\hbar c}\right) + \Delta S. \qquad (3)$$

Here $k = |\mathbf{k}|$, $\Delta S = S_1 - S_2$ is the total cross section for diffraction by a "black" ring-shaped screen, identical to the absorption cross section and equal to the area of the screen, and S_2 is the area of the opening in the screen.

For a very thin solenoid with a finite magnetic flux Φ and a very thin "black" screen, we have

$$\sigma_{el} = 4S\sin^2\left(\frac{e\Phi}{2\hbar c}\right), \qquad (4)$$

where $S = S_1 \approx S_2$. We note that the same formula also holds in the absence of a screen covering a thin solenoid [4, 5]. Similar results were obtained subsequently in [11, 12].

It is obvious that in the presence of a "black" screen the wave function of the charged particles inside the region where the magnetic field intensity is different from zero tends to zero. Classically, the magnetic-field dependence of the effective scattering cross section for a charged particle is surprising under these conditions. Thus, the Aharonov–Bohm effect is a purely quantum phenomenon associated with the wave properties of the microparticles.

At the same time, in the quantum theory of scattering there exists a physical quantity which is proportional to the force acting on the scattering particle and is expressed in terms of the expectation value of the force operator with respect to the scattering state. This is the so-called momentum-transfer scattering cross section [13–15]. The dependence of the momentum-transfer cross section on the magnetic flux with the magnetic field being inaccessible to the charged particles would be completely improbable and would contradict the law of conservation of momentum. In the present work it is shown that the momentum-transfer elastic scattering cross section of a toroidal solenoid covered by a ring-shaped, completely absorbing, screen does not depend on the magnetic flux inside the solenoid. In our view, this fact finally removes the "paradox" that is usually associated with the Aharonov–Bohm effect.

2. MOMENTUM TRANSFER AND THE MOMENTUM-TRANSFER ELASTIC-SCATTERING CROSS SECTION

If the scattering center is irradiated by a flux of particles *I*, each particle possessing momentum **p**, then a force **F** equal to the momentum transmitted to the scatterer per unit time in absorption and elastic scattering processes acts on the scatterer. This force is described by the formula [16]

$$\mathbf{F} = I[\mathbf{p}\sigma_{\mathrm{abs}} + \int\sigma_{\mathrm{el}}(\theta,\phi)(\mathbf{p} - \mathbf{p}')d\Omega], \qquad (5)$$

where σ_{abs} is the absorption cross section, σ_{el} is the elastic scattering cross section, $\mathbf{p}'$ is the momentum of the scattered particle, and θ and ϕ are the polar and azimuthal angles.

The longitudinal force acting on a scatterer has the form

$$F^{(\parallel)} = \mathbf{F}\frac{\mathbf{p}}{p} = Ip(\sigma_{\mathrm{abs}} + \sigma_{\mathrm{tr}}), \qquad (6)$$

where $p = |\mathbf{p}|$, σ_{tr} is the momentum-transfer elastic scattering cross section, which is defined as [16–18]

$$\sigma_{\mathrm{tr}} = \int\sigma_{\mathrm{el}}(\theta,\phi)(1 - \cos\theta)d\Omega. \qquad (7)$$

For azimuthal symmetry, only a longitudinal momentum is transferred to the scatterer, so that

$$\mathbf{F} = \mathbf{n}F^{(\parallel)},$$

where $\mathbf{n} = \mathbf{p}/p$ is a unit vector in the direction of the initial momentum. Then

$$\sigma_{tr} = 2\pi\int_{-1}^{1}|f(\theta)|^2(1-\cos\theta)d(\cos\theta), \qquad (8)$$

where $f(\theta)$ is the elastic scattering amplitude.

Let N particles with momentum $\mathbf{p}$ be located in the volume V. Then the intensity I can be represented in the form

$$I = \frac{N}{V}\frac{p}{m}, \qquad (9)$$

where m is the mass of a particle. On the basis of the formula (6), according to the third law of mechanics the force

$$\mathbf{f} = -\frac{\mathbf{F}}{N} = -\frac{p^2}{mV}\mathbf{n}(\sigma_{tr}+\sigma_{abs}) \qquad (10)$$

should act on a scattering particle located in the volume V. In what follows we shall be interested in the contribution of elastic scattering

$$\mathbf{f}_{el} = -\frac{p^2}{mV}\mathbf{n}\sigma_{tr}. \qquad (11)$$

According to Ehrenfest's theorem [19], on the basis of quantum mechanics the vector $\mathbf{f}_{el}$ should be regarded as the expectation value of the operator corresponding to the external force acting on the scattering particle. In the case at hand the averaging must be performed over a stationary state in the scattering problem, normalized to the macroscopic volume V. Thus

$$\mathbf{f}_{el} = \frac{1}{V}\int\psi_{\mathbf{p}}^{(+)*}\hat{\mathbf{f}}\psi_{\mathbf{p}}^{(+)}d^3\mathbf{r}, \qquad (12)$$

where $\psi_{\mathbf{p}}^{(+)}$ is a wave function which belongs to the continuous spectrum and is asymptotically a superposition of a plane wave and a diverging spherical wave [18], and $\hat{\mathbf{f}}$ is the force operator. Comparing the expressions (11) and (12), we find [14, 15]

$$\sigma_{tr} = -\frac{1}{2\mathscr{E}}\int\psi_{\mathbf{p}}^{(+)*}(\hat{\mathbf{f}}\cdot\mathbf{n})\psi_{\mathbf{p}}^{(+)}d^3\mathbf{r}. \qquad (13)$$

Here $\mathscr{E}$ is the kinetic energy of the particle. The relation (13) should be viewed as a generalization of Ehrenfest's theorem to the scattering state. The result (13) was first obtained (in a somewhat different form) in [13] on the basis of the generalized optical theorem for the change in quantum-mechanical quantities in a scattering process (see also [20]).

The operator for the force $\hat{\mathbf{f}}$, acting on a particle with charge e in an arbitrary external electromagnetic field, has the form [19]

$$\hat{\mathbf{f}} = e\mathbf{E} + \frac{e}{2c}\hat{\mathbf{v}}\times\mathbf{H} - \frac{e}{2c}\mathbf{H}\times\hat{\mathbf{v}}, \qquad (14)$$

where $\mathbf{E}$ is the intensity of the electric field, $\mathbf{H}$ is the intensity of the magnetic field,

$$\hat{\mathbf{v}} = -\frac{i\hbar\nabla}{m} - \frac{e}{mc}\mathbf{A} \qquad (15)$$

is the velocity operator of the charged particle, and $\mathbf{A}$ is the vector potential. In the absence of an electric field (toroidal solenoid) the operator $\hat{\mathbf{f}}$ corresponds simply to the Lorentz force in a magnetic field. The formulas (12) and (13) show that the momentum-transfer cross section for elastic scattering of electrons by an unscreened toroidal solenoid is different from zero only because electrons penetrate into the toroid, in the region where the magnetic field is localized. For a toroid with a "black" ring-shaped screen, electrons do not reach the region where the magnetic field is concentrated, and the momentum-transfer cross section should be determined only by the form of the screen and should not depend on the magnitude of the magnetic flux. It is shown below that these results are confirmed by direct calculations of the momentum-transfer cross section in the eikonal approximation, as well as in a unitary model of scattering with a sharp jump in the partial amplitudes at the boundaries of the scatterer.

3. MOMENTUM-TRANSFER CROSS SECTION FOR FRAUNHOFER DIFFRACTION

If the de Broglie wavelength of a particle is small compared with the geometric size of the target, then by analogy to the theory of Fraunhofer diffraction in wave optics [21] the eikonal approach can be used to describe scattering by small angles. The general eikonal formula for the elastic scattering amplitude has the structure

$$f(\mathbf{q}) = \frac{ik}{2\pi}\int d^2\boldsymbol{\rho}\,\eta(\boldsymbol{\rho})\exp(-i\mathbf{q}\cdot\boldsymbol{\rho}), \qquad (16)$$

where

$$\eta(\boldsymbol{\rho}) = 1 - S(\boldsymbol{\rho}), \quad |S(\boldsymbol{\rho})| \le 1. \qquad (17)$$

Here $\boldsymbol{\rho}$ is a two-dimensional vector perpendicular to the initial momentum $\hbar\mathbf{k}$ and $\mathbf{q} = \mathbf{k}' - \mathbf{k}$ is the change in the wave vector at scattering (for small scattering angles $\mathbf{q}\perp\mathbf{k}$, $|\mathbf{q}| = k\theta$, where θ is the scattering angle). In the presence of a magnetic field the complex parameter $S(\boldsymbol{\rho})$ acquires the additional phase factor [4, 5]

$$S(\boldsymbol{\rho}) = S^{(0)}(\boldsymbol{\rho})\exp\left(i\frac{e}{\hbar c}\int_{-\infty}^{\infty}A_z(\boldsymbol{\rho},z)dz\right), \qquad (18)$$

where e is the electric charge of the particle and A_z is the projection of the vector potential onto the direction of the initial momentum, and the integration extends over the rectilinear trajectory of a particle.

We shall assume that small angles $\theta \ll 1$ make the main contribution to the momentum-transfer elastic-scattering cross section, just as in the case of the total elastic scattering cross section.[1] For small scattering angles the element of solid angle

$$d\Omega \approx d^2\mathbf{q}/k^2,$$

and

$$1 - \cos\theta \approx \mathbf{q}^2/2k^2.$$

On the basis of the definition (7), using the relation (16) for the scattering amplitude we obtain as a result the eikonal formula for the momentum-transfer elastic-scattering cross section

$$\sigma_{tr} = \frac{1}{2k^4}\int |f(\mathbf{q})|^2\mathbf{q}^2 d^2\mathbf{q} = \frac{1}{2k^2}\int \left|\frac{d\eta(\boldsymbol{\rho})}{d\boldsymbol{\rho}}\right|^2 d^2\boldsymbol{\rho}. \quad (19)$$

At the same time the eikonal formulas for the integral elastic-scattering cross section and the total interaction cross section have the form [18]

$$\sigma_{el} = \int d^2\boldsymbol{\rho}|\eta(\boldsymbol{\rho})|^2,$$
$$\sigma_{tot} = 2\int d^2\boldsymbol{\rho}\,\mathrm{Re}\,\eta(\boldsymbol{\rho}). \quad (20)$$

For Fraunhofer diffraction by a "black" screen Σ, whose plane is perpendicular to the initial momentum, the partial amplitude $\eta(\boldsymbol{\rho})$ is 1 at points inside the screen and 0 outside the screen. Here

$$\sigma_{el} = \sigma_{tot}/2 = S,$$

where S is the area of the screen. We underscore that if the width b of the boundary layer is small compared to the linear dimensions of the screen, the total elastic-scattering cross section is virtually independent of the structure of the boundary. The situation is entirely different for the momentum-transfer cross section. The momentum-transfer diffraction cross section is determined completely by the character of the screen boundary near which the partial amplitude $\eta(\boldsymbol{\rho})$ changes and to which momentum is actually transferred. If the "black" screen possesses a diffuse boundary with effective width

$$b \gg 1/k = \lambda/2\pi,$$

inside which $\eta(\boldsymbol{\rho}) = 1$ and outside which $\eta(\boldsymbol{\rho}) = 0$, then, as is obvious from the relations (19),

$$\sigma_{tr} \sim \frac{L}{2bk^2}, \quad (21)$$

where L is the perimeter of the screen.

[1] This is incorrect for diffraction by a screen with a sharp boundary (see [16]).

4. MOMENTUM-TRANSFER CROSS SECTION FOR ELASTIC SCATTERING OF ELECTRONS BY A TOROIDAL SOLENOID

4.1. Fraunhofer Diffraction by a Toroidal Solenoid without a Screen

We shall calculate the momentum-transfer cross section for elastic scattering of charged particles (electrons) by an unscreened circular toroidal solenoid whose magnetic field is perpendicular to the initial momentum. Let R be the radius of the "hole" of the toroid and b the radius of a loop of the solenoid (the radius of the "bagel"). We shall assume that the de Broglie wavelength of an electron is small compared with the linear dimensions of the solenoid, i.e.,

$$R \gg 1/k, \quad b \gg 1/k;$$

the transverse dimensions of the incident wave packet are

$$\Delta \gg R, \quad \Delta \gg b.$$

Under these conditions scattering by small angles plays the main role and the eikonal approach can be used [4, 5].

Figure 2 displays the projection of a circular toroidal solenoid on a plane perpendicular to the magnetic field and passing through the initial momentum of the electron (the z axis). The magnetic flux inside the solenoid is, evidently,

$$\Phi = \pi b^2 H, \quad (22)$$

where H is the magnetic field intensity. According to the relations (17) and (18), the partial amplitude $\eta(\rho)$ in the case at hand is (compare [4, 5])

$$\eta(\rho) = 1 - \exp\left(\frac{ie}{\hbar c}\int_{-\infty}^{\infty} A_z(\rho, z)dz\right), \quad (23)$$

where $\rho = |\boldsymbol{\rho}|$ is the impact parameter.

As is well known, irrespective of the gauge of the vector potential $\mathbf{A}$, the circulation of the vector potential $\oint \mathbf{A}\,d\mathbf{l}$ along an arbitrary contour is equal to the magnetic flux threading this contour. Since the cross section for scattering by any finite angle is gauge-invariant, with no loss of generality Eq. (23) can be rewritten as

$$\eta(\rho) = 1 - \exp\left(i\frac{e}{\hbar c}\Phi(\rho)\right), \quad (24)$$

where Φ is the magnetic flux through a plane bounded by a straight line with impact parameter ρ and a semicircle of infinite radius. As a result, we have

$$\eta(\rho) = \eta_1 = 0, \quad \rho > R + 2b,$$
$$\eta(\rho) = 1 - \exp\left(i\frac{e}{\hbar c}\Phi(x)\right), \quad R \le \rho < R + 2b, \quad (25)$$
$$\eta(\rho) = \eta_2 = 1 - \exp\left(i\frac{e}{\hbar c}\Phi\right), \quad \rho < R.$$

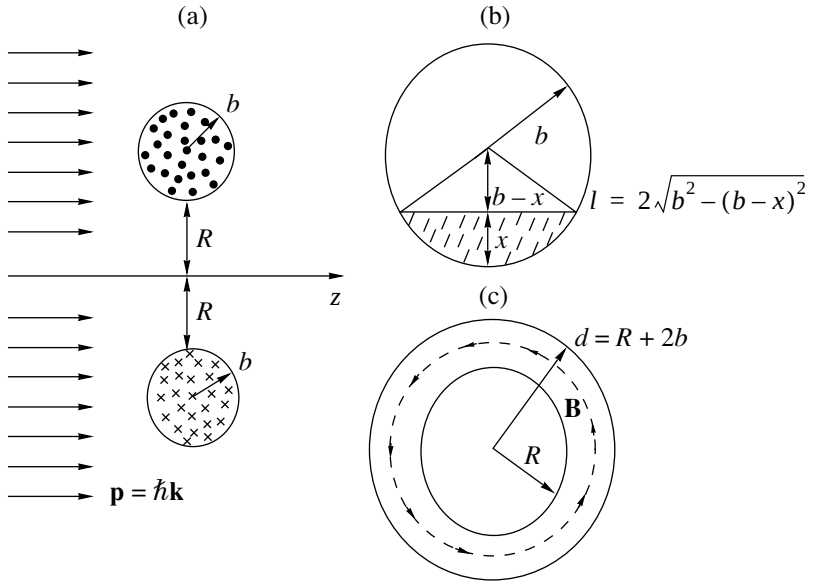

Fig. 2. Scattering of electrons by a toroidal solenoid. (a, b) section of a circular toroidal solenoid in a plane passing through the symmetry axis *z*; *R* is the radius of the "hole" of the toroid, *b* is the radius of a loop, *x* is the height of a segment, cut off by the straight lines intersecting the magnetic-field region and parallel to the *z* axis. A wide beam of electrons with momentum $\mathbf{p} = \hbar\mathbf{k}$ parallel to the symmetry axis of the toroid is incident on the toroidal solenoid. The magnetic field is perpendicular to the plane of the figure: in the upper loop it is oriented out of this plane and in the lower loop it is oriented into the plane. (c) section of a circular toroidal solenoid in a plane perpendicular to the symmetry axis and passing through the center of the toroid.

Here $e = -|e|$ is the electron charge,

$$\Phi(x) = \Phi\left(1 - \frac{S(x)}{\pi b^2}\right), \quad 0 \le x = (\rho - R) \le 2b, \quad (26)$$

$S(x)$ is the area of the hatched circular segment of height x in Fig. 2 and Φ is the total magnetic flux inside the toroidal solenoid, which is determined according to the relation (22); in the configuration shown in Fig. 2 $\Phi > 0$.

For a problem with axial symmetry the relation (19) for the momentum-transfer cross section becomes

$$\sigma_{tr} = \frac{\pi}{k^2}\int_0^\infty \left|\frac{d\eta(\rho)}{d\rho}\right|^2 \rho d\rho. \quad (27)$$

Substituting into the formula (27) the values of the partial amplitudes (24) and (25) we find

$$\sigma_{tr} = \frac{\pi}{k^2}\left(\frac{e}{\hbar c}\right)^2 \int_0^{2b} \left(\frac{d\Phi(x)}{dx}\right)^2 (R + x)dx. \quad (28)$$

It is easy to see that

$$\frac{d\Phi}{dx} = -\frac{\Phi}{\pi b^2}\frac{dS(x)}{dx} = -\frac{\Phi}{\pi b^2} l(x),$$

where

$$l(x) = 2\sqrt{b^2 - (b - x)^2} = 2\sqrt{2bx - x^2} \quad (29)$$

is the length of the chord of the circular segment of height x. Using Eq. (29) an elementary integration gives

$$\sigma_{tr} = \frac{16}{3\pi k^2}\left(\frac{e\Phi}{\hbar c}\right)^2 \left(\frac{R}{b} + 1\right). \quad (30)$$

For a very thin solenoid ($b \ll R$) the formula for the momentum-transfer cross section becomes

$$\sigma_{tr} = \frac{16R}{3\pi k^2 b}\left(\frac{e\Phi}{\hbar c}\right)^2. \quad (31)$$

It should be underscored that the expression (31) is valid only if

$$R \gg b \gg 1/k = \lambda/2\pi.$$

The situation where $b < 1/k$ will be considered below (Section 4.3).

We can see that the momentum-transfer cross section for elastic scattering of electrons by an unscreened toroidal solenoid is determined by the range of impact parameters ($R < \rho < R + 2b$) where the electron trajectories intersect the magnetic field inside the solenoid. We note that, using Eqs. (22) and (26), the expression (28) for the momentum-transfer scattering cross section can be rewritten as

$$\sigma_{tr} = \pi\int_0^{2b} (R + x)\left(\frac{el(x)H}{pc}\right)^2 dx, \quad (32)$$

where $p = \hbar k$ is the modulus of the electron momentum and $l(x)$ is the length of the electron path in a magnetic field and is given by Eq. (29). The quantity

$$\theta_0 = \frac{el(x)H}{pc} \tag{33}$$

is the classical angle of deflection of a charged particle on a rectilinear path $l(x)$ as a result of the action of the Lorentz force

$$\mathbf{F} = \frac{e}{c}\mathbf{v} \times \mathbf{H},$$

perpendicular to the momentum $\mathbf{p}$ and the magnetic field $\mathbf{H}$ ($\theta_0 = Fl(x)/pv$). Thus, the relations (28), (30), and (31) are of a classical character.

It has been shown in [17] that in the eikonal approximation the momentum-transfer cross section for scattering by a centrosymmetric potential is described by the classical formula

$$\sigma_{tr} = \pi\int_0^\infty \theta_0^2(\rho)\rho d\rho \approx 2\pi\int_0^\infty (1 - \cos\theta_0(\rho))\rho d\rho$$

(although the classical theory may not be applicable for calculating the differential and total cross sections for elastic scattering!). We have shown that this result is also valid in the presence of a magnetic field.

We underscore that in contrast to the momentum-transfer cross section the total cross section for the elastic scattering of an electron by a very thin toroidal solenoid ($b \ll R$) is determined primarily by the trajectories that do not intersect the region of localization of the magnetic field (impact parameters $\rho < R$; only the vector potential is different from zero along the path of the electrons!). Using the relations (25), Eqs. (20) lead to the result obtained in [4, 5]:

$$\sigma_{\text{el}} = \sigma_{\text{tot}} = |\eta_2|^2\int d^2\boldsymbol{\rho} = 4\pi R^2 \sin^2\left(\frac{e\Phi}{2\hbar c}\right)$$

(see also Eq. (4) with $S_2 = \pi R^2$).

4.2. Fraunhofer Diffraction by a Toroidal Solenoid with a "Black" Screen

As shown in [4, 5], for a toroidal solenoid having an arbitrary form and covered with a "black" screen with an opening which is projected onto the "hole" of the toroid, the diffraction pattern depends on the magnetic flux Φ, even though charged particles do not penetrate into the solenoid, i.e., into the region where the magnetic field is concentrated; the dependence of the elastic scattering cross section on the magnetic field is determined by the dimensions and shape of the opening. The total cross section for elastic scattering by a screened toroidal solenoid is given by Eq. (3) (see the relation (14) in [5]).

We shall now find the momentum-transfer cross section for elastic scattering of electrons by such a solenoid, using the relation (19). According to the formulas (17) and (18), the partial amplitudes $\eta(\boldsymbol{\rho})$ in the case at hand have the form

$$\eta(\boldsymbol{\rho}) = 1 - S^{(0)}(\boldsymbol{\rho})\exp\left(i\frac{e}{\hbar c}\Phi(\boldsymbol{\rho})\right), \tag{34}$$

where, just as in Eq. (24), $\Phi(\boldsymbol{\rho})$ is the magnetic flux encompassed by the closed contour consisting of the rectilinear trajectory with impact parameter $\boldsymbol{\rho}$ and a semicircle with radius tending to infinity; the impact parameter here is a two-dimensional vector perpendicular to the initial electron momentum $\mathbf{p} = \hbar\mathbf{k}$. It is easy to see that if the impact parameters correspond to rectilinear trajectories passing on the outside of the screen, then

$$S^{(0)}(\boldsymbol{\rho}) = 1, \quad \Phi(\boldsymbol{\rho}) = 0, \quad \eta(\boldsymbol{\rho}) = \eta_1 = 0.$$

If the impact parameters correspond to rectilinear trajectories passing through an opening in the screen, then

$$S^{(0)}(\boldsymbol{\rho}) = 1, \quad \Phi(\boldsymbol{\rho}) = \Phi, \quad \eta(\boldsymbol{\rho}) = \eta_2,$$

where η_2 is given by the third formula in Eqs. (25). Since η_1 and η_2 are constants, according to Eq. (19) the momentum-transfer cross section for elastic scattering of electrons (and any charged particles) by a screened solenoid is determined only by the trajectories which intersect the ring-shaped screen. Since according to the conditions of the problem the electrons do not penetrate into the toroidal solenoid located behind the screen, the trajectories which in the absence of the screen would pass through the region where the magnetic field is concentrated correspond to the values

$$S^{(0)}(\boldsymbol{\rho}) = 0, \quad \eta(\boldsymbol{\rho}) = 1.$$

For trajectories intersecting the screen near its inner boundary,

$$\eta(\boldsymbol{\rho}) = 1 - S^{(0)}(\boldsymbol{\rho})\exp\left(i\frac{e\Phi}{\hbar c}\right), \tag{35}$$

and for trajectories intersecting the screen near its outer boundary

$$\eta(\boldsymbol{\rho}) = 1 - S^{(0)}(\boldsymbol{\rho}), \tag{36}$$

and inside the screen $S^{(0)}(\boldsymbol{\rho}) = 0$ while near its boundaries $1 > |S^{(0)}(\boldsymbol{\rho})| > 0$.

We can see that the quantity

$$\left|\frac{d\eta(\boldsymbol{\rho})}{d\boldsymbol{\rho}}\right|^2 = \left|\frac{dS^{(0)}(\boldsymbol{\rho})}{d\boldsymbol{\rho}}\right|^2, \tag{37}$$

appearing in Eq. (19) for the momentum-transfer cross section, does not depend on the magnetic flux inside the solenoid. Thus, the momentum-transfer cross section for elastic scattering of electrons by a toroidal solenoid covered by an impenetrable ring-shaped screen, in contrast to the total scattering cross section, does not depend on the magnetic flux and is identical to the

momentum-transfer diffraction cross section of the same screen. The magnitude of the momentum-transfer cross section, as already mentioned in Section 3, depends on the structure of the transitional layer and is proportional to the total length of the outer and inner boundaries of the ring (see the estimate (21)).

We note that taking account of the relation (2) for the scattering amplitude the momentum-transfer cross section for elastic scattering of charged particles by a toroidal solenoid, covered by a screen Σ_1 with an opening Σ_2, is simply the sum of the momentum-transfer cross sections for elastic scattering by continuous screens Σ_1 and Σ_2. Indeed, in calculating the momentum-transfer cross section on the basis of the eikonal formula (19) with amplitude (2) the contribution of the interference of the amplitudes $f_{\Sigma_1}(\mathbf{q})$ and $f_{\Sigma_2}(\mathbf{q})$, which possess the diffraction structure (16), vanishes, since according to the condition of the problem the derivatives $d\eta_1/d\boldsymbol{\rho}$ and $d\eta_2/d\boldsymbol{\rho}$ are different from zero only near the boundaries of the corresponding regions Σ_1 and Σ_2, i.e., in the nonintersecting ranges of the impact parameters. Ultimately, the dependence on the magnetic flux in the solenoid also vanishes.

4.3. Unitary Model of Diffraction by a Toroidal Solenoid

We assumed above that the small-angle range $\theta \ll 1$ makes the main contribution not only to the total cross section for the elastic scattering but also to the momentum-transfer cross section. This is correct if the de Broglie wavelength $\lambda = 2\pi/k$ is small compared to the target size L and the width b of the transitional layer where the partial amplitudes vary. For a very sharp target boundary, so that $L \gg \lambda > b$, the results obtained for the total elastic scattering cross section by the eikonal approach remain valid but when calculating the momentum-transfer cross section the range of angles $\theta \sim 1$ where the eikonal approximation is inapplicable now plays a large role.

To take into account the contribution of large scattering angles we shall examine, following [16], the model of diffraction of electrons by an infinitely thin toroidal solenoid with magnetic flux Φ; this model completely reproduces the results of the eikonal approach for the differential scattering cross section for small angles and for the total scattering cross section, and at the same time it strictly satisfies the condition of unitarity of the S matrix. In this model the amplitude for elastic scattering of electrons by an unscreened circular toroid with "hole" radius $R \gg 1/k$ and magnetic field in a plane perpendicular to the initial momentum has the form

$$f(\theta) = \frac{i}{2k}\left[1 - \exp\left(i\frac{e\Phi}{\hbar c}\right)\right]\sum_{l=0}^{n}(2l+1)P_l(\cos\theta), \quad (38)$$

where $P_l(\cos\theta)$ are Legendre polynomials and n is the integral part of the quantity $kR \gg 1$. It is easy to see that the amplitude (38) satisfies the optical theorem:

$$\sigma_{el} = \frac{4\pi}{k}\mathrm{Im}\, f(0) = 2\pi\int_{-1}^{1}|f(\theta)|^2 d(\cos\theta) = 2\frac{\pi}{k^2}(n+1)^2\left[1 - \cos\left(\frac{e\Phi}{\hbar c}\right)\right], \quad (39)$$

i.e.,

$$\sigma_{el} \approx 4\pi R^2 \sin^2(e\Phi/2\hbar c)$$

(see the relation (4)).

In calculating the momentum-transfer cross section it is convenient to use the well-known relation for the Legendre polynomials [22] and represent the scattering amplitude (38) in the form

$$f(\theta) = \frac{i}{2k}(n+1)\frac{P_n(\cos\theta) - P_{n+1}(\cos\theta)}{1-\cos\theta} \times\left[1 - \exp\left(i\frac{e\Phi}{\hbar c}\right)\right]. \quad (40)$$

Since the Legendre polynomials are orthonormal, we find [16]

$$\sigma_{tr} = 2\pi\int_{-1}^{1}|f(\theta)|^2(1-\cos\theta)d(\cos\theta) = \frac{4\pi}{k^2}(n+1)\sin^2\left(\frac{e\Phi}{2\hbar c}\right) \approx \frac{4\pi R}{k}\sin^2\left(\frac{e\Phi}{2\hbar c}\right). \quad (41)$$

The expression (41) differs substantially from the result (31) for a solenoid with finite width $b \ll R$, $b \gg 1/k$.

Now let the toroidal solenoid be covered by a "black" screen which completely absorbs electrons and has the form of a circular ring with outer radius a and inner radius d ($a > R$, $d < R$). Once again, we assume that the plane of the screen is perpendicular to the initial momentum of the electrons $\mathbf{p} = \hbar\mathbf{k}$. Then

$$ka \gg 1, \quad kd \gg 1.$$

In the unitary model under discussion, the amplitude for elastic scattering of electrons by a screened toroidal solenoid can be written in the form (compare with Eq. (2))

$$f(\theta) = \frac{i}{2k}\sum_{l=0}^{n}(2l+1)P_l(\cos\theta) - \frac{i}{2k}\sum_{l=0}^{m}(2l+1)P_l(\cos\theta)\exp\left(i\frac{e\Phi}{\hbar c}\right), \quad (42)$$

or

$$f(\theta) = f_{\rm scr}(\theta) + \frac{i}{2k}\sum_{l=0}^{m} P_l(\cos\theta)\left[1 - \exp\left(i\frac{e\Phi}{\hbar c}\right)\right], \quad (43)$$

where

$$f_{\rm scr}(\theta) = \frac{i}{2k}\sum_{l=m+1}^{n} (2l+1)P_l(\cos\theta)$$

is the amplitude for diffraction by the "black" ring [16]. Here n and m are the integer parts of the parameters ka and kd, respectively, where $n \gg 1$ and $m \gg 1$.

Using Eqs. (42) and (43), we obtain for the angle-integrated cross section for elastic scattering

$$\sigma_{\rm el} = \frac{\pi}{k^2}(n+1)^2 + \frac{\pi}{k^2}(m+1)^2 - \frac{2\pi}{k^2}(m+1)^2\cos\left(\frac{e\Phi}{\hbar c}\right) = 4\pi d^2\sin^2\left(\frac{e\Phi}{2\hbar c}\right) + \pi(a^2 - d^2). \quad (44)$$

We underscore that the expression (44) completely agrees with Eq. (3) obtained in [4, 5] in the eikonal approach (in this case $S_2 = \pi d^2$, $\Delta S = \pi(a^2 - d^2)$).[2]

We shall now determine the momentum-transfer cross section for elastic scattering of electrons by a screened toroidal solenoid. The elastic scattering amplitude (42) can be represented, similarly to Eq. (40), in the form

$$f(\theta) = \frac{i}{2k}\frac{P_n(\cos\theta) - P_{n+1}(\cos\theta)}{1 - \cos\theta} - \frac{i}{2k}\exp\left(i\frac{e\Phi}{\hbar c}\right)\frac{P_m(\cos\theta) - P_{m+1}(\cos\theta)}{1 - \cos\theta}, \quad (45)$$

where $n > m$, $(n - m) \gg 1$. Combining in the formula for the momentum-transfer cross section

$$\sigma_{\rm tr} = 2\pi\int_{-1}^{1} |f(\theta)|^2(1 - \cos\theta)d(\cos\theta)$$

the equivalent expressions (42) and (45) for the amplitude $f(\theta)$, we find that as a result of the orthogonality of the Legendre polynomials of different orders, the interference term containing the magnetic flux vanishes after integration over the scattering angles. Then

$$\sigma_{\rm tr} = \frac{\pi(n+1)}{k^2} + \frac{\pi(m+1)}{k^2} = \frac{\pi}{k}(a + d). \quad (46)$$

This expression is, evidently, identical to the momentum-transfer diffraction cross section of the "black" ring; indeed, the elastic-scattering amplitude for such a ring is described by Eqs. (42), (43), or (45) with $\Phi = 0$.

Thus, in the unitary model of diffraction, considered here, with sharp jumps in the partial amplitudes, just as in the eikonal approach, the momentum-transfer cross section for the elastic scattering of charged particles by a toroidal solenoid, protected by a completely absorbing screen which prevents the electrons from penetrating into the solenoid, does not depend on the magnetic field inside the solenoid. The contribution of diffraction by the opening in the screen to the momentum-transfer cross section is compensated by the contribution of the interference of the amplitudes for elastic scattering by the opening and the screen itself to the momentum-transfer cross section [see Eq. (43)].

5. CONCLUDING REMARKS

The results presented above are, in our view, of fundamental importance for understanding the Aharonov–Bohm effect and are of a general character. According to Ehrenfest's theorem for scattering states [see the relation (13)] the dependence of the momentum-transfer scattering cross section on the magnetic field inside the solenoid can be due only to the direct force exerted by the magnetic field on the charged particles, and the specific calculations confirm this conclusion. For scattering of charged particles by a screened toroidal solenoid, the incident beam is divided by a ring-shaped screen into two coherent parts, which go around the magnetic flux in the solenoid, the interference between which results in the dependence of the differential and total elastic-scattering cross sections on the magnetic field inside the solenoid. This is a purely quantum effect associated with the wave properties of microparticles. But in the momentum-transfer cross section for elastic scattering by a screened solenoid, proportional, just as in the classical theory, to the average transfer of the longitudinal momentum to the scattered particle, the magnetic-flux dependence completely vanishes. At the same time, in the absence of the screen the momentum-transfer cross section for scattering of charged particles by a solenoid is due only to the part of the beam that passes through the region containing the magnetic field.

[2] The amplitude (42) satisfies the unitarity condition

$$\frac{4\pi}{k}{\rm Im}\, f(0) = 2\pi a^2 - 2\pi d^2\cos\left(\frac{e\Phi}{\hbar c}\right) = \sigma_{\rm el} + \Delta S,$$

where $\sigma_{\rm el}$ is determined according to (44), and $\Delta S = \pi(a^2 - d^2)$ is the absorption cross section of the "black" ring.

REFERENCES

1. Y. Aharonov and D. Bohm, Phys. Rev. **115**, 485 (1959).
2. Y. Aharonov and D. Bohm, Phys. Rev. **123**, 1511 (1961).
3. Y. Aharonov and D. Bohm, Phys. Rev. **130**, 1625 (1963).
4. V. L. Lyuboshitz and Ya. A. Smorodinskiĭ, Preprint No. P2-11189, OIYaI (Joint Institute for Nuclear Research, Dubna, 1978).
5. V. L. Lyuboshitz and Ya. A. Smorodinskiĭ, Zh. Éksp. Teor. Fiz. **75**, 40 (1978) [Sov. Phys. JETP **48**, 19 (1978)].
6. H. J. Lipkin, Phys. Rev. D **23**, 1466 (1981).
7. A. Tonomura, T. Matsuda, R. Suzuki, *et al.*, Phys. Rev. Lett. **48**, 1443 (1982).

8. A. Tonomura, N. Osahabe, T. Matsuda, *et al.*, Phys. Rev. Lett. **56**, 792 (1986).
9. N. Osahabe, T. Matsuda, T. Kawasaki, *et al.*, Phys. Rev. A **34**, 815 (1986).
10. M. Peshkin and A. Tonomura, *The Aharonov–Bohm Effect* (Springer-Verlag, Berlin, 1989).
11. G. N. Afanasjev, J. Phys. A **21**, 2095 (1988).
12. G. N. Afanasjev and V. M. Shilov, J. Phys. A **22**, 5195 (1989).
13. E. Gerjuoy, J. Math. Phys. **6**, 993 (1965).
14. R. A. Brown, Phys. Lett. A **115**, 78 (1986).
15. R. A. Brown, J. Phys. A **20**, 3309 (1987).
16. V. L. Lyuboshitz, Yad. Fiz. **55**, 1996 (1992) [Sov. J. Nucl. Phys. **55**, 1108 (1992)].
17. M. I. D'yakonov and A. V. Khaetskiĭ, Zh. Éksp. Teor. Fiz. **99**, 1066 (1991) [Sov. Phys. JETP **72**, 590 (1991)].
18. L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-Relativistic Theory* (Nauka, Moscow, 1989, 4th ed.; Pergamon, Oxford, 1977, 3rd ed.), Parag. 126, 131, 152.
19. D. I. Blokhintsev, *Quantum Mechanics* (Vysshaya Shkola, Moscow, 1961; Reidel, Dordrecht, 1964), Parag. 32, 56.
20. A. I. Baz', Ya. B. Zel'dovich, and A. M. Perelomov, *Scattering, Reactions and Decays in Nonrelativistic Quantum Mechanics* (Nauka, Moscow, 1971, 2nd ed.; Israel Program for Scientific Translations, Jerusalem, 1966), Chap. II, Parag. 4.
21. L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields* (Nauka, Moscow, 1988; Pergamon, Oxford, 1975), Parag. 61.
22. I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products* (GIFML, Moscow, 1963; Academic, New York, 1980).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 4, 2000, pp. 682–690.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 4, 2000, pp. 787–797.

ATOMS, SPECTRA, RADIATION

Scattering of Charged Particles by Atoms

V. I. Radchenko

Ural State Technical University, Yekaterinburg, 620002 Russia
e-mail: rad@nich.ustu.ru

Received March 30, 2000

Abstract—A closed variant of the Born approximation for calculating differential scattering cross sections in ion-atom collisions is developed. An expression in terms of the matrix elements J_{ij} with respect to the single-electron states of the atom is found for the matrix element describing the target atom in the formula for the differential cross section. The matrix elements J_{ij} are averaged over the relative orientation of the momentum transferred in the collision and the symmetry axis of the electronic orbitals of the target atom, using the single-electron Rutaan–Hartree–Fock wave functions. The algebraic representation of the matrix elements J_{ij} makes it possible to perform calculations for atoms with any value of Z. The model developed is used to calculate the cross sections σ_Σ and characteristic scattering angles θ_c for the process of electron loss by H^- ions with energy E = 0.1–100 MeV in targets consisting of atoms with Z = 2–54. It is shown that $\sigma_\Sigma \propto E^{-1}$ and $\theta_c \propto E^{-1/2}$ for all Z, and for fixed E the behavior of $\sigma_\Sigma(Z)$ and $\theta_c(Z)$ is determined by the order of filling of the electronic shells of the target atoms (the ionization potential). The computational results are analyzed and compared with the experimental data and the results of other calculations. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

One of the basic problems in the physics of ion-atom collisions is to establish the dependence of the total scattering cross section for incident particles A scattered by the atomic particles B of the target in a particular process on the atomic number Z of the target particles and the kinetic energy E of the incident particles: $\sigma = \sigma(Z, E)$. Here the term "atomic particles" is understood to be atoms and their ions in the ground and excited states.

Ordinarily, the energy behavior of the cross sections with Z = const is studied for a limited set of targets, which from the experimental standpoint is due to the difficulty of producing gaseous, beam, or plasma targets with strictly controllable thickness for arbitrary chemical elements and from the theoretical standpoint it is due to the sharp increase with increasing Z of the difficulty of describing target atomic particles (and incident particles). These factors play an even greater role in explaining the absence of sufficiently complete experimental and theoretical dependences of the cross sections on the atomic number of the target with fixed kinetic energy of the incident particles:

$$\sigma = \sigma(Z, E = \text{const}). \quad (1)$$

It is obvious that the mutually complementary experimental and theoretical results of investigations of the dependences (1) make it possible to perform a purposeful search for targets suitable for solving diverse applied problems and to check or determine the range of applicability of the theoretical models.

A wide class of ion-atomic collision processes involving two particles A and B can be described by the formula

$$A(\alpha_i) + B(\beta_i) \longrightarrow A(\alpha_f) + B(\Sigma). \quad (2)$$

Here α_i, β_i and α_g, β_f are the initial and final states of the particles, respectively; the symbol Σ means that all possible final states β_f of the target B, which belong to a discrete spectrum (including the ground state) and continuous spectrum are used in the calculation. The formula (2) covers elastic scattering and excitation of the particles A as well as the loss of electrons by particle A. Electron capture, charge transfer, and electron exchange will be neglected.

We shall assume that the velocity of the incident particles A lies in the range bounded on one side by the condition of applicability of the Born approximation and on the other by the requirement that the nonrelativistic theory be applicable. If the interaction of the colliding particles in the process (2) is described by a Coulomb potential, then in the closed Born approximation (closure approximation) the differential scattering cross section for particles A scattered by the angle ν into a solid angle do in the center-of-mass system can be written, using the sum rule over final states of the target particles, in the form [1–3]

$$\frac{d\sigma_{\alpha_f}(\nu)}{do} = \frac{4a_0^2}{(\bar{q}a_0)^4}\left(\frac{M}{m}\right)^2\frac{\bar{k}_f}{k_i}\left|F^{\mathrm{A}}_{\alpha_f\alpha_i}(\bar{q})\right|^2 M(\bar{\mathbf{q}}), \quad (3)$$

where a_0 is the radius of the first Bohr orbit; m and M are the electron mass and the reduced mass of the col-

liding particles; $F^{A}_{\alpha_f\alpha_i}$ is the atomic form factor of the system A; $\mathbf{k}_i$ and $\mathbf{k}_f$ are the wave vectors of the particle A in the center-of-mass system before and after the collision; $\mathbf{q} = \mathbf{k}_f - \mathbf{k}_i$; $\bar{\mathbf{q}}$ and $\bar{\mathbf{k}}_f$ are the average (with respect to β_f) values of the corresponding vectors, which are determined by the method indicated in [2]. The quantity $M(\mathbf{q})$ in Eq. (3) is the matrix element determined for the initial state of the particle B:

$$M(\mathbf{q}) = \langle\psi_{\beta_i}|\left|Z - \sum_{b=1}^{N}\exp(i\mathbf{q}\cdot\mathbf{r}_b)\right|^2|\psi_{\beta_i}\rangle, \qquad (4)$$

where ψ_{β_i} is the wave function of the initial state of particle B; N is the number of electrons belonging to the particle B; $\mathbf{r}_b$ is the radius vector of the bth electron. The problem of obtaining the dependences (1) using the closure approximation is related with the difficulty of calculating the matrix elements (4) for the particles B with a large number of electrons.

A remarkable feature of the relation (3) is that it can be factorized into three cofactors: the first one contains the fundamental constants m, a_0, the parameters of the problem M and k_i, and the variable quantities $\bar{\mathbf{q}}$ and $\bar{k}_f$; the second cofactor is the form factor F^{A} (and/or the incoherent scattering function [1, 2]) and describes only the particle A; and, the third cofactor is the matrix element $M(\bar{\mathbf{q}})$ and contains quantities characterizing only the particle B. Thus, determining the expression making it possible to calculate the matrix element (4) solves, in principle, the problem of determining the dependences $\sigma(Z, E)$ for any processes of the type (2).

In the present paper an algebraic expression is obtained for the matrix element (4) that makes it possible to perform calculations of the interaction cross sections of particles in processes of the type (2) for atomic target particles with arbitrary atomic number Z. Calculations of the cross sections (1) for the range Z = 2–54 were performed for the process $(\bar{1}\,0) + (\bar{1}\,1)$ of electron loss by the negative hydrogen ions H^- with energy E = 10 MeV.

2. THEORY

As we have seen above, the problem of calculating the scattering cross sections in the closed variant of the Born approximation (3) requires finding the matrix element $M(\mathbf{q})$. Let $M(\mathbf{q})$ be determined for the particle B, which in the general case can be an ion with nuclear charge Z and N electrons. We shall write the operator part of the matrix element $M(\mathbf{q})$ in the expanded form

$$\left|Z - \sum_{b=1}^{N}\exp(i\mathbf{q}\cdot\mathbf{r}_b)\right|^2 = Z^2 - Z\sum_{b=1}^{N}[\exp(i\mathbf{q}\cdot\mathbf{r}_b) + \exp(-i\mathbf{q}\cdot\mathbf{r}_b)] + \sum_{b,c=1}^{N}\exp(i\mathbf{q}\cdot(\mathbf{r}_c - \mathbf{r}_b)). \qquad (5)$$

We shall assume that the wave function of the particle B is a Slater determinant of single-electron wave functions, and the particle itself at the moment of collision can be in an arbitrary state. We shall designate the single-electron states by the index i, which represents the entire set of quantum numbers required for a complete description of the state: $|i\rangle \equiv |nlm\sigma\rangle$, where n and l are the principal and orbital angular momentum quantum numbers, and m and σ are the projections of the orbital and spin angular momenta of the electron. The use of the letter i to denote $(-1)^{1/2}$ and the single electron state will not lead to any confusion, since a state will be labeled by the letter i only as an index.

We substitute the expression (5) into the expression for the matrix element $M(\mathbf{q})$:

$$M(\mathbf{q}) = Z^2 - Z\sum_{b=1}^{N}(\langle\psi|\exp(i\mathbf{q}\cdot\mathbf{r}_b)|\psi\rangle + \langle\psi|\exp(-i\mathbf{q}\cdot\mathbf{r}_b)|\psi\rangle) + N + \sum_{\substack{b,c=1\\(b\neq c)}}^{N}\langle\psi|\exp(i\mathbf{q}\cdot(\mathbf{r}_c - \mathbf{r}_b))|\psi\rangle. \qquad (6)$$

Since $|\psi|^2$ is a symmetric function under the inversion of the vector $\mathbf{r}_b$, we have

$$\langle\psi|\exp(i\mathbf{q}\cdot\mathbf{r}_b)|\psi\rangle = \langle\psi|\exp(-i\mathbf{q}\cdot\mathbf{r}_b)|\psi\rangle, \qquad (7)$$

using which we obtain for the matrix element (6)

$$M(\mathbf{q}) = Z^2 + N - 2Z\sum_{b=1}^{N}\langle\psi|\exp(i\mathbf{q}\cdot\mathbf{r}_b)|\psi\rangle + 2\sum_{b=1}^{N-1}\sum_{c=b+1}^{N}\langle\psi|\exp(i\mathbf{q}\cdot(\mathbf{r}_c - \mathbf{r}_b))|\psi\rangle. \qquad (8)$$

We shall now consider the matrix element in the third term of the relation (8) under the summation sign,

expanding ψ as a Slater determinant and remembering that the normalization factor of the function ψ is $(N!)^{-1/2}$:

$$\langle\psi|\exp(i\mathbf{q}\cdot\mathbf{r}_b)|\psi\rangle = \frac{1}{N!} \times\int\left[\sum_{s=1}^{N!} P_s^{(t)}(-1)^s\psi^*_{t_1}(\mathbf{r}_1)\ldots\psi^*_{t_i}(\mathbf{r}_b)\ldots\psi^*_{t_N}(\mathbf{r}_N)\right] \times\exp(i\mathbf{q}\cdot\mathbf{r}_b) \times\left[\sum_{u=1}^{N!} P_u^{(\nu)}(-1)^u\psi_{\nu_1}(\mathbf{r}_1)\ldots\psi_{\nu_j}(\mathbf{r}_b)\ldots\psi_{\nu_N}(\mathbf{r}_N)\right]d\tau, \quad (9)$$

where the integration extends over the configuration space of N electrons of the particle B; $P_s^{(t)}$ is an operator which interchanges elements of the ordered set t_1, t_2, …, t_N, consisting of the elements 1, 2, …, N; the symbol t in the interchange operator $P_s^{(t)}$ indicates subsequent functions with elements t_k to which the action of the operator extends, and similarly for the operator $P_u^{(\nu)}$. In Eq. (9) the interchange operators act on the indices enumerating the single-electron states of the particle B.

Since the single-electron wave functions are orthogonal to one another, in the expression (9) only the following terms remain different from zero:

$$\langle\psi|\exp(i\mathbf{q}\cdot\mathbf{r}_b)|\psi\rangle = \frac{1}{N!}\int\exp(i\mathbf{q}\cdot\mathbf{r}_b) \times\sum_{s=1}^{N!} P_s^{(t)}\left|\psi_{t_1}(\mathbf{r}_1)\right|^2\ldots\left|\psi_{t_i}(\mathbf{r}_b)\right|^2\ldots\left|\psi_{t_N}(\mathbf{r}_N)\right|^2 d\tau. \quad (10)$$

If the bth electron with radius vector $\mathbf{r}_b$ is fixed in the ith state (the wave function with index t_i), then the operator $P_s^{(t)}$ will interchange the remaining $N-1$ indices of states for single-electron wave functions, each of which depends on one of the $N-1$ remaining radius vectors different from $\mathbf{r}_b$. The number of ways that $N-1$ electrons can be placed into $N-1$ states without any repetitions is $(N-1)!$, so that

$$\langle\psi|\exp(i\mathbf{q}\cdot\mathbf{r}_b)|\psi\rangle = \frac{1}{N!}\sum_{i=1}^{N}\int\exp(i\mathbf{q}\cdot\mathbf{r}_b)\left|\psi_{t_i}(\mathbf{r}_b)\right|^2 \times\sum_{s=1}^{(N-1)!} P_s^{(t)}\left|\psi_{t_1}(\mathbf{r}_1)\right|^2\ldots\left|\psi_{t_N}(\mathbf{r}_N)\right|^2 d\tau. \quad (11)$$

The integral over the product of $N-1$ cofactors of the form $\left|\psi_{t_1}(\mathbf{r}_1)\right|^2\ldots\left|\psi_{t_N}(\mathbf{r}_N)\right|^2$ is 1 because the single-electron wave functions are orthonormal. The number of terms consisting of products of this type is, evidently, $(N-1)!$, so that

$$\langle\psi|\exp(i\mathbf{q}\cdot\mathbf{r}_b)|\psi\rangle = \frac{1}{N}\sum_{i=1}^{N} J_i(\mathbf{q}), \quad (12)$$

where

$$J_i(\mathbf{q}) \equiv J_{ii} = \langle\psi|\exp(i\mathbf{q}\cdot\mathbf{r}_b)|\psi\rangle = \int\exp(i\mathbf{q}\cdot\mathbf{r}_b)\left|\psi_i(\mathbf{r})\right|^2 dV. \quad (13)$$

It is easy to see that $J_i(-\mathbf{q}) = J_i^*(\mathbf{q}) = J_i(\mathbf{q})$, and hence J_i is a real function. We note that the right-hand side of the equation (13) is independent of the index b, as should be the case, so that

$$\sum_{b=1}^{N}\langle\psi|\exp(i\mathbf{q}\cdot\mathbf{r}_b)|\psi\rangle = \sum_{i=1}^{N} J_i(\mathbf{q}). \quad (14)$$

We shall now consider the matrix element in the fourth term in the formula (8) just as was done in the derivation of the expressions (9)–(11). But, since the exponential operator of this matrix element contains the radius vectors of the bth and cth electrons, this matrix element will contain two groups of terms (the others are equal to zero because of the orthogonality of the single electron wave functions):

$$\langle\psi|\exp(i\mathbf{q}\cdot(\mathbf{r}_c-\mathbf{r}_b))|\psi\rangle = \frac{1}{N!}\int\exp(i\mathbf{q}\cdot(\mathbf{r}_c-\mathbf{r}_b)) \times\sum_{s=1}^{N!} P_s^{(t)}\left|\psi_{t_1}(\mathbf{r}_1)\right|^2\ldots\left|\psi_{t_i}(\mathbf{r}_c)\right|^2\ldots\left|\psi_{t_j}(\mathbf{r}_b)\right|^2 \ldots\left|\psi_{t_N}(\mathbf{r}_N)\right|^2 d\tau + \frac{1}{N!}\int\exp(i\mathbf{q}\cdot(\mathbf{r}_c-\mathbf{r}_b)) \times\sum_{s=1}^{N!} P_s^{(t)}\left|\psi_{t_1}(\mathbf{r}_1)\right|^2\ldots\psi^*_{t_i}(\mathbf{r}_c)\psi^*_{t_j}(\mathbf{r}_b) \ldots\psi_{t_i}(\mathbf{r}_b)\psi_{t_j}(\mathbf{r}_c)\ldots\left|\psi_{t_N}(\mathbf{r}_N)\right|^2(-1)^{3s+1}d\tau. \quad (15)$$

The appearance of the cofactor $(-1)^{2s+1}\equiv -1$ in the second term in the expression (15) is due to the fact that the terms of the Slater determinant appear in Eq. (15) as a sum of pairwise products, differing from one another only by one interchange (and by the complex conjugation sign).

Let us now fix the bth and cth electrons in the ith or jth states and interchange the indices of the states for the remaining single-electron wave functions (with the order of the radius vectors of the electrons remaining the same). The indices are chosen from the indices corresponding to the remaining $N-2$ states (except the ith and jth states). In other words we perform all possible interchanges of the remaining $N-2$ electrons over the

remaining $N-2$ states. The number of arrangements without repetition is $(N-2)!$. Ultimately, we find

$$\langle\psi|\exp(i\mathbf{q}\cdot(\mathbf{r}_c-\mathbf{r}_b))|\psi\rangle = \frac{1}{N(N-1)} \times \sum_{\substack{i,j=1\\(i\neq j)}}^{N}[J_i(\mathbf{q})J_j(\mathbf{q})-J_{ij}(\mathbf{q})J_{ji}(-\mathbf{q})], \tag{16}$$

where

$$J_{ij}(\mathbf{q}) = \langle\psi_i|\exp(i\mathbf{q}\cdot\mathbf{r})|\psi_j\rangle = \int\psi_i^*(\mathbf{r})\exp(i\mathbf{q}\cdot\mathbf{r})\psi_j(\mathbf{r})dV, \tag{17}$$
$$J_{ji}(-\mathbf{q}) = J_{ij}^*(\mathbf{q}).$$

The wave functions $\psi_i(\mathbf{r})$ and $\psi_j(\mathbf{r})$ from Eq. (17) are either symmetric or antisymmetric with respect to inversion of the vector $\mathbf{r}$. If the states i and j have the same symmetry under inversion of the vector $\mathbf{r}$, then it follows from Eq. (17) that $J_{ij}(\mathbf{q}) = J_{ij}(-\mathbf{q})$. However, if the states i and j have the opposite symmetry, then $J_{ij}(\mathbf{q}) = -J_{ij}(-\mathbf{q})$. Consequently, in any case, using Eq. (17) we have

$$|J_{ij}(\mathbf{q})|^2 = [\pm J_{ji}^*(-\mathbf{q})][\pm J_{ji}(-\mathbf{q})] = |J_{ij}(-\mathbf{q})|^2 = |J_{ji}(\mathbf{q})|^2. \tag{18}$$

The right-hand side of Eq. (16) once again is independent of the indices b and c, and consequently we find, taking account of Eqs. (13) and (17), that

$$2\sum_{b=1}^{N-1}\sum_{c=b+1}^{N}\langle\psi|\exp(i\mathbf{q}\cdot(\mathbf{r}_c-\mathbf{r}_b))|\psi\rangle = \sum_{\substack{i,j=1\\(i\neq j)}}^{N}J_iJ_j - \sum_{\substack{i,j=1\\(i\neq j)}}^{N}|J_{ij}|^2. \tag{19}$$

Here we used the equality

$$\sum_{b=1}^{N-1}\sum_{c=b+1}^{N}1 = \frac{N(N-1)}{2}.$$

We now substitute the relations (14) and (19) into Eq. (8):

$$M(\mathbf{q}) = Z^2+N-2Z\sum_{i=1}^{N}J_i + \sum_{\substack{i,j=1\\(i\neq j)}}^{N}(J_iJ_j-|J_{ij}|^2). \tag{20}$$

No simplifying assumptions were made in deriving the expression (20). However, it is obvious that the functions J_{ij} appearing in it are simpler than the matrix elements summed in Eq. (8).

The expression (20) can be simplified by using the relation

$$\sum_{\substack{i,j=1\\(i\neq j)}}^{N}J_iJ_j = 2\sum_{i=1}^{N-1}J_i\sum_{j=i+1}^{N}J_j = \left(\sum_{i=1}^{N}J_i\right)^2-\sum_{i=1}^{N}J_i^2, \tag{21}$$

which can be easily obtained noting that, on the one hand,

$$\sum_{i=1}^{N-1}J_i\sum_{j=i+1}^{N}J_j = \sum_{i=1}^{N-1}J_i\left(\sum_{j=i}^{N}J_j-J_i\right) = \sum_{i=1}^{N-1}J_i\sum_{j=i}^{N}J_j-\sum_{i=1}^{N}J_i^2 = \sum_{i=1}^{N}J_i\sum_{j=i}^{N}J_j-\sum_{i=1}^{N}J_i^2, \tag{22}$$

while on the other hand

$$\sum_{i=1}^{N-1}J_i\sum_{j=i+1}^{N}J_j = \sum_{i=2}^{N}J_i\sum_{j=1}^{i-1}J_j = \sum_{i=2}^{N}J_i\left(\sum_{j=1}^{i}J_j-J_i\right) = \sum_{i=1}^{N}J_i\sum_{j=1}^{i}J_j-\sum_{i=1}^{N}J_i^2. \tag{23}$$

Adding Eqs. (22) and (23) we obtain the expression (21), substituting which into Eq. (20) we find

$$M(\mathbf{q}) = Z^2+N-2ZD+D^2-G-\sum_{\substack{i,j=1\\(i\neq j)}}^{N}|J_{ij}|^2, \tag{24}$$

$$D = \sum_{i=1}^{N}J_i, \quad G = \sum_{i=1}^{N}J_i^2. \tag{25}$$

We shall now find the matrix elements J_{ij}. The wave functions from Eq. (17) have the form [4, 5] $\psi_i(\mathbf{r}) = R_{nl}(r)Y_{lm}(\theta,\varphi)\chi(\sigma)$, where $R_{nl}(r)$ is a radial wave function; $Y_{lm}(\theta,\varphi) = \Theta_{lm}(\theta)\Phi_m(\varphi)$ is a spherical function; $\Phi_m(\varphi) = (2\pi)^{-1/2}e^{im\varphi}$; and, $\chi(\sigma)$ is a spin function. The presence of spin functions in the expression for J_{ij} leads to the appearance of the Kronecker delta function $\delta_{\sigma\sigma'}$.

The spherical coordinates $\mathbf{r}$, θ, and φ of the electrons in the target atom are defined using the right-hand vector triplet ($\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$) centered on the nucleus of the atom. The $\mathbf{z}$-axis is conventionally chosen as the polar axis and is the axis of symmetry of the single-electron

orbitals of the target atom. In the same spherical coordinate system the direction of the wave vector **q** is given by the polar angle α and the azimuthal angle β. Then the scalar product in the argument of the exponential in Eq. (17) will be

$$(\mathbf{q}\cdot\mathbf{r}) = qr[\cos\theta\cos\alpha + \sin\theta\sin\alpha\cos(\varphi-\beta)]. \quad (26)$$

The relative spatial orientation of the vectors **q** and **z** can be arbitrary, so that the differential scattering cross sections (3) and hence the matrix elements $M(\mathbf{q})$ must be averaged over the directions **q** relative to the **z** axis. The probability of finding the vector **q** inside the solid angle $d\Omega_q = \sin\alpha\, d\alpha\, d\beta$ is $d\Omega_q/4\pi$. The problem of averaging $M(\mathbf{q})$ must be simplified. To this end, we shall use the approximate statistical method, according to which the average value of $M(\mathbf{q})$ is taken to be the value obtained by averaging the matrix elements J_{ij} and substituting them into the formula (24) in order to calculate $M(\mathbf{q})$.

The calculation and averaging of the matrix elements J_{ij} can be performed analytically. The integration can be done efficiently with respect to the variables in the following order: β, φ, α, θ, and r. We shall continue to use the previous notation for the average matrix element J_{ij}. Omitting the laborious but essentially simple calculations, using tabulated integrals, we present for J_{ij} an expression found after integrating over the angular variables:

$$J_{ij} = \frac{1}{q}\int_0^\infty R_{nl}R_{n'l}\sin(gr)\,r\,dr. \quad (27)$$

We note that in the integration the Kronecker symbols $\delta_{mm'}$ and $\delta_{ll'}$ with respect to the magnetic and orbital quantum numbers appear, and the matrix elements J_{ij} themselves do not depend on m, which was used in writing down the formula (27). The single-electron wave functions R_{nl} and $R_{n'l'}$, appearing in the Slater determinant and the expression (27), can be written in the form [5]

$$R_{nl} = \sum_c a_c r^{n_c-1}\exp(-\zeta_c r),$$
$$R_{n'l'} = \sum_d a'_d r^{n'_d-1}\exp(-\zeta'_d r). \quad (28)$$

Substituting the functions R_{nl} and $R_{n'l}$ in the form (28) into the integral (27), performing the elementary integration, and using Newton's binomial formula, we obtain the final relation for the average matrix element J_{ij}:

$$J_{ij} = \sum_{c,d}\frac{a_c a'_d (n_{cd}-1)!}{\zeta_{cd}[\zeta_{cd}(1+x_{cd})]^{n_{cd}}}\sum_{s=0}^{S} C_{n_{cd}}^{2s+1}(-x_{cd})^s, \quad (29)$$

where $n_{cd} = n_c + n'_d$, $\zeta_{cd} = \zeta_c + \zeta'_d$, $x_{cd} = q^2a_0^2/\zeta_{cd}^2$, and $C_{n_{cd}}^{2s+1}$ are the binomial coefficients, and the upper limit of the summation is

$$S = \begin{cases} \dfrac{n_{cd}-1}{2}, & \text{if } n_{cd} \text{ is an odd number} \\ \dfrac{n_{cd}}{2}-1, & \text{if } n_{cd} \text{ is an even number.} \end{cases} \quad (30)$$

The radius r in the formulas (28) is usually measured in atomic units, as a result of which the factor a_0 appears in the expression for x_{cd} from Eq. (29).

3. CALCULATION OF THE CROSS SECTIONS. ANALYSIS OF THE RESULTS

We now turn to the calculation of the sums over i and j in Eq. (24) for $M(\mathbf{q})$, using the average matrix elements J_{ij}. It is convenient to switch from summation over the state numbers i and j to summation over the corresponding quantum numbers n, l, m, and σ, and to introduce the occupation function for the state $|nlm\sigma\rangle$:

$$\mu_{nlm\sigma} = \begin{cases} 1, & \text{if the state is occupied} \\ 0, & \text{if the state is unoccupied.} \end{cases} \quad (31)$$

Let n_e be the principal quantum number of the outer shell of the particle B, i.e., the maximum value of n from the set of occupied states. Then we obtain for the sum D from Eqs. (24) and (25) the expression (in what follows, $J_i = J_{nl}$)

$$D = \sum_{n=1}^{n_e}\sum_{l=0}^{n-1} J_{nl}\sum_{m=-l}^{l}\sum_{\sigma=-1/2}^{1/2}\mu_{nlm\sigma}, \quad (32)$$

where we took account of the fact that the matrix elements J_{ij} do not depend on the quantum number m and the value of the spin σ, so that here and below the numbers m and σ are dropped from the indices of J. We find a similar relation for the sum G from Eqs. (24) and (25).

We represent the last sum in the equality (24), using the property (18) and the presence of the Kronecker symbol $\delta_{ll'}$ in the formula for the average matrix elements $J_{ij} = J_{nl,n'l}$, in the form

$$2V = \sum_{\substack{i,j=1\\(i\neq j)}}^{N}|J_{ij}|^2 = 2\sum_{nlm\sigma}\sum_{\substack{n'lm\sigma\\(n'>n)}}|J_{nl,n'l}|^2, \quad (33)$$

where

$$V = \sum_{n=1}^{n_e-1}\sum_{n'=n+1}^{n_e}\sum_{l=0}^{n-1} J_{nl,n'l}^2\sum_{m=-l}^{l}\sum_{\sigma=-1/2}^{1/2}\mu_{nlm\sigma}\mu_{n'lm\sigma}. \quad (34)$$

The modulus symbol for the matrix element J in Eq. (34) is omitted since the matrix elements (29) are real.

In the present work the total cross sections $\sigma_\Sigma = \sigma_{\bar{1}0} + \sigma_{\bar{1}1}$ and the characteristic scattering angles θ_c of H^- ions with energy E = 0.1–100 MeV scattered by atoms with Z = 2–54 were calculated, on the basis of the theory described above, for the process $(\bar{1}\,0) + (\bar{1}\,1)$ in which one and two electrons are lost (in what follows, the cross sections σ_Σ and the angles θ_c calculated using the closed Born approximation will not be given any distinguishing indices). The characteristic scattering angle θ_c of the particles is the value of the angle θ for which the function $\sin\theta\, d\sigma(\theta)/d\Omega$, is maximum, where θ is the scattering angle in the laboratory coordinate system. The equation (3), summed over all possible final states $\alpha_f \neq \alpha_i$ of the incident particle and written in the laboratory coordinate system, was used to calculate σ_Σ and θ_c. After summing over α_f the squared modulus of the form factor of the incident particle A in Eq. (3) is replaced by the incoherent scattering function [1, 2], which is determined by the wave function of the ground state of the particle A. In the present work, just as in [3], the ground state of the H^- ions was described by the Chandrasekar function.

The matrix elements (4) from Eq. (3) were determined according to Eq. (24), which, we recall, is exact and in which the expressions (32)–(34) were used to find the corresponding terms. Finally, the functions J_{ij} were calculated from Eqs. (32) and (34) and from the analogous formula for the sum G (see Eq. (24)) using the relation (29). We note that as $q \longrightarrow 0$ the matrix elements $J_{ii} \longrightarrow 1$, and $J_{ij} \longrightarrow 0$; therefore, for the functions D, G, and V for target atoms with Z in the same limit $q \longrightarrow 0$ we obtain $D \longrightarrow Z$, $G \longrightarrow Z$, and $V \longrightarrow 0$ (see Eqs. (25) and (33)). A check of these limits serves as a criterion for the correctness of the computational algorithm and the wave-function parameters introduced.

In addition, in the present work similar calculations of the cross sections σ_Σ^d and angles θ_c^d were performed for the same ranges of the energies E and numbers Z in the dipole-moment approximation for describing the target atom (the indicator cross sections and angles are labeled with an additional index d). It is assumed in the dipole-moment approximation that during an ion-atom collision the atom can be treated as an electric dipole with an effective dipole moment d. The theory for the dipole-moment approximation is described in detail in [6, 7], so that only the results of the corresponding calculations will be presented here.

In both computational models the single-electron states of the target atom were described by the Rutaan–Hartree–Fock wave functions [5].

The cross sections and characteristic scattering angles for 10 MeV H^- ions as functions of Z, obtained on the basis of both theoretical approaches, are presented in the table and in Figs. 1 and 2. The value E = 10 MeV was chosen because, in the first place, the con-

Characteristic angles in scattering cross sections for the process $(\bar{1}0) + (\bar{1}1)$ in which 10 MeV H^- ions lose electrons on interacting with atomic targets with atomic numbers Z = 2–54 (θ_c and σ_Σ were calculated using the theory presented in the present paper, and the cross section σ_Σ^d was calculated in the dipole-moment approximation [6, 7])

Z	Target atom	θ_c, 10^{-6} rad	σ_Σ, 10^{-16} cm^2	σ_Σ^d, 10^{-16} cm^2
2	He	10.2	0.0546	0.0500
3	Li	8.12	0.199	0.464
4	Be	9.81	0.279	0.426
5	B	10.8	0.334	0.384
6	C	11.6	0.363	0.328
7	N	12.2	0.380	0.313
8	O	12.5	0.400	0.322
9	F	12.8	0.411	0.320
10	Ne	13.0	0.418	0.312
11	Na	9.28	0.604	0.684
12	Mg	11.0	0.746	0.757
13	Al	12.1	0.902	0.948
14	Si	14.0	1.01	0.887
15	P	16.6	1.10	0.888
16	S	19.0	1.19	0.916
17	Cl	22.2	1.25	0.884
18	Ar	25.1	1.31	0.867
19	K	11.4	1.60	1.41
20	Ca	12.8	1.85	1.55
21	Sc	13.8	1.92	1.39
22	Ti	14.5	1.87	1.34
23	V	15.3	2.01	1.30
24	Cr	18.4	1.89	1.26
25	Mn	16.7	2.05	1.25
26	Fe	17.4	2.09	1.23
27	Co	20.8	1.99	1.21
28	Ni	19.1	2.13	1.19
29	Cu	22.5	2.03	1.18
30	Zn	20.8	2.16	1.03
31	Ga	18.6	2.34	1.45
32	Ge	20.7	2.49	1.50
33	As	23.0	2.62	1.50
34	Se	24.6	2.75	1.52
35	Br	26.3	2.88	1.52
36	Kr	28.1	2.99	1.50
37	Rb	14.7	3.37	2.30
38	Sr	14.9	3.72	2.56
39	Y	16.4	3.91	2.26
40	Zr	17.9	4.04	1.97
41	Nb	22.8	4.01	1.86
42	Mo	24.6	4.10	2.05
43	Tc	21.9	4.36	2.03
44	Ru	27.0	4.29	2.00
45	Rh	28.2	4.37	1.97
46	Pd	33.9	4.28	1.95
47	Ag	30.4	4.50	1.91
48	Cd	27.6	4.72	2.07
49	In	27.8	4.53	–
50	Sn	24.5	5.24	2.35
51	Sb	25.4	5.46	2.38
52	Te	26.0	5.69	2.42
53	I	27.0	5.90	2.43
54	Xe	28.2	6.09	2.43

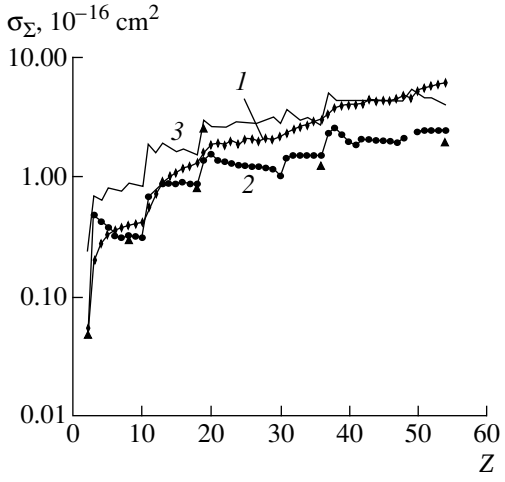

Fig. 1. Cross section for the loss of electrons by H^- ions with energy E = 10 MeV interacting with atoms with atomic numbers Z = 2–54: (*1*) calculation using the theory developed in the present paper; (*2*) calculation in the dipole-moment approximation [6, 7]; (*3*) calculation using Eq. (54); (▲) experimental data for E = 10.4 MeV [8] (the result for a potassium target is taken from [6] for E = 5.14 MeV and recalculated for energy E = 10 MeV assuming that the cross section for K and neighboring Ar targets have the same energy dependence).

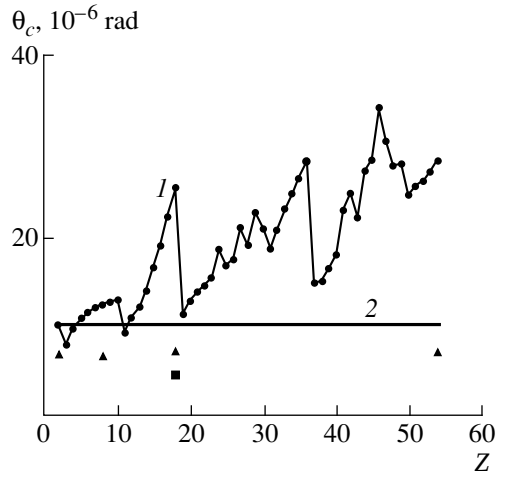

Fig. 2. Characteristic scattering angle for hydrogen atoms in the process where H^- ions with energy E = 10 MeV lose electrons versus the atomic number Z of the target: (*1*) computational results obtained using the theory developed in the present paper; (*2*) calculation in the dipole-moment approximation [6, 7]; (■) calculation in the three-particle Born approximation for the angle $\theta_{1/2}$ [9]; (▲) experimental data for $\theta_{1/2}$ from [10].

dition $v > 2Zv_0$ for the applicability of the Born approximation for target atoms with large Z (here v is the ion velocity, $v_0 = 2.19 \times 10^8$ cm/s, i.e., E must be increased; see [11, 12]) must be satisfied better; in the second place, the correction factors for calculating the cross sections in the dipole-moment approximation play a negligible role for this value of the energy [7]; and, in the third place, for this value of the energy experimental data are available for gas targets.

Both theoretical models show that for a fixed value of E the cross section $\sigma_\Sigma(Z)$ increases abruptly for atoms neighboring inert gases in the first and second groups of the periodic system. For each such transition the cross section increases, to a first approximation, by the same amount, so that the relative growth of the cross section is greatest for the transition from He to Li. This result is confirmed by the experimental investigations performed in [13] for E = 30–200 keV.

The values found in the dipole-moment approximation (Fig. 1) as well as in the free-collisions approximation [14–16] on the whole agree better with existing experimental data for inert-gas targets than the cross sections calculated using the theory presented in this paper. However, the closed Born approximation is a systematic theory, whose approximate character is due only to the natural limitations on the velocity of the incident particle and the use of the sum rules over the final states of the colliding particles. The final construction of the closed Born approximation, essentially associated with obtaining the exact algebraic expression for the average matrix element $M(q)$, will make it possible to determine the accuracy and range of applicability of this approximation. For He, Li, and Be atoms in the ground state and containing electrons only in spherically symmetric s orbitals, the matrix elements (4) do not depend on the relative orientation of the vectors $\mathbf{q}$ and $\mathbf{z}$, so that for these atoms the average matrix elements $M(\mathbf{q})$ and the corresponding scattering cross sections are exact in the closed Born approximation.

The following formula is proposed in [17] for describing the dependences $\sigma_{\bar{1}0}(Z, E)$:

$$\sigma_{\bar{1}0} = N_i \pi a_0^2 \frac{Z^{\alpha(Z)}}{v^\gamma u_i u(Z)}, \qquad (35)$$

where N_i is the number of equivalent electrons in the incident ion; $\alpha(Z)$, $\gamma \equiv 1$ are parameters; $u_i = \sqrt{I_i/I_0}$ is the average orbital velocity of the removed electron for the incident ion; $u(Z) = \sqrt{I(Z)/I_0}$ is the average orbital velocity of the outer electron for an atom in a medium with Z; I_0 = 13.6 eV; I_i is the binding energy of the electron in the ion shell; $I(Z)$ is the ionization potential of an atom in the medium; all velocities in Eq. (35) are in atomic units v_0. Figure 1 displays the computational results obtained using Eq. (35) for N_i = 1, I_i = 0.754 eV, and $\alpha(Z)$ = 0.75 (see [17]). The cross sections (35) with E = 10 MeV are 2–4 times greater than the experimental values for inert-gas targets. A general comparison of the dependence (35) with the measurement results (see below) leads to the conclusion that the exponent γ is a function of Z and E.

As one can see from Fig. 2, the behavior of the function $\theta_c(Z, E = \text{const})$ with increasing Z is determined by the sequence in which the electronic shells of the target atom are filled. However, according to the physical model which is the basis of the dipole-moment approximation, the angle θ_c^d does not depend on the type of target. Besides θ_c, the angles $\theta_{1/2}$, corresponding to the half-width of the differential scattering cross section of the particles at half-height were calculated using the closed Born approximation. The calculations showed that the theoretical dependence of the angles $\theta_{1/2}$ on the energy of the H^- ions and the atomic number Z of the target does not agree with the experimental facts. The importance of this situation is indicated in [2]. The point is that the closed Born approximation is based on the use of the sum rules over the final states of the colliding particles, and as a result the final states which do not satisfy the laws of conservation of energy and momentum are automatically included in the analysis. The error arising in the calculation of the differential scattering cross sections as a result of the use of the sum rules will be all the larger the smaller the average momentum transferred in a collision, i.e., when $\theta \longrightarrow 0$. However, the value of $d\sigma(\theta = 0)/d\Omega$ is used to determine the angle $\theta_{1/2}$, which is why the methodological error in calculating $\theta_{1/2}$ is higher than for θ_c. For these reasons the computational results for $\theta_{1/2}$ are not presented below.

The results of our calculations of the characteristic scattering angles are compared in Fig. 2 with the experimental values from [10] for the angles $\theta_{1/2}$, corresponding to the half-width at half-height of the spatial-angular distribution of hydrogen atoms, obtained in the $(\bar{1}\,0)$ process of neutralization of H^- ions in a CO_2 target. It should be noted that in [10] the measurements of $\theta_{1/2}$ were performed for a ribbon beam of particles. The experimental values of $\theta_{1/2}$ do not depend, within the limits of the measurement error, on the number Z of the target atom ([10]; see also [9, 18]) and fall substantially below the theoretical values of θ_c. This latter circumstance is explained primarily by the fact that the closed Born approximation and the dipole-moment approximation are two-particle approximations, i.e., the angle θ gives in these models the direction of motion of the center of mass of the particles formed in the process where the H^- ion loses one or two electrons. However, experimentally, the distribution of H^0 ions is measured for the process $(\bar{1}\,0)$. Analysis of the process $(\bar{1}\,0)$ in the three-particle Born approximation [9] shows that the form of the differential scattering cross section of hydrogen atoms is essentially independent of the target choice, and the computed value of the angle $\theta_{1/2} \propto E^{-1/2}$ and is approximately 1.5 times lower than the experimental data for a ribbon beam of particles (Fig. 2).

The energy dependences of the cross sections σ_Σ and the characteristic angles θ_c, calculated for He and Xe targets, are presented in Figs. 3 and 4. For target atoms with arbitrary Z the cross sections behave according to a law close to $\sigma_\Sigma \propto E^{-1}$, and the angles behave as $\theta_c \propto E^{-1/2}$ (with the exception of a helium target with $E < 1$ MeV; see Fig. 4). The behavior $\theta_c \propto E^{-1/2}$ is

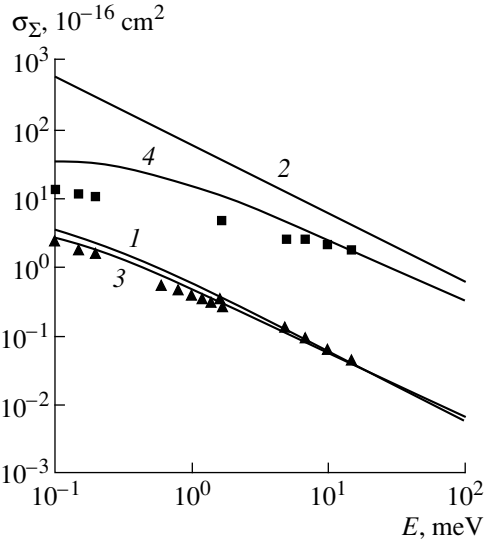

Fig. 3. Cross sections for loss of electrons by H^- ions versus the energy E for targets consisting of He and Xe atoms: (*1*) and (*2*) computational results obtained using the theory in the present paper for He and Xe atoms, respectively; (*3*) and (*4*) calculations in the dipole-moment approximation for He and Xe atoms, respectively; (▲) experimental data for He from [8, 13, 19]; (■) experimental data for Xe from [8, 13].

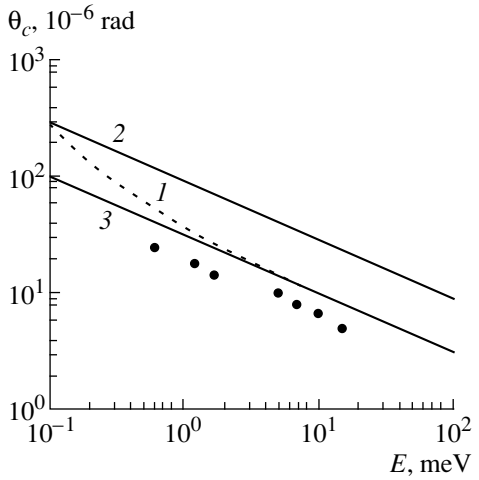

Fig. 4. Energy dependences of the characteristic scattering angles of hydrogen particles in a process in which H^- ions lose electrons for a target consisting of He and Xe atoms: (*1*) and (*2*) computational results obtained using the theory of the present paper for He and Xe atoms, respectively; (*3*) calculation in the dipole-moment approximation for a He target [6, 7]; (●) experimental data for the angle $\theta_{1/2}$ corresponding to the half-width at half height of the spatial-angular distribution of hydrogen atoms, appearing in the scattering of H^- ions scattered by a CO_2 target [10].

identical to the experimental behavior $\theta_{1/2} \propto E^{-1/2}$ [10, 18]. The cross sections σ_Σ and the angles θ_c found in the present work for a He target are essentially identical to the analogous calculations performed in [3]. For light target atoms (He, Li) the dependence $\sigma_\Sigma(E)$ agrees well with the experimental data in the entire range of applicability of the closed Born approximation with respect to E. As Z increases, the approximate experimental dependence $\sigma_\Sigma^{\exp} \propto E^{-n(Z)}$ becomes increasingly flatter, i.e., the exponent $n(Z)$ systematically decreases (from $n \approx 1$ for He to $n \approx 0.45$ for Xe). The discrepancy between the theoretical and experimental values for the cross sections increases with increasing Z and decreasing E. This is explained by the fact that the Born approximation neglects the motion of the electrons in the colliding particles and the dynamics of the ion-atom collisions.

4. CONCLUSIONS

(1) In the variant of the closed Born approximation developed above, the matrix element (4), responsible for the description of target atom, reduces, as a result of identity transformations, to an expression containing the matrix elements J_{ij} with respect to single-electron states of the atom (Eqs. (20) and (24)), and averaging of the matrix element $M(\mathbf{q})$ over the directions of the vector $\mathbf{q}$ relative to the $\mathbf{z}$-axis of the atom is performed approximately, specifically, by substituting into Eq. (24) the average (with respect to the relative orientation of the vectors $\mathbf{q}$ and $\mathbf{z}$) matrix elements J_i and J_{ij} (Eq. (29)). The single-electron Rutaan–Hartree–Fock single-electron wave functions were used to average the matrix elements J_{ij}. Ultimately, the relation for the differential scattering cross section was reduced to an algebraic expression, which makes it possible to perform calculations of the scattering cross sections of particles scattered by target atoms with arbitrary atomic number Z. For He, Li, and Be atoms, which in the ground state contain electrons in spherically symmetric s orbitals, the computational results for the matrix elements $M(\mathbf{q})$ are exact within the framework of the closed Born approximation.

(2) Systematic calculations were performed of the cross sections σ_Σ and characteristic angles θ_c for the process where H^- ions with energies $E = 0.1$–100 MeV lose electrons in collisions with atoms of a target with atomic numbers $Z = 2$–54. According to the calculations, when inert gas atom targets are replaced by targets with atoms of the neighboring alkali metal the cross section σ_Σ increases abruptly and the angle θ_c decreases; a beam of hydrogen atoms obtained by neutralizing H^- ions in a lithium vapor target will possess the best angular characteristics (if scattering of H^- ions and H^0 atoms without a change in the charge is neglected [3, 10]).

ACKNOWLEDGMENTS

I am deeply grateful to V.S. Kortov and A.V. Kruzhalov for encouragement and Yu.G. Lazarev for assisting in the debugging of the computational programs.

REFERENCES

1. M. Inokuti, Rev. Mod. Phys. **43**, 297 (1971).
2. Y. T. Lee and J. C. Y. Chen, Phys. Rev. A **19**, 526 (1979).
3. V. I. Radchenko, Zh. Éksp. Teor. Fiz. **103**, 40 (1993) [JETP **76**, 22 (1993)].
4. L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-Relativistic Theory* (Nauka, Moscow, 1974; Pergamon, Oxford, 1977).
5. E. Clementi and C. Roetti, At. Data Nucl. Data Tables **14**, 177 (1974).
6. V. I. Radchenko, Zh. Éksp. Teor. Fiz. **105**, 834 (1994) [JETP **78**, 445 (1994)].
7. V. I. Radchenko, D. A. Kozhukhov, and V. N. Kudryavtsev, Zh. Tekh. Fiz. **70** (2), 12 (2000) [Tech. Phys. **45**, 153 (2000)].
8. V. I. Radchenko and G. D. Ved'manov, Zh. Éksp. Teor. Fiz. **107**, 3 (1995) [JETP **80**, 1 (1995)].
9. J. A. Johnstone, Nucl. Instrum. Methods Phys. Res. B **52**, 1 (1990).
10. G. D. Ved'manov, Yu. G. Lazarev, and V. I. Radchenko, Zh. Tekh. Fiz. **70** (2), 81 (2000) [Tech. Phys. **45**, 221 (2000)].
11. N. Bohr, *Penetration of Atomic Particles through Matter* (Munksgaard, Copenhagen, 1948; Inostrannaya Literatura, Moscow, 1950).
12. I. S. Dmitriev, Ya. M. Zhileĭkin, and V. S. Nikolaev, Zh. Éksp. Teor. Fiz. **49**, 500 (1965) [Sov. Phys. JETP **22**, 352 (1965)].
13. C. J. Anderson, R. J. Girnius, A. M. Howald, and L. W. Anderson, Phys. Rev. A **22**, 822 (1980).
14. K. Riesselmann, L. W. Anderson, L. Durand, and C. J. Anderson, Phys. Rev. A **43**, 5934 (1991).
15. I. S. Dmitriev and V. S. Nikolaev, Zh. Éksp. Teor. Fiz. **44**, 660 (1963) [Sov. Phys. JETP **17**, 447 (1963)].
16. D. P. Dewangan and H. R. J. Walters, J. Phys. B **11**, 3983 (1978).
17. I. S. Dmitriev, Ya. F. Teplova, and Yu. A. Faĭnberg, Zh. Éksp. Teor. Fiz. **107**, 55 (1995) [JETP **80**, 28 (1995)].
18. B. A. D'yachkov, V. I. Zinenko, and G. V. Kazantsev, Zh. Tekh. Fiz. **47**, 416 (1977) [Sov. Phys. Tech. Phys. **22**, 245 (1977)].
19. D. P. Almeida, N. V. de Castro, F. L. Freite, *et al.*, Phys. Rev. A **36**, 16 (1987).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 4, 2000, pp. 691–697.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 4, 2000, pp. 798–805.

ATOMS, SPECTRA, RADIATION

Induced Superradiation for a Collection of Three-Level Atoms

R. A. Ismailov and A. Ya. Kazakov*
State University of Aerospace Instrument Building, St. Petersburg, 190000 Russia
**e-mail: akazak@mail.ru*
Received April 4, 2000

Abstract—*N* three-level atoms interact simultaneously with classical and quantum fields, which are in quasiresonance with various atomic transitions. The classical and quantum fields exchange photons by means of the atoms. It is shown that under certain conditions this process is collective. The number of photons in a quantized mode oscillates, and the amplitude of these oscillations is proportional to N^2. The frequency of the oscillations is determined by the frequencies of the classical and external fields. *© 2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The dynamics of two- and three-level atoms under the action of quasiresonance classical fields plays an important role in many problems of quantum optics and laser spectroscopy and has now been well studied (see, for example, [1, 2]). At the same time, the dynamics of quantum systems consisting of atoms interacting simultaneously with classical and quantized fields has been investigated much less well. In [3–9] various physical aspects of the dynamics of a two-level atom interacting simultaneously with classical and quantized fields have been discussed in [3–9]. In the present paper we examine the situation where a collection of three-level atoms interacts with external fields.

Let a three-level atom interact simultaneously with the classical fields I_2 and I_3 and a quantized mode I_q, which are quasiresonant with respect to various atomic transitions (see the level and field scheme in Fig. 1). It has been shown [6–8] that under certain conditions the exchange of photons between the classical and quantized fields occurs in such physical systems. An atom can viewed as a intermediary in such an exchange process. In the present paper we shall discuss the situation where $N > 1$ three-level atoms interact with classical and quantized fields. Our objective is to show that the exchange of photons becomes a collective process. A similar problem for a collection of two-level atoms was discussed in [10]. We note that although the number of atoms is assumed to be quite large, at the same time it is assumed to be much less than the Avogadro's number. Then the back effect of the atoms on the amplitude of the classical fields can be neglected.

We assume that all atoms are identical and we do not discuss spatial effects. This means that the atoms are localized in a region whose dimensions in the corresponding directions are much smaller than the wavelengths of the fields under study. The latter condition can be achieved by various methods. For example, the atoms can be placed on a plane perpendicular to the direction of propagation of the fields (see Fig. 2).

We shall begin with a discussion of the situation where the electromagnetic fields interact with a single atom. The interaction of the external fields with the atom can be described by means of the corresponding Rabi parameters, which depend on the product of the field amplitudes by the value of the corresponding matrix element. Ordinarily, the Rabi parameter of a classical field is much greater than that for a quantized field. We shall discuss our problem precisely under this assumption. The dynamics of our quantum system can be "split" into two parts. On the one hand, the interaction with the classical fields determines the corresponding "fast" oscillations in the behavior of the quantum system. On the other hand, a "slow" evolution of the quantum system, determined by the interaction of the atom with the quantized mode, is present against this background. Using an appropriate averaging procedure (the corresponding mathematical details were discussed in [9]), we shall derive a Hamiltonian that con-

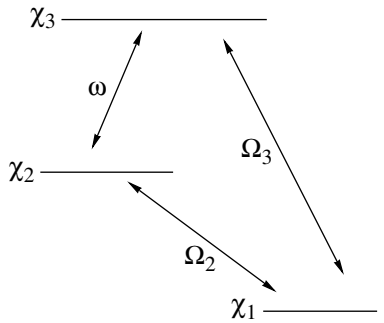

Fig. 1. Scheme of levels and fields (see explanation in the text).

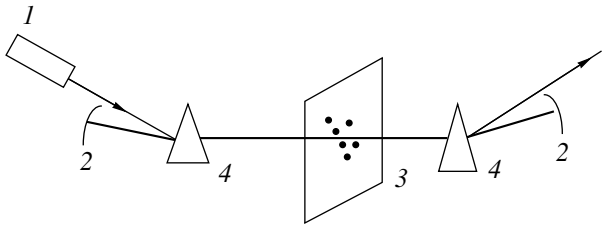

Fig. 2. Physical scheme: (*1*) source of external radiation, (*2*) resonator mirrors; (*3*) plane with pinned atoms; (*4*) optical components.

trols the "slow" dynamics of our system. Using the Bargmann–Fock formalism we shall obtain a solution for the corresponding initial problem. Then, we shall transfer these results to the case $N > 1$. Finally, we shall calculate the number of photons in the quantized mode $n(t)$ and we shall show that it oscillates with an amplitude proportional to N^2. This fact shows that the interaction of atoms with the quantized mode in our situation is a collective process. Of course, the results obtained are valid only under the assumptions made, including taking account of the above-indicated relation between the Rabi parameters of the classical and quantized fields.

We shall briefly discuss the physical realizability of the scheme shown in Fig. 1. It is well known that for free atoms interacting with external fields, in the dipole approximation, the three transitions described in this scheme cannot simultaneously be optically open. However, it can be assumed that one of the atomic transitions interacts with an external classical field in the quadrupole approximation. A different variant of this scheme of levels and fields obtains in three-level systems based on (asymmetric) quantum wells; here there are no symmetric restrictions and all three atomic transitions can be optically active in the dipole approximation. We shall not distinguish these two possibilities, we shall discuss them on the basis of a general approach.

2. ONE ATOM

The dynamics of a three-level atom interacting with classical and quantum fields in the manner shown in the scheme in Fig. 1 can be described with the Hamiltonian

$$H = \omega a^+ a + \mathrm{diag}\{\kappa_3, \kappa_2, \kappa_1\} + \zeta(a^+ J_- + a J_+) + \begin{pmatrix} 0 & 0 & \mu_3 \exp(-i\Omega_3 t) \\ 0 & 0 & \mu_2 \exp(-i\Omega_2 t) \\ \mu_3 \exp(-i\Omega_3 t) & \mu_2 \exp(-i\Omega_2 t) & 0 \end{pmatrix} \quad (1)$$

(we employ the rotating-wave approximation which is standard in quantum optics). Here

$$J_- = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad J_+ = J_-^T,$$

κ_m, $m = 1, 2, 3$ are the energies of the atomic levels, Ω_3, Ω_2, and ω are, respectively, the frequencies of the classical monochromatic fields and the quantized mode, and the parameters μ_3, μ_2, and ζ characterize the interaction of an atom with these fields. The parameters μ_3 and μ_2 are the Rabi parameters of the classical fields. The condition $\kappa_1 = 0$ can be attained by means of a shift along the energy scale. The Hamiltonian (1) operates in the space $F \otimes C^3$, where F is a Fock space and C is the complex plane. We are interested in the solution of the corresponding Schrödinger equation

$$i\frac{\partial \Psi(t)}{\partial t} = H\Psi(t).$$

Let $J_0 = \mathrm{diag}\{1, -1, 0\}$. The substitution

$$\Psi(t) = \exp\left[-i\omega t\left(a^+ a + \frac{J_0}{2}\right)\right]\Phi(t) \quad (2)$$

singles out the optical frequency, and we obtain

$$i\frac{\partial \Phi(t)}{\partial t} = \left[\zeta(a^+ J_- + a J_+) + \begin{pmatrix} \Delta_3 - \nu_3 & 0 & \mu_3 \exp[i\nu_3 t] \\ 0 & \Delta_2 - \nu_2 & \mu_2 \exp[i\nu_2 t] \\ \mu_3 \exp[-i\nu_3 t] & \mu_2 \exp[-i\nu_2 t] & 0 \end{pmatrix}\right]\Phi(t). \quad (3)$$

Here and below

$$\nu_3 = \omega/2 - \Omega_3, \quad \nu_2 = -\omega/2 - \Omega_2,$$
$$\Delta_3 = \kappa_3 - \Omega_3, \quad \Delta_2 = \kappa_2 - \Omega_2.$$

We shall discuss first the solution of the following initial problem:

$$i\frac{\partial \Xi(t)}{\partial t} = \begin{pmatrix} \Delta_3 - \nu_3 & 0 & \mu_3 \exp[i\nu_3 t] \\ 0 & \Delta_2 - \nu_2 & \mu_2 \exp[i\nu_2 t] \\ \mu_3 \exp[-i\nu_3 t] & \mu_2 \exp[-i\nu_2 t] & 0 \end{pmatrix}\Xi(t), \quad (4)$$
$$\Xi(0) = E,$$

where E is the unit matrix. The matrix $\Xi(t)$ describes the dynamics of a three-level atom interacting only with classical fields. Making the substitution

$$\tilde{\Xi}(t) = \exp[\mathrm{diag}\{-i\nu_3, -i\nu_2, 0\}t]\Xi(t)$$

we find

$$i\frac{\partial\tilde{\Xi}(t)}{\partial t} = \Lambda\tilde{\Xi}(t), \quad \Lambda = \begin{pmatrix} \Delta_3 & 0 & \mu_3 \\ 0 & \Delta_2 & \mu_2 \\ \mu_3 & \mu_2 & 0 \end{pmatrix}.$$

The matrix Λ has three real eigenvalues λ_1, λ_2, and λ_3 (we are not interested in their precise values). Generally speaking, they are different and are of the same order of magnitude as μ_3 and μ_2 (or, in different terms, R_2 and R_3). They are much greater than R_q, the effective Rabi parameter of the quantized mode. We now introduce the corresponding eigenvectors

$$e_m = (e_{m3}, e_{m2}, e_{m1})^T, \quad m = 1, 2, 3,$$

normalized to one, so that

$$e_{m3}^2 + e_{m2}^2 + e_{m1}^2 = 1.$$

It follows from the form of the matrix Λ at these vectors are orthogonal and form a basis in C^3. It is easy to show that in the general situation (when $\mu_3\mu_2 \neq 0$)

$$e_{m2}e_{m3} \neq 0, \quad e_{m1}e_{m2}e_{m3} \neq 0.$$

It follows from our arguments that the solution of the initial problem (4) is described by the relation

$$\Xi(t) = \exp[\mathrm{diag}\{\nu_3, \nu_2, 0\}it]U \times \exp[-\mathrm{diag}\{\lambda_3, \lambda_2, \lambda_1\}it]U^{-1}, \quad (5)$$

where

$$U = \begin{pmatrix} e_{33} & e_{23} & e_{13} \\ e_{32} & e_{22} & e_{12} \\ e_{31} & e_{21} & e_{11} \end{pmatrix}$$

is a unitary matrix, $U^{-1} = U^T$ (here we use the normalization of the vectors e_m).

We shall decouple the dynamics of the quantum system by substituting into Eq. (3)

$$\Phi(t) = \Xi(t)\varphi(t). \quad (6)$$

Then we obtain for the function $\varphi(t)$

$$i\frac{\partial\varphi(t)}{\partial t} = \zeta\Xi^{-1}(t)(a^+J_- + aJ_+)\Xi(t)\varphi(t). \quad (7)$$

The left-hand side of Eq. (7) is proportional to R_q. This means that this equation describes the "slow" evolution of our system. The relation (6) describes the wave function in the form of a product of "fast" and "slow" cofactors.

To describe the dynamics of a quantum system in the leading asymptotic order (relative to the small parameter $R_q/(R_2, R_3)$) we must average Eq. (7). Let

$$\nu = \nu_3 - \nu_2 = \omega + \Omega_2 - \Omega_3.$$

Then the operators $\Xi^{-1}(t)(a^+J_- + aJ_+)\Xi(t)$ (we omit their exact expressions which are complicated) have harmonics with frequencies $\pm[\nu \pm (\lambda_m - \lambda_g)]$, where $j, m = 1, 2, 3$. Eliminating the "fast" harmonics (see the details in [9]), we obtain the average Hamiltonian

$$H_{an} = \zeta\langle\Xi^{-1}(t)(a^+J_- + aJ_+)\Xi(t)\rangle,$$

which controls the "slow" evolution of the quantum system. There are two possible cases.

(1) $|\nu \pm (\lambda_m - \lambda_j)| \ll \mu_{2,3}$ for some $m \neq j$. In this case we arrive (to within simple transformations) at the standard Janes–Cummings model.

(2) In what follows we shall discuss a different situation, where

$$|\nu| \ll \mu_{2,3}. \quad (8)$$

Now we have

$$H_{av} = \zeta[a\exp(-i\nu t) + a^+\exp(i\nu t)]S,$$

where

$$S = U\mathrm{diag}\{e_{33}e_{32}, e_{23}e_{22}, e_{13}e_{12}\}U^{-1}, \quad \mathrm{rank}S = 3.$$

We are interested in the solution of the corresponding Schrödinger equation

$$i\frac{\partial\varphi(t)}{\partial t} = H_{av}\varphi(t).$$

The numbers $s_m = e_{m2}e_{m3}$ are the eigenvalues of the matrix S. We introduce the corresponding eigenvectors e_m, $m = 1, 2, 3$. These vectors form a basis in C^3. Then

$$\varphi(t) = \sum_{m=1}^{3}\eta_m(t)e_m,$$

and the functions $\eta_m(t)$ assume values in a Fock space. They are solutions of the equation

$$i\frac{\partial\eta_m(t)}{\partial t} = \zeta s_m[a\exp(-i\nu t) + a^+\exp(i\nu t)]\eta_m(t). \quad (9)$$

Such equations have been discussed in [8, 9]. Their solutions were constructed using the Bargmann–Fock formalism. The wave function is represented as an analytical function $f(z)$ and $a \longrightarrow d/dz$, $a^+ \longrightarrow z$ (see details, e.g., in [11]). Let

$$\varphi(0, z) = (\varphi_3(0, z), \varphi_2(0, z), \varphi_1(0, z))^T$$

be the initial state, where the functions $\varphi_m(0, z)$, $m = 1, 2, 3$, are the components of the vector $\varphi(0, z)$. Then

$$\varphi(0, z) = \sum_{m=1}^{3}Q_m(z)e_m,$$

where

$$(Q_3, Q_2, Q_1) = (\varphi_3(0, z), \varphi_2(0, z), \varphi_1(0, z))U,$$

and the solution of the corresponding initial problem for Eq. (9) is described by the relation

$$\eta_m(t, z) = \exp\left[\zeta s_m(1 - e^{i\nu t})\frac{z}{\nu} + \frac{\zeta^2 s_m^2(e^{-i\nu t} + i\nu t - 1)}{\nu^2}\right] \times Q_m\left[z - \frac{\zeta s_m(1 - e^{-i\nu t})}{\nu}\right].$$

We note that the operators $\Xi(t)$ (see the relation (5)) and $\exp[-i\omega t(a^+a + J_0/2)]$ are unitary. Using the relations (2) and (6) we arrive at the conclusion that the wave functions $\Psi(t)$ and $\varphi(t)$ are related by a unitary transformation. Next, the Fock operators a^+ and a commute with the matrix operators $\Xi(t)$ and $\exp[-i\omega t J_0/2]$. Therefore if the operator G is a polynomial in a^+a and has no matrix structure, then

$$\langle G\rangle = \langle\overline{\Psi(t)}, G\Psi(t)\rangle = \int dz d\bar{z}\exp(-z\bar{z})\overline{\varphi(z, t)}G\varphi(z, t) = \sum_{m=1}^{3}\int dz d\bar{z}\exp(-z\bar{z})\overline{\eta_m(z, t)}G\eta_m(z, t).$$

For example, let

$$\varphi_3(0, z) = \varphi_2(0, z) = 0, \quad \varphi_1(0, z) = \psi(z),$$

so that only the level one is occupied initially. Then

$$Q_3(z) = \psi(z)\sin\chi\sin\tau, \quad Q_2(z) = \psi(z)\cos\chi\sin\tau, \quad Q_1(z) = \psi(z)\cos\tau,$$

and the parameters τ and χ are determined by the relations

$$\sin\chi\sin\tau = e_{13}, \quad \cos\chi\sin\tau = e_{12}, \quad \cos\tau = e_{11}.$$

If

$$\psi(z) = z^m/\sqrt{m!},$$

i.e., the quantized mode has precisely m photons initially, we find

$$\langle n(t)\rangle = \langle a^+a\rangle = m + \zeta^2\frac{4\sin^2\nu t/2}{\nu^2}\sum_{m=1}^{3}(e_{m1}e_{m2}e_{m3}). \tag{10}$$

3. N ATOMS

We shall write out the Hamiltonian describing the interaction of N identical three-level atoms with classical fields and the quantized mode according to the scheme shown in Fig. 1. It operates in the space

$$\Im = F \otimes (\otimes_{k=1}^{N} C^3).$$

The space $\Im$ is formed by the linear combinations of the elements $f|e_{\sigma_1}, e_{\sigma_2}, \ldots, e_{\sigma_k}, \ldots, e_{\sigma_N}\rangle$, where $f \in F$, the numbers σ_k are 1, 2, or 3, and e_k are the eigenvectors, introduced above, of the matrix Λ. We interpret σ as a set of N such numbers and call the vectors $e_{\sigma k}$ "components" of the wave function. We shall use below the following convenient notations: the matrix operator $A^{(k)}$ operates as a matrix operator A on the kth component of the wave function:

$$A^{(k)}f|e_{\sigma_1}, e_{\sigma_2}, \ldots, e_{\sigma_k}, \ldots, e_{\sigma_N}\rangle = f|e_{\sigma_1}, e_{\sigma_2}, \ldots, Ae_{\sigma_k}, \ldots, e_{\sigma_N}\rangle.$$

We note that $A^{(k)}$ and $B^{(m)}$ commute, if $m \neq k$. Our Hamiltonian can be written in the form

$$H = \omega a^+a + \sum_{k=1}^{N}\left\{\begin{pmatrix}\kappa_3 & 0 & 0\\ 0 & \kappa_2 & 0\\ 0 & 0 & 0\end{pmatrix}^{(k)} + \zeta(a^+J_-^{(k)} + aJ_+^{(k)}) + \begin{pmatrix}0 & 0 & \mu_3\exp(-i\Omega_3 t)\\ 0 & 0 & \mu_2\exp(-i\Omega_2 t)\\ \mu_3\exp(i\Omega_3 t) & \mu_2\exp(i\Omega_2 t) & 0\end{pmatrix}^{(k)}\right\}.$$

The optical frequency can be singled out by the transformation

$$\Psi(t) = \exp\left[-i\omega t\left(a^+a + \sum_{k=1}^{N}J_0^{(k)}/2\right)\right]\Phi(t). \tag{11}$$

Then

$$i\frac{\partial\Phi(t)}{\partial t} = \sum_{k=1}^{N}\left[\zeta(a^+J_-^{(k)} + aJ_+^{(k)}) + \begin{pmatrix}\Delta_3 - \nu_3 & 0 & \mu_3\exp[i\nu_3 t]\\ 0 & \Delta_2 - \nu_2 & \mu_2\exp[i\nu_2 t]\\ \mu_3\exp[-i\nu_3 t] & \mu_2\exp[-i\nu_2 t] & 0\end{pmatrix}^{(k)}\right]\Phi(t).$$

The substitution

$$\Phi(t) = \prod_{k=1}^{N} \Xi^{(k)}(t)\varphi(t) \tag{12}$$

separates the "fast" and "slow" parts of the wave function. For the slow part we find

$$i\frac{\partial\varphi(t)}{\partial t} = \zeta\prod_{m=1}^{N}[\Xi^{(m)}(t)]^{-1} \times \sum_{k=1}^{N}(a^{+}J_{-}^{(k)} + aJ_{+}^{(k)})\prod_{n=1}^{N}\Xi^{(n)}(t)\varphi(t) = \zeta\sum_{k=1}^{N}[\Xi^{-1}(t)(a^{+}J_{-} + aJ_{+})\Xi(t)]^{(k)}\varphi(t).$$

Just as in the case of one atom, the latter operator possesses harmonics of frequencies $\pm[\nu \pm (\lambda_m - \lambda_j)]$, m, $j = 1$, 2, and 3, and we shall discuss the case $|\nu| \ll \mu_{2,3}$. The averaging procedure (similar to that described above) can be implemented in each term of this sum separately, and we obtain

$$i\frac{\partial\varphi(t)}{\partial t} = H_{av}\varphi(t), \tag{13}$$

$$H_{av} = \zeta[a\exp(-i\nu t) + a^{+}\exp(i\nu t)]\sum_{k=1}^{N}S^{(k)}. \tag{14}$$

An important fact follows from this relation: the average Hamiltonian (14) is the product of the Fock operator

$$\zeta[a\exp(-i\nu t) + a^{+}\exp(i\nu t)]$$

and the matrix operator

$$\sum_{k=1}^{N}S^{(k)}.$$

Therefore our problem splits into a set of one-dimensional problems when the problem was written in the basis of eigenvectors of this matrix operator. This operator is a sum of matrix operators, in which only the corresponding three-dimensional blocks are nontrivial and which commute with one another. Switching to the appropriate basis, we obtain

$$\varphi(t) = \sum\sigma\eta_{\sigma}(t)|e_{\sigma_1}e_{\sigma_2}\ldots e_{\sigma_N}\rangle.$$

Substituting the latter relation into Eq. (13), we find

$$i\frac{\partial\eta_{\sigma}(t)}{\partial t} = \zeta|\sigma|\left(\sum_{k=1}^{N}s_{\sigma_k}\right) \times [a\exp(-i\nu t) + a^{+}\exp(i\nu t)]\eta_{\sigma}(t),$$

where

$$|\sigma| = \left(\sum_{k=1}^{N}s_{\sigma_k}\right).$$

We note that

$$|\sigma| = A(\sigma)s_3 + B(\sigma)s_2 + [N - A(\sigma) - B(\sigma)]s_1,$$

where

$$A(\sigma) = \left(\sum_{k=1}^{N}\delta_{3,\sigma_k}\right), \quad B(\sigma) = \left(\sum_{k=1}^{N}\delta_{2,\sigma_k}\right).$$

The latter equation is the same as Eq. (9) to within a factor. Thus we can reduce the problem to a problem which has already been solved. If initially the wave function

$$\varphi(0, z) = \sum_{\sigma}Q_{\sigma}(z)|e_{\sigma_1}e_{\sigma_2}\ldots e_{\sigma_N}\rangle,$$

then

$$\eta_{\sigma}(t) = \exp\left[\frac{\zeta|\sigma|(1 - e^{i\nu t})z}{\nu} + \frac{\zeta^2|\sigma|^2(e^{-i\nu t} + i\nu t - 1)}{\nu^2}\right] \times Q_{\sigma}\left[z - \frac{\zeta|\sigma|(1 - e^{-i\nu t})}{\nu}\right].$$

We note that this relation includes the dependence on σ only through $|\sigma|$.

For simplicity, here we shall discuss only one special case of the initial problem. Let only the level 1 be initially occupied for each atom, and let the quantized mode contain precisely m photons. Then

$$Q_{\sigma} = z^{m}(m!)^{-1/2}[\sin\chi]^{A(\sigma)}[\cos\chi]^{B(\sigma)} \times [\sin\tau]^{A(\sigma)+B(\sigma)}[\cos\tau]^{N-A(\sigma)-B(\sigma)},$$

where τ and χ have the same meaning as previously.

We note that the operators

$$\prod_{k=1}^{N}\Xi^{(k)}(t) \text{ and } \exp\left[-i\omega t\left(a^{+}a + \sum_{k=1}^{N}J_0^{(k)}/2\right)\right] \tag{15}$$

are unitary, and in accordance with the relations (11) and (12) the wave functions $\Psi(t)$ and $\varphi(t)$ are related by unitary operators. Next, the Fock operator $a^{+}a$ commutes with the operators (15). Therefore if the operator

G is a polynomial in a^+a (and does not have a matrix structure), then

$$\langle G\rangle = \langle\overline{\Psi(t)}, G\Psi(t)\rangle = \int dzd\bar{z}\exp(-z\bar{z})\overline{\varphi(z,t)}G\varphi(z,t).$$

Finally, for such operators G we obtain

$$\langle G\rangle = \sum_{A=1}^{N}C_N^A\sum_{B=1}^{N-A}C_N^B[\sin\chi]^{2A}\times[\cos\chi]^{2B}[\sin\tau]^{2A+2B}[\cos\tau]^{2N-2A-2B}\times\int dzd\bar{z}\exp(-z\bar{z})\overline{\varphi_\sigma(z,t)}G\varphi_\sigma(z,t)|_{A(\sigma)=A,\,B(\sigma)=B}.$$

Setting $G = a^+a$, we find

$$\langle n(t)\rangle = m + \zeta^2\frac{4\sin^2(\nu t/2)}{\nu^2}\times\left[N(N-1)\left(\sum_{m=1}^{3}(e_{m1})^2e_{m2}e_{m3}\right)^2 + N\sum_{m=1}^{3}(e_{m1}e_{m2}e_{m3})^2\right]. \quad (16)$$

It was noted above that

$$\left(\sum_{m=1}^{3}(e_{m1})^2e_{m2}e_{m3}\right)^2 \neq 0,$$

if $\mu_3\mu_2 \neq 0$. We omit the complicated explicit expressions for the cofactors

$$\left(\sum_{m=1}^{3}(e_{m1})^2e_{m2}e_{m3}\right)^2, \quad \sum_{m=1}^{3}(e_{m1}e_{m2}e_{m3})^2.$$

The relation (16) is the basic result of this work. It shows that the number of photons in the quantized mode oscillates, and the amplitude of these oscillations is proportional to N^2, where N is the number of atoms. This fact demonstrates the collective nature of the interaction of atoms with the quantized field. A similar result is valid for any choice of the initial state of the atoms (assuming that all atoms are in the same initial state).

4. CONCLUSIONS

We shall now discuss the results obtained. As shown in [6–8], when a single atom interacts with classical fields and a quantized mode simultaneously, photons can be exchanged between the classical and quantum modes. In the present paper we studied the interaction of N three-level atoms with classical fields and a quantized mode. We showed that in this case the exchange of photons acquires a collective character. Specifically, the amplitude of the oscillations of the number of photons in the quantized mode is proportional to N^2. The frequency of the oscillations is

$$\nu = \omega + \Omega_2 - \Omega_3,$$

where ω, Ω_2, and Ω_3 are, respectively, the frequencies of the quantized mode and the classical fields, and we assumed that $|\nu| \ll \mu_2$, μ_3 (μ_2 and μ_3 are the Rabi parameters of the classical fields). We recall that for "ordinary" superradiation (see [12, 13]) the temporal dynamics of the photons in the quantized mode depends on the number of photons N. The phenomenon described in this paper—collective transfer of photons from classical modes into a quantized mode and vice versa—can be understood as induced superradiation. The calculations leading to the final result (16) demonstrate the interference nature of this phenomenon.

An atom interacting with quasiresonance classical fields can be viewed as an atom "dressed with a field." Thus we have discussed the interaction of a quantized mode with a set of such atoms dressed with a field. For "ordinary" superradiation the initially inverted atoms are the source of photons in the quantized mode. In our case the classical fields are such a source. We note that in experiments realizing ordinary superradiation optical pumping is often used in order to prepare the initial states of the atoms, which then generate (in a definite time interval) a superradiation pulse. The difference of our scheme is that the transfer of photons into the quantized mode occurs when the atoms interact with classical fields and under the action of such an interaction. The classical fields "force" the atoms to transfer photons into the quantized mode, which is what justifies the term "induced superradiation" for this phenomenon. We underscore that our quantum system does not possess an additional integral of motion (in the case of ordinary superradiation the number of excitations in the system is an integral of motion).

In the present paper we assumed that the atoms and their initial states are identical. It is easy to show that these assumptions can be relaxed, taking account of the corresponding corrections. The condition (8) is not limiting from the physical standpoint.

As mentioned above, similar results are also valid for a collection of two-level atoms [10]. However, the situation studied in the present paper is more promising for applications. Indeed, for three-level atoms the frequency of the quantized mode is close (in optical scales) to the difference or sum of the frequencies of classical modes and can be quite far from the values of these frequencies. Consequently, this physical system must be regarded as a new source of coherent radiation. There is no need to "prepare" specially an inverted initial state of the atoms. In this sense, here, there is a parallel with the well-known phenomenon of "amplification without population inversion" [14].

ACKNOWLEDGMENTS

We thank N.B. Delon and D.A. Firsov for valuable remarks.

REFERENCES

1. V. C. Letokhov and V. P. Chebotaev, *Nonlinear Spectroscopy of Ultrahigh Resolution* (Nauka, Moscow, 1990).
2. S. Stenholm, *Foundations of Laser Spectroscopy* (Wiley, New York, 1984; Mir, Moscow, 1987).
3. C. K. Law and J. H. Eberly, Phys. Rev. A **43**, 6337 (1991).
4. P. Alsing, D.-S. Guo, and H. J. Carmichael, Phys. Rev. A **45**, 5135 (1992).
5. I. V. Jyotsna and G. S. Agarwal, Opt. Commun. **99**, 344 (1993).
6. A. Ya. Kazakov, Phys. Lett. A **206**, 229 (1995).
7. A. Ya. Kazakov, Opt. Spektrosk. **81**, 549 (1996) [Opt. Spectrosc. **81**, 498 (1996)].
8. A. Ya. Kazakov, Quantum Semiclassic. Opt. **10**, 753 (1998).
9. A. Ya. Kazakov, Teor. Mat. Fiz. **117**, 92 (1998).
10. R. A. Ismailov and A. Ya. Kazakov, Zh. Éksp. Teor. Fiz. **116**, 858 (1999) [JETP **89**, 454 (1999)].
11. A. M. Perelomov, *Generalized Coherent States and Their Applications* (Nauka, Moscow, 1987; Springer-Verlag, New York, 1986).
12. R. H. Dicke, Phys. Rev. **93**, 99 (1954).
13. A. V. Andreev, V. I. Emel'yanov, and Yu. A. Il'inskiĭ, *Cooperative Phenomena in Optics* (Nauka, Moscow, 1988).
14. Ya. I. Khanin and O. A. Kocharovskaya, Pis'ma Zh. Éksp. Teor. Fiz. **48**, 581 (1988) [JETP Lett. **48**, 630 (1988)].

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 4, 2000, pp. 698–705.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 4, 2000, pp. 806–815.

ATOMS, SPECTRA, RADIATION

Quantum Transmittance of Barriers for Composite Particles

F. M. Pen'kov

Joint Institute for Nuclear Research, Dubna, Moscow oblast, 141980 Russia
e-mail: penkov@thsun1.jinr.dubna.su

Received April 18, 2000

Abstract—The transmission of two bound particles through a repulsive barrier is studied. A simple mechanism for the appearance of barrier resonances, which results in anomalous barrier transmittance as compared with the transmission probability for structureless objects, is demonstrated. It is shown that the probabilities for two interacting particles to tunnel from a false vacuum can be much higher than previously thought. *© 2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Quantum tunneling through a barrier is one of the most often encountered problems in various fields of physics. The transmission of a structureless particle through a barrier is used as a basis for studying the physical processes of tunneling. However, most applications involve problems of the transmission of composite objects through a barrier. It is obvious that when the spatial size of the barrier is much larger than the characteristic size of the incident complex the transmission probability of a composite system should be essentially the same as the transmission probability of structureless particles. The situation changes radically when the size of the complex is comparable to the spatial width of the barrier. In this case the barrier transmittance can be high (see, for example, [1] and the literature cited there). The simplest mechanism arises when only some particles of a complex interact with the barrier, when the transmission probability depends on a mass that is smaller than the mass of the complex.

A new mechanism for resonance transmittance of a square barrier for a pair of particles bound together by an infinite one-dimensional square potential was recently discovered [2]. The method of expansion in terms of the target functions (the eigenstates of the particle pair) was used but the physical picture of the transmittance and convergence questions were not considered. In the present paper this effect is investigated for other types of interactions on the basis of numerical solutions of the initial two-dimensional Schrödinger equation.

The physical picture of radical transmittance of a barrier lies in the possibility of the formation of a barrier resonance, since the potential energy of the system can possess a local minimum which gives a metastable state of the complex. This requires at least two particles to interact with the barrier. The mechanism leading to the formation of such a resonance state can be very easily imagined. Let one of the particles pass through the barrier and let the forces binding the pair together be sufficient to hold the particles on different sides of the barrier. Then, such a resonance state will exist as long as one of the particles has not passed through the barrier. The width of the barrier will determine the lifetime of such a resonance. As shown below, in this case the probability of tunneling through the barrier can reach 1. A simple explanation of this effect is interference suppression of the reflected wave, since the presence of a barrier resonance can be simply described by an effective interaction along the variable of the motion of the center of mass of the pair, the spatial form of the interaction being a local minimum at the center of the barrier. Consequently, the suppression of the reflected wave can be explained by the well-known interference phenomenon used for antireflection in optics—the path difference between the wave reflected from the first peak and the wave reflected from the second peak must be one-half the wavelength.

In the present paper an identical pair of particles bound together by an oscillator interaction (referred to as an oscillator below) and passing through a one-dimensional repulsive barrier is used to demonstrate the transmittance effect analytically and numerically. The choice of an oscillator interaction in the pair is made, on the one hand, because of the extreme simplicity of the system, making it possible to reduce the three-dimensional scattering of a three-dimensional pair of particles by a one-dimensional barrier to scattering by a one-dimensional oscillator by a one-dimensional barrier. On the other hand this type of interaction is used in the literature [3] on the probability of induced decay of a false vacuum in collisions of high-energy particles (see, e.g., [4, 5]). It was indicated that transitions from a false vacuum can be explained on the basis of the quantum-mechanical tunneling of a pair of particles through a barrier, but a system where only one of the particles of the oscillator interacted with the barrier

1063-7761/00/9104-0698$20.00

was investigated. Here, however, it will be shown that when two particles interact with the barrier the same resonant transmittance effect as described in [2] arises, and consequently the probability of the passage of a pair can be much higher than in previously studied systems.

In the present paper three types of potential barriers are studied. The first type is a Gaussian barrier with parameters of the type "narrow" and "high" compared with the characteristic size and energy of excitation of the oscillator. This type of interaction is convenient for explaining the effect on the basis of an adiabatic expansion of a two-dimensional equation. The second type of potential barrier studied, which has the form of a Gaussian function but with parameters of type "wide" and "high" in the sense indicated above, was chosen from [3] in order to observe the possibility for a radical increase of the probability of induced decay of a false vacuum. Finally, the third type of potential barrier—a Coulomb form—was investigated in order to call attention to the possibility of resonance transmission through a barrier in problems of fusion of heavy nuclei.

2. EQUATIONS

Let us consider the transmission of a pair of identical particles, coupled by an oscillator interaction, with masses $m_1 = m_2 = m$ and coordinates $\mathbf{r}_1$ and $\mathbf{r}_2$ through a potential barrier $V_0(x_1) + V_0(x_2)$. The Hamiltonian of this system ($\hbar = 1$)

$$-\frac{1}{4m}\Delta R - \frac{1}{m}\Delta_r + \frac{m\omega^2}{4}r^2 + V_0\left(\mathbf{R} - \frac{\mathbf{r}}{2}\right) + V_0\left(\mathbf{R} + \frac{\mathbf{r}}{2}\right),$$

written in the coordinates of the center of mass of the pair $\mathbf{R} = (\mathbf{r}_1 + \mathbf{r}_2)/2$ and the internal variable corresponding to the relative motion $\mathbf{r} = (\mathbf{r}_1 - \mathbf{r}_2)$, describes the three-dimensional motion of a three-dimensional oscillator. Since the potential barrier depends only on one variable and the oscillator interaction is additive in the projections $\mathbf{r}$, the wave function factorizes and its nontrivial part, describing the scattering process, depends on only two variables. It is convenient to represent these variables in the form

$$x = \sqrt{\frac{m\omega}{2}}(x_1 - x_2), \quad y = \sqrt{\frac{m\omega}{2}}(x_1 + x_2).$$

The Schrödinger equation in these variables has the form

$$[-\partial_x^2 - \partial_y^2 + x^2 + V(x-y) + V(x+y) - E]\Psi = 0, \quad (1)$$

where the energy E is given in units of $\omega/2$, and in what follows the potential barrier

$$V(x \pm y) = \frac{2}{\omega}V_0\left(\frac{x \pm y}{\sqrt{2m\omega}}\right)$$

will be written in a form convenient for our discussions. The equation (1) must be supplemented with boundary conditions. Let the scattering process proceed from left to right and let the initial state of the oscillator correspond to the state n. Then the boundary conditions can be written in the obvious form

$$\begin{gathered}\Psi \longrightarrow \exp(ik_n y)\varphi_n(x) \\ - \sum_{j \le N} S_{nj}\exp(-ik_j y)\varphi_j(x), \quad y \longrightarrow -\infty, \\ \Psi \longrightarrow \sum_{j \le N} R_{nj}\exp(ik_j y)\varphi_j(x), \quad y \longrightarrow +\infty, \\ \Psi \longrightarrow 0, \quad x \longrightarrow \pm\infty.\end{gathered} \quad (2)$$

The wave functions $\varphi_j(x)$ of the oscillator correspond to the Schrödinger equation

$$(-\partial_x^2 + x^2 - \varepsilon_i)\varphi_i = 0 \quad (3)$$

with energy $\varepsilon_j = 2j + 1$ ($j = 0, 1, 2, \ldots$) and momenta $k_j = \sqrt{E - \varepsilon_j}$, and N is the number of the last open channel ($E - \varepsilon_{N+1} < 0$). In what follows we shall study an oscillator consisting of bosons whose spectrum is conveniently enumerated from 1. Thus, everywhere below, $\varepsilon_j = 4j - 3$ ($j = 1, 2, \ldots$).

We shall determine the transmission probability W_{ij} and the reflection probability D_{ij} as the ratio of the transmitted or reflected flux density to the flux density of the incident particles:

$$W_{ij} = |R_{ij}|^2\frac{k_j}{k_i}, \quad D_{ij} = |S_{ij}|^2\frac{k_j}{k_i}.$$

It is obvious that

$$\sum_{j \le N} W_{ij} + D_{ij} = 1.$$

The problem so posed of determining the transmission (reflection) probabilities requires solving the two-dimensional differential equation (1), whose numerical solutions will be presented below.

However, we shall first demonstrate the physical reasons for the appearance of resonance transmittance, which are most clearly seen on the basis of the well-known adiabatic expansion approach, which has proven itself well in various applications of the three-body problem (see, for example, the review in [6]).

We introduce the basis functions Φ_i, satisfying the equation

$$\begin{gathered}[-\partial_x^2 + x^2 + V(x-y) \\ + V(x+y) - \epsilon_i(y)]\Phi_i(x; y) = 0,\end{gathered} \quad (4)$$

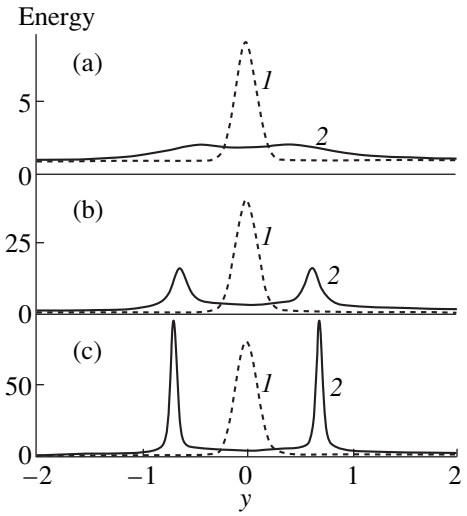

Fig. 1. Effective interaction energies of a bound pair interacting with a barrier for different amplitudes of the pair potentials V: A = (a) 1, (b) 5, and (c) 10. The curves *1* show the structureless particle approximation (the potential $2V(y)$). The curves *2* show the single-channel adiabatic approximation ($E_1(y)$). Explanations are given in the text.

and we use them for the expansion

$$\Psi(x, y) = \sum_i f_i(y)\Phi_i(x; y).$$

We substitute this expansion into Eq. (1) and project onto a basis. We obtain the system of equations

$$[(-\partial_y^2 + \epsilon_i - E)\delta_{ij} - Q_{ij}\partial_y - \partial_y Q_{ij} + P_{ij}]f_j = 0, \quad (5)$$

where the effective interaction in the channel i, $E_i = \epsilon_i + P_{ii}$, corresponds to the diagonal part of the interaction, and the functions obtained by projection, $Q_{ij} = \langle\Phi_i, \partial_y\Phi_j\rangle$ and $P_{ij} = \langle\partial_y\Phi_i, \partial_y\Phi_j\rangle$, correspond to coupling of the channels. The angular brackets denote integration over the entire domain of x. From the definition of these functions it follows that Q_{ij} is antisymmetric and P_{ii} is positive. As a rule, the coupling of the channels is weak and a limited number of equations is used to describe the scattering processes. For the discrete spectrum of Eq. (4), a good description of scattering is obtained by using all channels which are open with respect to energy [6]. In our case the spectrum of Eq. (4) is only discrete. For large $|y|$ the effective energy $E_i \longrightarrow \epsilon_i$, and $\Phi_i(x; y) \longrightarrow \varphi_i(x)$, which makes it possible to rewrite easily the boundary conditions (2) in the channel form.

The Numerov scheme (see, e.g., [7]) with step $h_x = 10^{-4}$ was used to solve Eq. (4) numerically. For large values of the variable x ($x_{\max} = 6$) the solution, starting at zero, was matched with the exponential asymptotic solution of Eq. (3). The quantities Q_{ij} and P_{ij} were constructed numerically. The numerical error in determining these quantities did not exceed 10^{-8}. The solution of the system of Eqs. (5) for determining the quantities S_{ij} and R_{ij} was based on constructing the fundamental Jost solutions (see, for example, [8]) for the left and right semiaxes with step $h_y = 10^{-4}$. The asymptotic behavior of the solutions was assumed for $y_{\max} = 10$. To analyze the discretization error and the error due to the finiteness of the region considered, the step was decreased by an order of magnitude and the size $y_{\max}$ was increased by a factor of 4. The analysis showed reliability to seven significant figures in all computational results, which are presented below, obtained on the basis of the adiabatic approximation.

A tridiagonal scheme for approximating the second derivatives with constant steps along x and y, $h_x = 0.025$ and $h_y = 0.005$, was used to solve Eq. (1) numerically. The finite sizes $|y_{\max}| = 12$ and $|x_{\max}| = 7$ of the region of the numerical calculations with the indicated degree of discretization gave an accuracy of not worse than three decimals in all computational results, presented below, based on Eq. (1).

3. RESULTS AND DISCUSSION

To demonstrate resonance transmittance we shall consider transmission through a potential barrier of the form

$$V(x) = \frac{A}{\sqrt{2\sigma\pi}}\exp\left(\frac{X^2}{2\sigma}\right), \quad X = x \pm y,$$

in the adiabatic approach. Here the amplitude A is a parameter describing the energy height of the barrier and σ determines the spatial width of the barrier.

In the approach chosen the quantum transmittance of the barrier is observed even in the one-channel approximation, i.e., in the Born–Oppenheimer approximation. Figure 1 displays the functions $E_1(y)$, obtained by solving Eq. (4) numerically with $\sigma = 0.01$ and amplitudes A = 1, 5, and 10 (a, b, c, respectively). These values of the parameters were chosen to demonstrate the appearance of a potential well, which gives the resonance features of the scattering processes. For comparison, the initial potential barriers with $x = 0$, i.e., $2V(y)$, which correspond to the scattering of structureless (or strongly bound) particles, are presented in the figure. For convenience in making comparisons, they are shifted by the binding energy of the pair.

Figure 2 shows the transmission probabilities of a pair through a barrier, which were obtained by solving Eq. (5) numerically with the corresponding to potentials shown in Fig. 1. It is obvious that for $A = 1$ the scattering of an oscillator is close to that of a structureless particle with twice the mass. For $A = 5$ a resonant component of scattering appears, and for $A = 10$ a distinct resonance is observed at the energy $E_r \approx 8.12$, which at the peak reaches the value $W_{11} = 1$. This behavior of the probability leads to the term "quantum transmittance of barriers." For comparison, we note that the probability

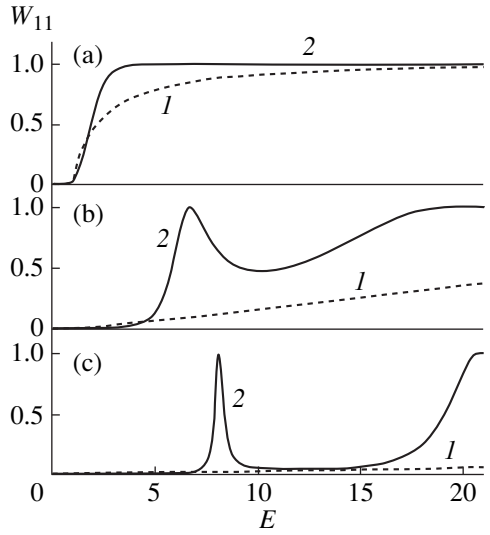

Fig. 2. Barrier transmission probabilities for a bound pair in the structureless particle approximation (curves *1*) and single-channel adiabatic approximation (curves *2*) for different amplitudes of the pair potentials V: a = (a) 1, (b) 5, and (c) 10. Explanations are given in the text.

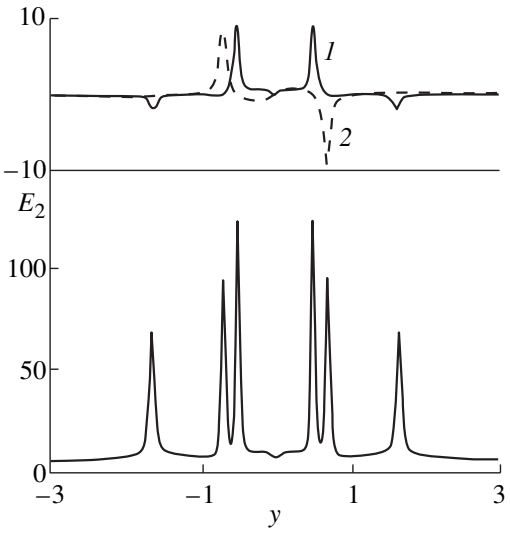

Fig. 3. Components of the second channel of the adiabatic approximation P_{12} (curve *1*) and Q_{12} (curve *2*) with the amplitude V of the pair potentials equal to 10. Explanations are given in the text.

of transmission through a barrier $2V(y)$ is approximately only 0.012.

The complete transmittance of the barrier could be somewhat surprising. Clear analogies with optical phenomena were indicated in the Introduction. In what follows we shall present simple expressions which are valid for square barriers and in the semiclassical approximation which show the possibility of complete transmittance. For discussion we chose a potential, shown in Fig. 1c, with two distinct peaks. Since the problem of transmission through a barrier in the one-dimensional case has been examined in many textbooks (see, for example, [9]), here we present only a scheme for solving the problem of transmission through a double-peak barrier. Denoting the three regions of classically allowed motion from left to right by the numbers 1, 2, and 3 and introducing superscripts for the amplitudes and probabilities of passage out of a region marked by a left-hand index into a region marked by a right-hand index, it is easy to obtain the expression

$$R^{(13)} = \frac{R^{(12)}R^{(23)}}{1 - S^{(21)}S^{(23)}}.$$

To simplify the expressions, the lower index for channel 1 is omitted here. Then the transmission probability through a double-peak barrier can be expressed in terms of the transmission probability for each peak:

$$W^{(13)} = \frac{W^{(12)}W^{(23)}}{1 + |S^{(21)}|^2|S^{(23)}|^2 - 2|S^{(21)}||S^{(23)}|\cos(\theta)}, \quad (6)$$

where θ is the twice the phase difference (or the action in the semiclassical approach) of the motion between the left- and right-hand peaks. Invariance under time reversal gives the principle of detailed balance (see, for example, [9]), which in our case leads to $|S^{(21)}| = |S^{(12)}|$.

For a symmetric potential ($W^{(12)} = W^{(23)}$) the transmission probability $W^{(13)}$ reaches a maximum for $\theta = 2\pi n$ (n = 1, 2, …). We note that this condition in the semiclassical approach determines the spectrum of bound states with infinitely wide peaks. Using the relation $|S^{(ij)}|^2 = 1 - W^{(ij)}$ it is easy to show that for these energies $W^{(13)} = 1$, i.e., complete transmittance obtains.

The parameters of the barrier potential V were chosen so that the resonance energy E_r would be higher than the energy of the second channel $\varepsilon_2 = 5$. This is necessary to demonstrate the assertion that inelastic processes will not change the resonance picture of transmittance.

Figure 3 shows the results obtained by solving Eq. (4) numerically for the second channel with $A = 10$ and $\sigma = 0.01$. It is evident that the coupling functions of the channels Q_{12} and P_{12} are approximately two orders of magnitude smaller than the diagonal values E_2. The effective energy E_2 has a more complicated form than E_1, and it can also engender additional resonances, a correct analysis of which requires taking account of a third channel (energies above 9).

Figure 4 shows the barrier transmission probabilities for a ground-state oscillator. The elastic peak W_{11} remains, though it is shifted by a relatively small amount ($E_r \approx 5.58$) and its width is substantially decreased—approximately by a factor of 3. Its maximum value ≈0.94 does not reach 1 because a second open channel is present. Here W_{12} and D_{12}, shown in the region of resonance energies in the inset, are approximately 0.03, and the complete barrier transmission

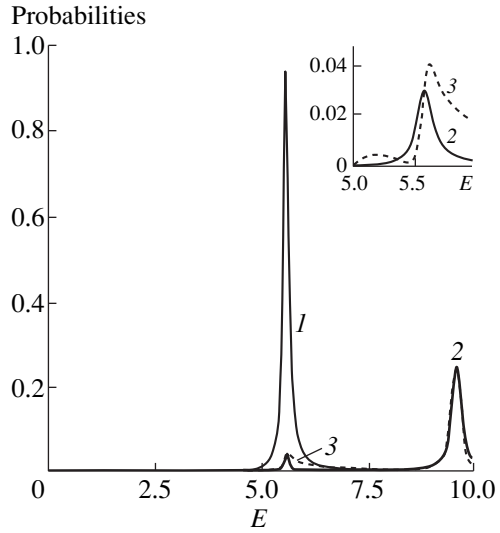

Fig. 4. Barrier transmission probabilities for a bound pair in the two-channel adiabatic approximation with amplitude V of the pair potentials equal to 10: W_{11} (*1*), W_{12} (*2*), D_{12} (*3*). Explanations are given in the text.

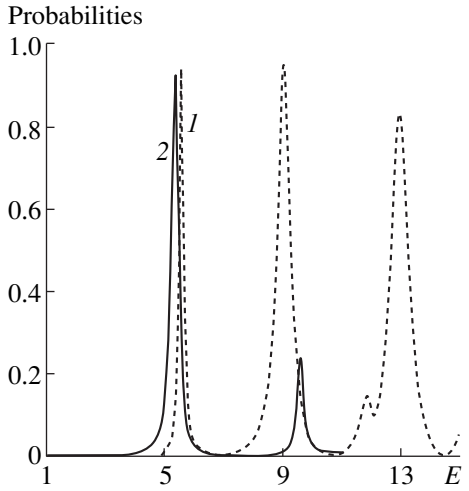

Fig. 5. Comparison of the barrier transmission probabilities of a bound pair with the exact solution of Eq. (1) (curve *1*) and the two-channel adiabatic approximation (curve *2*). Explanations are given in the text.

probability reaches about 0.97, which shows a substantial, though not 100%, transmittance. We note that the transmission probability for the barrier $2V(y)$ in this region is only about 0.0075. The quantity D_{11} (≈0.0007) is very close to 0, demonstrating the above-indicated optical effect in which the reflected wave is suppressed, even in the two-channel case.

The second peak in Fig. 4 at energy $E_r \approx 9.6$ is not reliable, since at these energies a third channel must be taken into account.

Of course, an N-channel approximation can be constructed. However, there always remains the question of the convergence of the method. Consequently, the probabilities of transmission of a pair through a barrier, found by solving the initial equation (1) numerically, will be demonstrated below. As noted above, the adiabatic approximation was required only to elucidate the transmittance mechanisms. To show how accurately the adiabatic approach describes scattering and resonance transmittance the barrier transmission probabilities of a pair in the adiabatic approximation and the results of a numerical calculation of the initial equation (1) with $A = 10$ and $\sigma = 0.01$ are presented in Fig. 5. The figure shows small differences near the first peak; this demonstrates the correctness of the two-channel adiabatic approximation in this energy range.

Thus, the adiabatic approximation qualitatively describes the dependence of the probability of transmission of a pair of particles through a barrier even if inelastic processes are possible. This shows that the explanation of resonance transmittance of barriers in such an approach is correct. To demonstrate resonance transmittance effects a type of potential barrier which can be characterized as "narrow" and "high" compared with the rms radius and excitation energy of the oscillator, respectively, was chosen. The resonance transmittance of barriers of this type fits well into the simple scheme, presented in the Introduction, explaining this effect. However, teal physical systems can possess more complicated barriers. Examples are the ordinary Coulomb barrier and the potential barrier of the form $(1/g^2)V(g^2X)$ with $g^2 \ll 1$ [3], arising in the study of the probability of induced decay of a false vacuum (see, e.g., [4, 5]). To demonstrate the resonance transmittance effects for such potential barriers the results of numerical calculations of two-dimensional equations of the type (1) with asymptotic conditions (2) are presented below.

As stated above, the model of quantum-mechanical transmission of an oscillator-coupled pair through a barrier has been used to study the induced decay of a false vacuum [3]. The case where only one of the particles interacted with the barrier was studied. On the basis of the resonance tunneling effect we can expect a large change in the tunneling picture when both incident particles are allowed to interact with the barrier. The barrier used in [3], after being put into a dimensionless form, is

$$V(x) = \frac{2}{g^2\omega}\exp\left(-\frac{g^2X^2}{\omega}\right), \quad X = x \pm y, \qquad (7)$$

where ω is the frequency of the oscillator and $g^2 \ll 1$ is a parameter in the false vacuum model. It is obvious that for constant g^2 and extremely small ω we return to the "narrow" barrier case examined above. The calculations in [3] studied the tunneling probability with $g^2 =$ 0.09, 0.06, 0.04, 0.03, 0.02, and 0.01 and fixed frequency $\omega = 1/2$. In our calculations we used the same value of ω but larger values of g^2. This is due to the

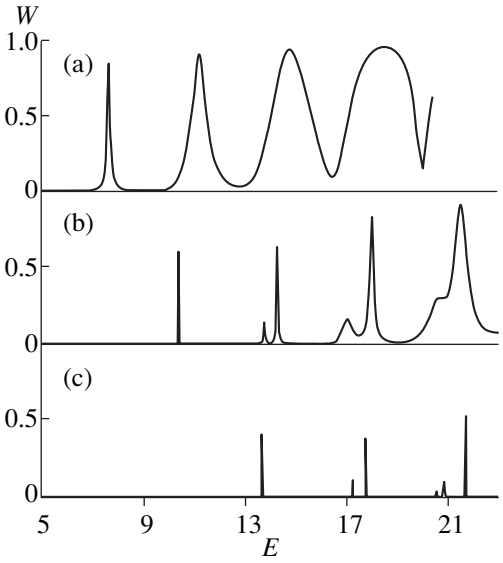

Fig. 6. Resonance behavior of the complete barrier transmission probability for a bound pair and the barrier (7). Calculation with model parameters for the false vacuum g^2 = (a) 0.5, (b) 0.3, and (c) 0.2. Explanations are given in the text.

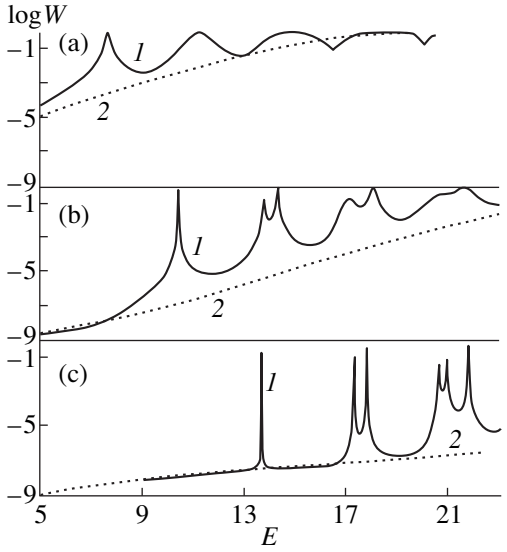

Fig. 7. Comparison of the complete barrier transmission probabilities for a bound pair. Curves *1*, exact solution for the potential (7). Curves *2*, structureless particle approximation. Calculation with parameters of the false-vacuum model g^2 = (a) 0.5, (b) 0.3, and (c) 0.2. Explanations are given in the text.

appearance of extremely (in the numerical calculation) narrow resonances in the energy dependence of the tunneling probability for a bound pair tunneling through a barrier even with $g^2 = 0.2$. Consequently, the results of numerical calculations of Eq. (1) with the barrier (7) and $g^2 = 0.5$, 0.3, and 0.2 (a, b, c, respectively) are presented in Figs. 6 and 7.

These figures also show the transmission probabilities for a bound pair from the ground state into all possible states, i.e.,

$$W = \sum_{j \le N} W_{1j},$$

for different values of g^2. The sharp resonance dependence of the tunneling probability shows the existence of the barrier resonances being discussed. We note that as g^2 decreases, the first resonance shifts in the direction of high energies, but g^2E_r decreases.

For $g^2 = 0.2$ the probability of resonance transmission of a pair is many times greater ($\sim 10^8$) then the transmission probability in the nonresonance region (background). Consequently, the computational results are presented in Fig. 7 in a logarithmic scale; these results not only demonstrate the indicated high probabilities but they also indicate that the background part of the curve coincides with the probability for the transmission of a structureless particle, i.e., the solution of Eq. (1) for a barrier potential $2V(y)$.

We used the adiabatic approximation above to explain the reasons for the resonance transmittance. However, the position of the resonances in this approach can be found after solving a system of one-dimensional differential equations. But, here, after the results of the numerical calculations of the two-dimensional equation (1) are presented, we shall show a simple scheme for the appearance of metastable barrier states, which, taking account of Eq. (6), lead to barrier transmittance. For this, we note that the potential $U(x, y) = V(x + y) + V(x - y) + x^2$ has a local minimum at $y = 0$ (the center of mass lies at the center of the barrier) and certain values $x = \pm x_0$. A maximum occurs at $x = 0$. Thus there are two potential "wells" separated by a barrier. The bound states of such a system split into even and odd states. The splitting is determined by the probability of penetration through the inner barrier. For the case $2V(x = 0, y = 0) \gg 1$ this shift can be very small and the spectrum of even states is determined simply by the spectrum of an isolated "well." To a first approximation, the position of the resonances can be described by an oscillator spectrum of bound states with $y = 0$, $x = x_0$:

$$E_{n_x n_y} = E_0 + 2\omega_x(1/2 + n_x) + 2\omega_y(1/2 + n_y). \quad (8)$$

Here n_x and n_y are oscillator quantum numbers, $E_0 = 2V(x_0)$, and the frequencies ω_x and ω_y are determined by the second derivatives at the point of the local minimum:

$$\omega_x = \sqrt{\partial_x^2 U(x, y)/2}\Big|_{x = x_0, y = 0},$$

$$\omega_y = \sqrt{\partial_y^2 U(x, y)/2}\Big|_{x = x_0, y = 0}.$$

Comparison of the positions of the first resonance with the variational estimate

g^2	E_{var}	E_r	$g^2E_r\omega/2$
0.5	7.649	7.62	1.30
0.3	10.416	10.38	0.779
0.2	13.680	13.65	0.683

It is easy to obtain the parameters of the oscillator model for the potential (7):

$$E_0 = \omega[1 + 2\ln(2/\omega)]/g^2,$$
$$\omega_x^2 = 4\ln(2/\omega), \quad (9)$$
$$\omega_y^2 = \omega_x^2 - 1.$$

We note that the expressions for the frequencies ω_x and ω_y do not depend on the parameter g^2 and impose restrictions on the parameters of the oscillator. In order for $U(x, y)$ to have a local minimum ω must be less than 1.557.

We shall now compare the positions of the resonances shown in Fig. 6 with $g^2 = 0.2$ and the computational results obtained using Eqs. (8) and (9), which make it possible to understand the existence of small but noticeable satellite resonances. The position of the first resonance $E_r = 13.65$ is described well by the energy of an oscillator in the ground state: $E_{00} = 13.92$. The second group of resonances corresponds to a single excitation of an oscillator along either x or y: $E_{01} = 18.18$, $E_{10} = 18.63$. They correspond to resonances with energies 17.24 and 17.74. The third group of resonances is engendered by a double excitation, $E_{02} = 22.45$, $E_{11} = 22.90$, and $E_{20} = 23.34$, with corresponding resonant energies: 20.58, 20.88, and 21.72. Thus, the simple oscillator model of a metastable barrier state gives the correct qualitative picture for the appearance of resonances. Comparing this picture with Fig. 6 shows that the highest tunneling probabilities correspond to metastable states with minimum excitation along the center-of-mass coordinate.

The expansion near the equilibrium point does not exhaust the possibilities of the oscillator model. The agreement between the resonance energy and the energy of the metastable state of an oscillator can be improved by a simple variational procedure. For this, we shall assume that the position of the minimum x_0 and the frequencies ω_x and ω_y are unknown quantities, which are determined by the minimum of the average value of the total Hamiltonian

$$\bar{H} = \langle\phi_x\phi_y| - \partial_x^2 - \partial_y^2 + x^2 + V(x-y) + V(x+y)|\phi_y\phi_x\rangle$$

with respect to the normalized eigenfunctions of an oscillator in the ground state:

$$\phi_x = \left(\frac{\omega_x}{2\pi}\right)^{1/4}\exp\left[-\frac{\omega_x}{4}(x-x_0)^2\right],$$
$$\phi_y = \left(\frac{\omega_y}{2\pi}\right)^{1/4}\exp\left[-\frac{\omega_y}{4}y^2\right].$$

Varying $\bar{H}$ with respect to x_0, ω_x, and ω_y gives a system of three nonlinear equations, which are not presented here because they are too complicated. Two of the three equations can be solved analytically: $x_0 = x_0(\omega_x, \omega_y)$, $\omega_y = \sqrt{\omega_x^2 - 1}$. Thus, there remains to estimate numerically a function of one variable $\bar{H} = \bar{H}(\omega_y)$, whose minimum determines the variational estimate E_{var} for the first resonance. We note that the variational relation between ω_x and ω_y is identical to the relation in the case of the expansion (9). For comparison, the values of E_{var} and the positions E_r of the first resonances, shown in Fig. 6, are presented in the table. Considering the numerical accuracy indicated above, the agreement is good.

For $g^2 \ll 1$ the variational expressions simplify and make possible the expansion

$$E_{var}^{as} \sim \frac{\omega}{g^2}\left[1 + 2\ln\left(\frac{2}{\omega}\right)\right] + \sqrt{4\ln\left(\frac{2}{\omega}\right)}$$
$$+ \sqrt{4\ln\left(\frac{2}{\omega}\right) - 1} + O(g^2),$$

which is identical up to $O(g^2)$ with the energy E_{00} obtained by a simple expansion near the minimum $U(x, y)$. Thus, the estimate of the resonance spectrum using Eqs. (8) and (9) is asymptotic in the limit $g^2 \longrightarrow 0$. Specifically, we can give the position, in the limit $g^2 \longrightarrow 0$, of the first resonance in units of g^2, i.e., the quantity $g^2E\omega/2$, which is used in [3]:

$$\frac{g^2E\omega}{2} \longrightarrow \frac{\omega^2}{2}\left[1 + \ln\left(\frac{2}{\omega}\right)\right].$$

For $\omega = 1/2$ this energy tends to approximately 0.472. The trending of the resonance energy to this limit is reflected in the fourth column in the table. A smooth computed curve (Fig. 2 in [3]) of the transmission probability in the energy range from 1.2 to 2 is presented in [3], where the transmission of a pair through a barrier was investigated. It follows from the calculations presented above that when the interaction of both particles in a pair with a potential barrier is taken into account this curve becomes strongly nonmonotonic.

The barriers examined above had a Gaussian form. To complete the picture we shall present below calcula-

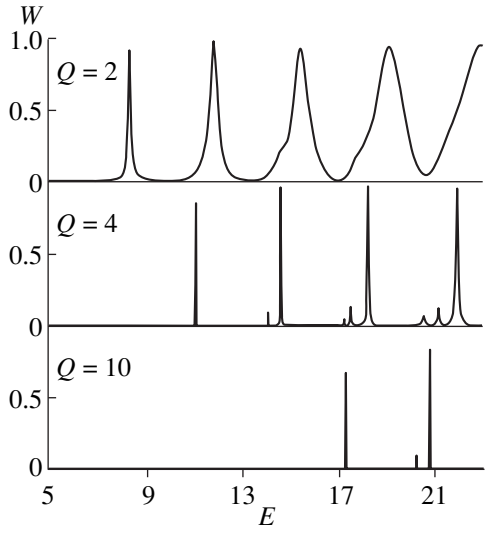

Fig. 8. Resonance behavior of the complete transmission probability of a bound pair through a Coulomb-type barrier for different values of Q. Explanations are given in the text.

tions for a Coulomb barrier cut off at small and large distances:

$$V(x) = \begin{cases} Q/X_{\min} : |X| \le X_{\min} \\ Q/|X| : X_{\min} \le |X| \le X_{\max}, \quad X = x \pm y \\ Q/X_{\max} : |X| \le X_{\max}. \end{cases} \quad (10)$$

The cutoff at small distances is introduced in order to simulate a nuclear Coulomb barrier, of course, within the limitations imposed by one-dimensional scattering. For such a cutoff the concept of "barrier height" for a one-dimensional model of scattering makes sense. For a stronger analogy, a barrier with width $|X| = X_{\min}$ should be small in the spatial units of the problem, i.e., compared with the rms size of the oscillator. The cutoff at large distances is introduced in order to use asymptotic expressions of the form (2). The quantity $X_{\max}$ must be greater than 1 in order to simulate a barrier with low transmittance. Here we chose $X_{\min} = 0.1$ and $X_{\max} = 5$. The quantity Q determines the energy height of the barrier. The computational results for $Q = 2$, 4, and 10 are presented in Fig. 8.

In this case a distinct picture of resonance tunneling of a bound pair is also observed. We shall not present here an analysis of the oscillator model for the positions of the resonances because of the strongly model nature of the potential chosen. We indicate only the obvious manifestations of the satellite resonances and the clearly noticeable equidistant spacing of the main resonances.

4. CONCLUSIONS

The barrier transmittance mechanism examined above for a bound pair of particles is clearly manifested for all potential barriers investigated. On the basis of the first observation of resonance transmittance [2] for square barriers and non-oscillator type coupling in a pair, it can be inferred that the resonance transmittance of barriers for compound particles can be observed for a very wide class of interactions. Consequently, quantum transmittance effects could occur in various fields of physics. An example is that the possibility of describing the decay of a false vacuum by quantum tunneling of a pair of particles bound by an oscillator interaction leads to resonance barrier transmittance and much higher decay probabilities for the false vacuum. The resonance transmittance of a one-dimensional truncated Coulomb barrier raises the problem of investigating the possible transmittance of a real Coulomb barrier in the physics of merging of heavy ions.

ACKNOWLEDGMENTS

I am deeply grateful to A.K. Motovilov for the idea of using projection in the numerical scheme and Yu.M. Chuvil'skiĭ for a preliminary discussion of the results obtained together with an indication of the importance of investigating the mechanism of resonance transmittance for Coulomb barriers.

REFERENCES

1. B. N. Zakhariev and A. A. Suzko, *Direct and Inverse Problems, Potentials in Quantum Scattering* (Springer-Verlag, Berlin, 1990).
2. N. Saito and Y. Kayanuma, J. Phys.: Condens. Matter **6**, 3759 (1994).
3. G. F. Bonini, A. G. Cohen, C. Rebbi, and V. A. Rubakov, Phys. Rev. D **60**, 076004 (1999); hep-ph/9901226.
4. V. A. Rubakov and M. E. Shaposhnikov, Usp. Fiz. Nauk **166**, 493 (1996) [Phys. Usp. **39**, 461 (1996)]; hep-ph/9603208.
5. A. N. Kuznetsov and P. G. Tinyakov, Phys. Rev. D **56**, 1156 (1997); hep-ph/9703256.
6. S. I. Vinitskiĭ and L. I. Ponomarev, Fiz. Élem. Chastits At. Yadra **13**, 1336 (1982) [Sov. J. Part. Nucl. **13**, 557 (1982)].
7. G. I. Marchuk, *Methods of Numerical Physics* (Nauka, Moscow, 1989).
8. J. R. Taylor, *Scattering Theory: the Quantum Theory of Nonrelativistic Colluions* (Wiley, New York, 1972; Mir, Moscow, 1975).
9. L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-Relativistic Theory* (Nauka, Moscow, 1974; Pergamon, Oxford, 1977).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 4, 2000, pp. 706–712.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 4, 2000, pp. 816–823.

ATOMS, SPECTRA, RADIATION

Double-Electron Ionization of a Quantum System in a Laser Field: Rescattering and Interparticle Correlations

E. A. Volkova, A. M. Popov*, and O. V. Tikhonova

Skobel'tsyn Institute of Nuclear Physics,
Lomonosov Moscow State University, Moscow, 119899 Russia
**e-mail: popov@mics.msu.su*

Received April 26, 2000

Abstract—Double-electron ionization of a model two-electron quantum system, describing a one-dimensional negative hydrogen ion, by a high-intensity laser pulse is investigated by direct numerical integration of the nonstationary Schrödinger equation. The possibility of interpreting the data obtained on the basis of the rescattering concept is analyzed. It is demonstrated that electron–electron correlations play an important role in the description of double-electron ionization phenomena. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

After the discovery of the production of multiply charged ions when a high-power laser field interacts with individual atoms [1], numerous experiments were performed (see, for example, [2–7]) in order to determine the mechanism for the many-fold increase in the photoelectron yield as compared with the model where a multiply charged ion is formed as a result of a sequence of independent, single-electron photoionizations. However, the mechanism leading to the formation of multiply charged ions in an intense radiation field remains unknown. The most popular model for describing multielectronic ionization is the rescattering model proposed in [8, 9]. According to this model the detachment of a second electron from an atom occurs as a result of a collision of the primary photoelectron, which undergoes almost free oscillations in the laser field, with the parent ion. One of the characteristic features of the double-electron ionization effect is the formation of a so-called "knee" in the dependence of the probability of production of doubly charged ions on the radiation intensity [2]. The possibility of describing the "knee" on the basis of a rescattering model is, apparently, at present the main argument in favor of this theory.

The rapid advance in computational techniques has opened up new possibilities in the investigation of the dynamics of atomic systems in strong light fields. The numerical integration of the nonstationary Schrödinger equation for an atom in the field of a light wave without any simplifying assumptions makes it possible not only to compare the computational data with the predictions of different analytical models but also to obtain detailed information about the flow of the process at all of its stages, which is inaccessible for modern laboratory experiments.

The capabilities of modern computers make it possible to simulate the dynamics of the ionization of a real three-dimensional helium atom [10, 11]. Such calculations require integration of a five-dimensional nonstationary Schrödinger equation, which makes it impossible to perform systematic calculations in a wide range of laser pulse parameters. Consequently, at present, for purposes of simulating the dynamics of double-electron ionization it is preferable to limit the analysis to the evolution of model two-electron one-dimensional quantum systems in a laser radiation field.

The question of the possibility of investigating the dynamics of real atomic-molecular systems on the basis of one-dimensional models has been discussed many times in the literature (see, for example, [12, 13]). Thus, qualitative agreement between the results of one- and three-dimensional calculations of the ionization dynamics of a single-electron quantum system with a short-range potential in a strong laser field was demonstrated in [13]. It was found that when calculating the threshold for the stabilization effect not only qualitative but also quantitative agreement with the computational data are observed. For two-electron systems (helium atom) qualitative agreement of the computational results for three- and one-dimensional systems follows from a comparison of the data obtained in [11] and [14–17], respectively.

The evolution of another two-electron system—the negative hydrogen ion—in a wave field is analyzed in detail in [18–21]. However, even though a detailed picture of double ionization was obtained in [16, 20] for the helium atom and the negative hydrogen ion, the mechanism of photodetachment of two electrons in a strong laser field was never completely determined. In both [16] and [20] computer experiments were performed for laser parameters corresponding to multiphoton ionization. At the same time, the appearance of rescattering, resulting in the production of doubly charged ions, is ordinarily attributed to tunneling ion-

ization. Consequently, in the present work the double ionization regime for a model negative hydrogen ion is studied for Keldysh parameters $\gamma \le 1$. It is shown that the appearance of electron rescattering by the parent center followed by electron-impact ionization cannot describe the double-electron ionization effect. It is pointed out that quantum electron–electron correlations play an important role in the process of laser action on the atom and that in consequence single-electron wave functions cannot be used to describe the dynamics of the dielectronic system in a strong light field.

2. NUMERICAL MODEL

Following [20], the Hamiltonian of the one-dimensional two-particle quantum system simulating a negative hydrogen ion was written in the form (using the atomic system of units)

$$H_0 = \sum_{i=1}^{2} (T_i + V(x_i)) + V_{12}(x_1, x_2). \quad (1)$$

Here T_i is the kinetic energy of the ith electron,

$$V(x_i) = -\frac{1}{\sqrt{\alpha^2 + x_i^2}},$$

is the energy of interaction of this electron with the nucleus,

$$V_{12} = \frac{1}{\sqrt{\alpha^2 + (x_1 - x_2)^2}}$$

is the energy of interaction of the electrons with one another, and α is a smoothing parameter. In the present work, just as in [20], it is assumed that $\alpha = 0.92$ Å. For this value of the parameter α, the system contains a single bound stationary state with ionization potential $I \sim 1.1$ eV; the binding energy of the second electron is $E_H = -11.45$ eV. This value is the ionization potential of a model one-dimensional hydrogen atom with the value used for the smoothing parameter. More detailed information about the structure of the energy spectrum and wave functions of stationary states of the system under study is contained in [20].

The interaction of the system with the field of an electromagnetic wave was studied in the dipole approximation

$$W(x_1, x_2, t) = (x_1 + x_2)\varepsilon(t)\cos(\omega t), \quad (2)$$

where ω is the frequency of the electric field of the wave and $\varepsilon(t)$ is the envelope of the pulse.

The nonstationary Schrödinger equation with the Hamiltonian

$$H = H_0(x_1, x_2) + W(x_1, x_2, t) \quad (3)$$

was integrated for values of the laser photon energy $\hbar\omega = 2$ eV ($\omega = 0.0735$) with intensities in the range $P = 10^{11}$–4×10^{14} W/cm^2. In this intensity range the Keldysh parameter

$$\gamma = \frac{\omega\sqrt{2I}}{\varepsilon}$$

varies from 10 to 0.2, i.e., for intensities in the range $P \ge 10^{13}$ W/cm^2 ionization proceeds in a tunneling or above-barrier regime.

All calculations were performed for a pulse with a smooth trapezoidal shape

$$\varepsilon(t) = \begin{cases} \varepsilon_0 \sin^2 \dfrac{\pi t}{2t_f}, & t \le t_f \\ \varepsilon_0, & t_f \le t \le t_f + t_p \\ \varepsilon_0 \sin^2 \dfrac{\pi[t - (2t_f + t_p)]}{2t_f}, & t_f + t_p \le t \le 2t_f + t_p, \end{cases} \quad (4)$$

with leading edge duration $t_f = 2T$ and "shelves" $t_p = 5T$, where $T = 2\pi/\omega$ is the period of the optical oscillations.

Following [20], the nonstationary Schrödinger equation was integrated on a grid in the coordinates

$$\xi = \frac{x_1 + x_2}{\sqrt{2}}, \quad \eta = \frac{x_1 - x_2}{\sqrt{2}}.$$

The range of integration was chosen to be a rectangle $\xi \in (-\xi_{max}, \xi_{max})$, $\eta \in (-\eta_{max}, \eta_{max})$, where $\xi_{max} = 120$ Å and $\eta_{max} = 250$ Å. Such a symmetric, in contrast to [20], spatial integration region for the nonstationary Schrödinger equation makes it possible to perform simulation in the "passive" electron approximation. In this approximation the effect of the field on one of the electrons is "switched off," and the interaction operator with the electromagnetic field is written in the form [19]

$$W = x_1\varepsilon(t)\cos(\omega t). \quad (5)$$

The second electron is "passive:" its energy varies only as a result of the interaction with the first, "active" electron. Thus the investigation of the dynamics of the dielectronic system with an interaction operator in the form (5) makes it possible to investigate the "passive" electron approximation and to determine the contribution of energy exchange between electrons in the dielectronic ionization phenomenon. Generally speaking, energy exchange between electrons in a two-electron, strongly correlated system, such as a negative hydrogen ion, cannot be reduced to classical (or semiclassical) scattering of the "active" electron by the "passive" electron. However, if the two-electron ionization is indeed the result of scattering of a photoelectron by the parent atom, then the "passive" electron model should give good agreement with the results of the solution of the Schrödinger equation with the exact Hamiltonian (3).

The probabilities of one- and two-electron ionization were calculated as follows using the wave function $\Psi(x_1, x_2, t)$ obtained by solving the nonstationary

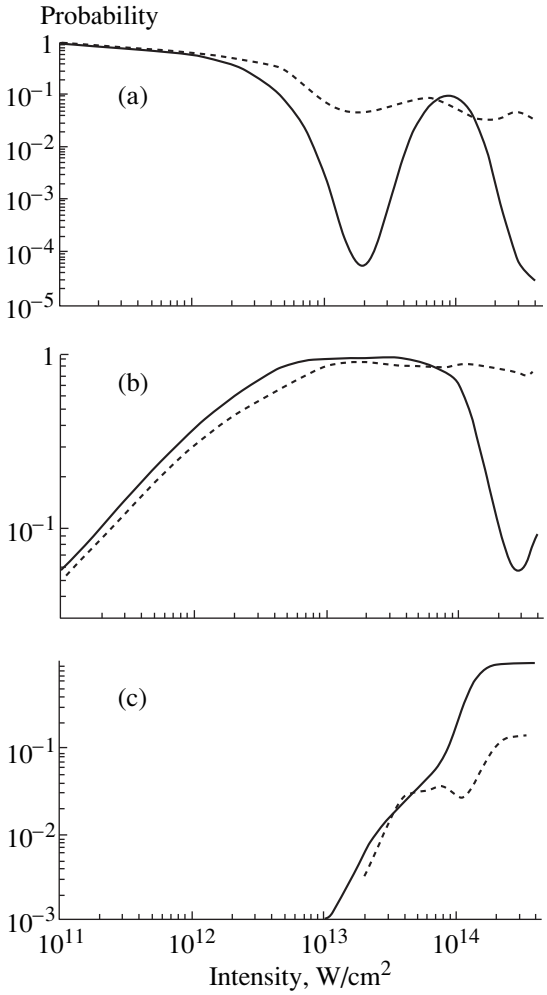

Fig. 1. Intensity dependence of the probability of finding a system in the initial unionized state (a) and probabilities of single (b) and double (c) ionization by the end of the laser pulse. Solid curves, exact calculation, broken curves, calculation in the "passive" electron approximation.

Schrödinger equation numerically. The probability of finding the system in the initial (unionized) state is

$$W_0(t) = \left|C_0(t)\right|^2 = \left|\langle \Psi(x_1, x_2, t)\left|\varphi_0(x_1, x_2)\right\rangle\right|^2, \quad (6)$$

where $\varphi_0(x_1, x_2)$ is the wave function of the stationary state of the system. The probabilities of one- and two-electron ionization were calculated using the function

$$\tilde{\Psi}(x_1, x_2, t) = \Psi(x_1, x_2, t) - C_0(t)\varphi_0(x_1, x_2)\exp\left\{-\frac{i}{\hbar}E_0 t\right\}, \quad (7)$$

where E_0 is the energy of the stationary state $\varphi_0(x_1, x_2)$.

Since the calculation of the two-electron wave functions corresponding to the stationary states of the one- and two-electron continuum is a separate, and difficult problem, the one- and two-electron ionization probabilities were calculated with respect to the wave function $\tilde{\Psi}(x_1, x_2, t^*)$ using a method similar to that of [16]. The wave functions of the one-electron continuum are different from zero if the coordinate of one of the electrons is close to zero. For states belonging to the two-electron continuum, the wave functions are different from zero in a region far from both coordinate axes. Consequently, probabilities of single- and two-electron ionization are given by

$$W_1 = \int_{-a}^{a} dx_1 \int_{-\infty}^{\infty} dx_2 \left|\tilde{\Psi}(x_1, x_2, t^*)\right|^2 + \int_{-a}^{a} dx_2 \int_{-\infty}^{\infty} dx_1 \left|\tilde{\Psi}(x_1, x_2, t^*)\right|^2 - \int_{-a}^{a} dx_1 \int_{-a}^{a} dx_2 \left|\tilde{\Psi}(x_1, x_2, t^*)\right|^2, \quad (8)$$

$$W_2 = 1 - W_0 - W_1.$$

In calculating W_1 and W_2 it was assumed that $a = 5$ Å, and the time $t^* = 11T$, which corresponds to two periods after the laser pulse ends.

3. SIMULATION RESULTS

The main result of the simulation, describing the probability of finding the system in the initial state at the end of the laser action, and the probability of one- and two-electron ionization is displayed in Figs. 1a–1c. The solid curves correspond to the exact calculation, and the broken curves correspond to the calculations performed in the "passive" electron approximation, when the electric field acts only on one "active" electron. For weak fields ($P \le 3 \times 10^{12}$ W/cm^2) the data obtained on the probability of single-electron ionization are qualitatively identical to one another, and the probability of two-electron ionization is negligibly small. It is important to note that because of the identity principle it is impossible to determine which electron leaves the system and which electron remains unionized. The electrons are equivalent, and a characteristic "cross" describing the single-electron ionization regime (Fig. 2a) arises in the spatial distribution $\rho = |\Psi(x_1, x_2)|^2$. One would think that in the "passive" electron regime only the electron on which the field acts should predominately occur, while the probability of ionization of the "passive" electron will be small. Our calculations show that this is not so. For comparatively low intensities the energy acquired by the "active" electron from the field is small, and as a result of the interelectronic interaction there is enough time for this energy to be distributed between both electrons approximately equally. As a result, the photodetachment of the "active" or "passive" electron is essentially equally proba-

ble. This is confirmed by the distribution $|\Psi(x_1, x_2)|^2$ shown in Fig. 2b and by the temporal dynamics of the expectation values of the coordinate of each electron in the quantum state. These quantities are calculated using the formula

$$\langle x_i(t)\rangle = \int x_i |\Psi(x_1, x_2, t)|^2 dx_1 dx_2, \quad i = 1, 2,$$

and are presented in Fig. 3. As one can see, the oscillations of the "active" electron in the field of the wave lead to excitation of antiphase oscillations of the "passive" electron and subsequent ionization of one of the electrons. Similar data on the magnitude of the acceleration d^2x_i/dt^2 of the "active" and "passive" electrons, obtained in [19], also show that energy exchange between electrons plays an important role in the range of fields corresponding to single-electron ionization. Under conditions of an ultrashort pulse the predominant directions of emission of the "active" and "passive" electrons are found to be opposite to one another (see also Fig. 2b). We also note that for intensities $P \le 3 \times 10^{12}$ W/cm^2 the calculations of the probability of single ionization in the "passive" electron approximation differ by not more than 30% from the exact results.

For radiation intensities above $P = 3 \times 10^{12}$ W/cm^2 the passive-electron approximation gives much too high a value for the ionization probability. For $P = 10^{13}$ W/cm^2 the exact calculation gives essentially complete ionization, while in the passive-electron model the probability of nonionization is ~5–7%. Actually, this means that for such intensities the interelectronic exchange of energy can no longer distribute the energy equally between the "active" and "passive" electrons. The computational results for the quantum-mechanical averages $\langle x_i(t)\rangle$ with $P = 2 \times 10^{13}$ W/cm^2, presented in Fig. 4, confirmed this conclusion: the dynamics of the "active" and "passive" electrons during the pulse is completely different. The amplitude of the oscillations of the "passive" electron is approximately an order of magnitude smaller than the amplitude of the oscillations of the "active" electron. We note that for exact calculations the dynamics of the quantum-mechanical averages $\langle x_i(t)\rangle$ is the same for both electrons, since the Hamiltonian (3) is symmetric under the interchange of variables $x_1 \longleftrightarrow x_2$.

Dielectronic ionization of the system was found in our calculations for intensities $P \ge 2 \times 10^{13}$ W/cm^2. The intensity dependence of the dielectronic ionization probability (see Fig. 1c) in the range $P = 4$–7×10^{13} W/cm^2 demonstrates a characteristic "knee" [2], usually attributed to the rescattering of the photoelectron by the parent atom [8, 9]. For intensities $P \ge 8 \times 10^{13}$ W/cm^2 the probability of dielectronic ionization increases rapidly and saturates at a value corresponding to $W_2 \approx 1$. Calculations of the dynamics of the system in the "passive" electron approximation also lead to the presence of two-electron ionization, which, evidently, is due to energy exchange between the electrons. Energy

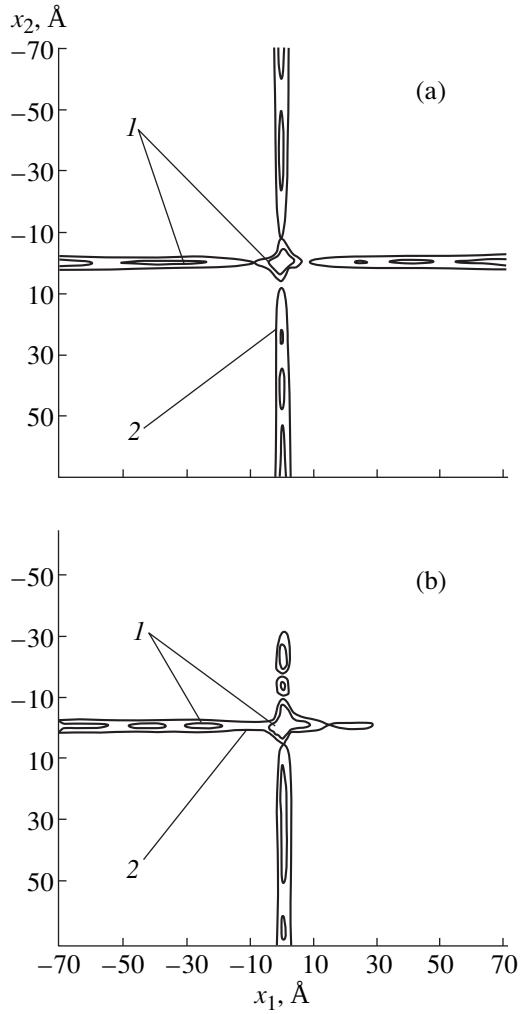

Fig. 2. Spatial structure of the wave function $|\Psi(x_1, x_2)|^2$ by the end of the laser pulse with intensity $P = 3 \times 10^{12}$ W/cm^2: (a) exact calculation, (b) "passive" electron approximation. The lines of constant probability density correspond to the values 10^{-4} (*1*) and 10^{-3} (*2*).

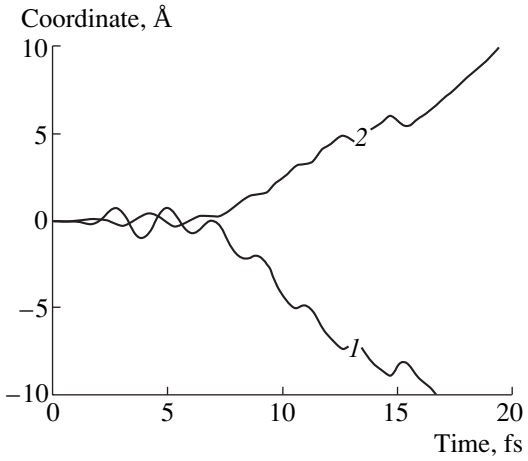

Fig. 3. Time dependence of the expectation values of the coordinates of the "active" (*1*) and "passive" (*2*) electrons in the quantum state. The radiation intensity is $P = 3 \times 10^{12}$ W/cm^2.

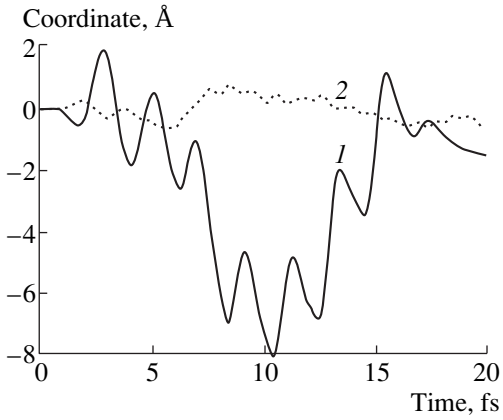

Fig. 4. Same as in Fig. 3 but for $P = 2 \times 10^{13}$ W/cm^2.

exchange between electrons can be interpreted as "rescattering," but such a definition of rescattering, generally speaking, is not the same as the classical interpretation of the phenomenon [9]. It is also important that the agreement between the exact calculations and calculations performed in the "passive" electron approximation is observed only for $P \leq 5 \times 10^{13}$ W/cm^2, corresponding to the "knee" in the dependence $W_2(P)$. For high intensities, in the "passive" electron model the probability of double ionization remains ~0.05–0.1, while the exact value ≈1. This means that in a strong field double ionization arises predominantly as a result of the direct action of the field of the wave on both electrons and is not associated with rescattering. One would think that the rescattering, as understood classically [9], could be responsible for double ionization at intensities corresponding to the "knee" in the dependence $W_2(P)$. However, this is not so. Indeed, in the tunneling ionization regime the maximum kinetic energy of a photoelectron, with which it returns to the parent atom in the half-period of the field of the electromagnetic wave, is

$$E_{\max} \simeq 3U_p = 3\varepsilon^2/4\omega^2$$

and under our conditions with $P = 3 \times 10^{13}$ W/cm^2 we have $E_{\max} \simeq 3.8$ eV, which is much less than not only the ionization potential but also the excitation potentials of all states of a one-dimensional hydrogen atom. In our view this situation makes it impossible to interpret the data obtained as rescattering of a photoelectron by the parent atom. In our calculations the ionization of the parent atom by electron impact becomes possible for $P \geq 10^{14}$ W/cm^2, i.e., when the "passive" electron model cannot describe the data obtained in exact calculations.

Thus, for the conditions of our calculations the observed double ionization of a two-electron system cannot be interpreted as photodetachment of one (outer) electron followed by electron-impact ionization of the second (inner) electron in a half-period of the electromagnetic wave.

It is important to note that any attempt to describe two-electron ionization on the basis of the concept of rescattering actually assumes the possibility of introducing single-electron wave functions describing the state of each electron in the ionization process. This assumption is justified only in the self-consistent field approximation. However, it is doubtful that this approximation is correct for describing the dynamics of ionization of multielectronic atoms in a strong electromagnetic field [14, 16, 21]. An attempt to use the nonstationary Hartree equations for describing photoionization of a model one-dimensional negative hydrogen atom [21] showed that these equations describe qualitatively correctly the dynamics of the process only in weak fields, when the detachment of only one electron is being considered. When two-electron ionization is substantial, the self-consistent field approximation does not describe the dynamics of the system even qualitatively. Consequently, the interaction between the electrons, taken into account exactly when solving the two-particle Schrödinger equation, cannot be described as the electrostatic interaction of two distributed charges.

The impossibility of describing the state of a two-electron system as a product of single-particle functions can be characterized quantitatively by means of the correlation coefficient introduced in [22, 23]

$$K(t) = [\mathrm{Sp}(\rho^2)]^{-1},$$

where

$$\rho(x_1, x_2, t) = \int \Psi^*(x_1, \xi, t)\Psi(\xi, x_2, t)d\xi$$

is the density matrix of the two-particle system.

If it is assumed that in the initial state the single-electron orbitals are equivalent, then for any moment in time

$$\Psi(x_1, x_2, t) = \Phi(x_1, t)\Phi(x_2, t)$$

and $K(t) \equiv 1$, i.e., the system is uncorrelated. Calculations of the time dependence of the correlation coefficient using the wave function obtained from the two-particle Schrödinger equation are displayed in Fig. 5 and demonstrate that the correlation coefficient is substantially different from 1, and the value of K increases in the course of the laser action all the more rapidly the higher the intensity of the radiation.

In summary, the data on the dynamics of ionization of a model negative hydrogen ion, which were obtained by comparing the exact calculations and calculations performed in the "passive" electron approximation, show that in the region of the "knee" in the dependence $W_2(P)$ ($P \sim 2$–5×10^{13} W/cm^2) double ionization occurs as a result of exchange of energy between electrons, but this energy exchange cannot be understood semiclassically as scattering of one of the electrons by the atomic core.

We also note that the degree of correlation of the electronic system is also important for describing ion-

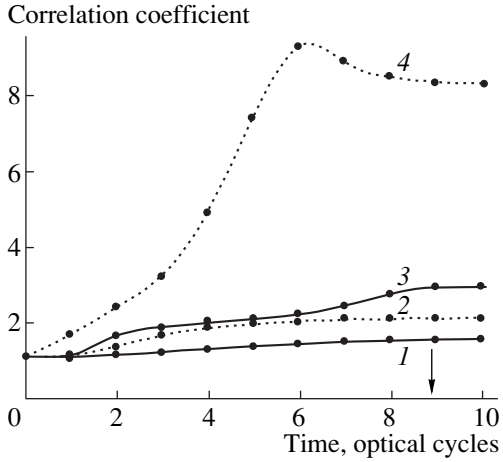

Fig. 5. Time dependence of the correlation coefficient of a two-electron system for various values of the radiation intensities (W/cm^2): (*1*) 3×10^{12}, (*2*) 10^{13}, (*3*) 10^{14}, (*4*) 2×10^{14}. The arrow marks the end of the laser pulse.

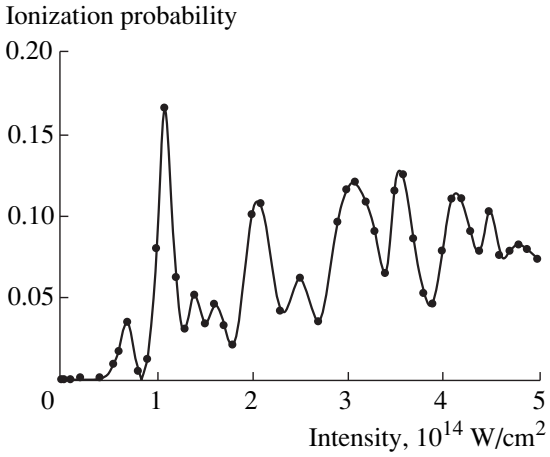

Fig. 6. Intensity dependence of the ionization probability for a model one-dimensional hydrogen atom. The parameters of the laser pulse correspond to Fig. 1.

ization in strong fields ($P \geq 10^{14}$ W/cm^2), when the probability of double ionization is close to 1. One would think that the data on the dynamics of two-electron ionization, which are presented in Fig. 1c show that the second electron leaves the atom as a result of the direct action of the field of the wave on it. Since the negative hydrogen ion is a system with a weakly bound outer electron, its double ionization should be sequential, and the dependence of the double ionization probability on the intensity will be close to the intensity dependence of the ionization probability of a hydrogen atom. However, calculations of the ionization probability of a one-electron atom with the same potential and similar pulse parameters show that this is not so (see Fig. 6). In the intensity range investigated the ionization probability as a function of the intensity has a pronounced resonance structure, probably associated with the appearance of multiphoton resonances between the ground (initial) state and one of the excited states taking account of their Stark shift in the field of the electromagnetic wave. The probability of ionization does not exceed 0.1–0.15. At the same time, the resonance structure of the dependence $W_2(P)$ is not observed for double ionization of the negative hydrogen ion for $P \geq 10^{14}$ W/cm^2, and the probability of double ionization is close to 1. This difference in the data shows that the description of the double ionization process for the system under study in terms of single-electron states does not adequately describe the picture of the process in strong fields.

4. CONCLUSIONS

In summary, our calculations of the dynamics of the ionization of the simplest two-electron system—a one-dimensional negative hydrogen atom—by the field of an electromagnetic wave show that energy exchange between electrons plays an important role for intensities corresponding to single- and double-electron ionization. At the same time, this energy exchange cannot be interpreted on the basis of a semiclassical rescattering picture consisting of photodetachment of one electron and knocking out of the second electron by the first electron in the process of scattering by the parent atom. The data obtained attest to strong interelectronic correlations in the system, which are manifested in the impossibility of describing the dynamics of the system in a strong field by means of single-electron wave functions.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (projects nos. 00-02-16046, 00-15-96554, support of leading scientific schools).

REFERENCES

1. A. L'Huiller, L. A. Lompre, G. Mainfray, and C. Manus, Phys. Rev. A **27**, 2503 (1983).
2. D. N. Fittinghoff, P. R. Bolton, B. Chang, and K. C. Kulander, Phys. Rev. Lett. **69**, 2642 (1992).
3. B. Walker, E. Mevel, Baouri Yang, *et al.*, Phys. Rev. A **48**, R894 (1993).
4. B. Walker, B. Sheehy, L. F. DiMauro, *et al.*, Phys. Rev. Lett. **73**, 1227 (1994).
5. P. Dietrich, K. H. Burnett, M. Ivanov, and P. B. Corcum, Phys. Rev. A **50**, R3585 (1994).
6. W. Sandner, H. Rottke, U. Eichmann, *et al.*, Laser Phys. **7**, 223 (1997).
7. S. Larochelle, A. Talebpour, and S. L. Chin, J. Phys. B **31**, 1201 (1998).
8. M. Yu. Kuchiev, Pis'ma Zh. Éksp. Teor. Fiz. **45**, 319 (1987) [JETP Lett. **45**, 404 (1987)].
9. P. B. Corcum, Phys. Rev. Lett. **71**, 1994 (1993).

10. J. Parker, K. T. Taylor, C. W. Clark, and S. Blodgett-Ford, J. Phys. B **29**, L33 (1996).
11. K. T. Taylor, J. S. Parker, D. Dundas, *et al.*, Laser Phys. **9**, 98 (1999).
12. E. A. Volkova and A. M. Popov, Laser Phys. **5**, 1065 (1995).
13. E. A. Volkova, A. M. Popov, and O. V. Tikhonova, Zh. Éksp. Teor. Fiz. **108**, 436 (1995) [JETP **81**, 235 (1995)].
14. M. S. Pindzola, D. C. Griffin, and C. Bottcher, Phys. Rev. Lett. **66**, 2305 (1991); M. S. Pindzola, P. Gavras, and T. W. Gorczyca, Phys. Rev. A **51**, 3999 (1995).
15. D. G. Lappas, A. Sanpera, J. B. Watson, *et al.*, J. Phys. B **29**, L619 (1996).
16. D. Bauer, Phys. Rev. A **56**, 3028 (1997).
17. W.C. Liu, J. H. Eberly, S. L. Haan, and R. Grobe, Phys. Rev. Lett. **83**, 520 (1999).
18. R. Grobe and J. H. Eberly, Phys. Rev. Lett. **68**, 2905 (1992); Phys. Rev. A **48**, 4664 (1993).
19. R. Grobe and J. H. Eberly, in *Proceedings of the 4th NATO Advanced Research Workshop on Super-Intense Laser-Atom Physics, Moscow, 1995*, Ed. by H. G. Muller and M. V. Fedorov (Kluwer, Dordrecht, 1996), p. 221.
20. E. A. Volkova, A. M. Popov, and O. V. Tikhonova, Zh. Éksp. Teor. Fiz. **114**, 1618 (1998) [JETP **87**, 875 (1998)].
21. A. M. Popov, O. V. Tikhonova, and E. A. Volkova, Laser Phys. **9**, 124 (1999).
22. R. Grobe, K. Rzazewski, and J. H. Eberly, J. Phys. B **27**, L503 (1994).
23. K. Rzazewski, in *Proceedings of the 4th NATO Advanced Research Workshop on Super-Intense Laser-Atom Physics, Moscow, 1995*, Ed. by H. G. Muller and M. V. Fedorov (Kluwer, Dordrecht, 1996), p. 213.

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 4, 2000, pp. 713–723.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 4, 2000, pp. 824–834.

ATOMS, SPECTRA, RADIATION

Interference Effects in the Scattering of Identical Atoms in a Laser Radiation Field

S. G. Rautian

Institute of Automatics and Electrometry, Siberian Division, Russian Academy of Sciences, Novosibirsk, 630090 Russia
e-mail: fractal@iae.nsk.su

Received May 22, 2000

Abstract—The properties of the collision integral in a quantum Boltzmann-type kinetic equation are studied under the conditions of spatially nonuniform distributions of colliding particles interacting with an external electromagnetic field. The components of the nonlinear resonances and the velocity distribution of the excited atoms, which are due to polarization transitions, are determined on the basis of the Kazantsev collision integral.

1. INTRODUCTION

In [1] the collision integral was found for the case of dipole–dipole interactions in the scattering of two identical atoms in a resonant radiation field. It was found that in addition to the standard terms the Kazantsev collision integral contains terms which are proportional to the product of the field-induced polarizations of the colliding atoms. These terms describe the scattering channel where the polarizations are quenched in the collision and one of the atoms is in an excited state while the other is in the ground state. The cross section for such a process is equal to the product of the scattering amplitudes with and without the exchange of excitation, i.e., it is related with the interference of these two channels. The physical manifestations of the indicated polarization transitions (the term which we shall use for them) were not studied in [1], just as they are not studied in general in the literature. One problem addressed in the present paper is to determine their role in the nonlinear spectroscopy of gases (Section 4).

The polarization transitions predicted by Kazantsev are by no means specific to the conditions assumed in [1]. In [2] it is shown that similar effects can also occur in collisions of nonidentical polarized atoms, and as a result of a collision the polarization of the atoms is quenched and a polarization is generated on a third transition. In this respect, there is an analogy with the standard scheme of nonlinear spectroscopy: the resonance interaction of a bichromatic field with neighboring transitions generates polarization on a third, forbidden transition (the Raman scattering schemes, two-photon fluorescence, and two-photon absorption; see, e.g., [3, 4]).

The situations described above raise a general question for the theory of the kinetic equation. The problem is that the polarizability of a buffer partner in a collision means that it is strongly spatially nonuniform. Indeed, the condition

$$kv_T \gg \Gamma$$

for Doppler width kv_T (k is the wave number, $v_T = \sqrt{2T/m}$ is the average thermal velocity, T is the temperature, and m is the mass of the atom) large compared with the homogeneous width Γ is equivalent to the condition

$$\lambda/2\pi \ll v_T/\Gamma \sim l_{fr}.$$

Therefore the scale of the spatial nonuniformity of the external field (λ is the wavelength), transferred to the polarized particle, is much less than the mean-free path l_{fr}. The ratio l_{fr}/λ of the scales under real conditions of nonlinear spectroscopy of gases can reach several orders of magnitude. At the same time, in the theory of the kinetic equation it is usually assumed that the thermostat is spatially uniform even in spectroscopic problems [3, 4]. The question of the role of the spatial nonuniformity of the distributions of the colliding particles is analyzed in Sections 2 and 3.

2. COLLISION INTEGRAL FOR A SPATIALLY NONUNIFORM GAS

We shall write the collision integral $\mathbf{S}$ in the following standard form (see, e.g., [3, 4])

$$\mathbf{S} = -\mathbf{S}^{(1)} + \mathbf{S}^{(2)}, \tag{2.1}$$

$$\mathbf{S}^{(1)} = \frac{i}{\hbar}\mathrm{Sp}_b\{\mathbf{T}(\boldsymbol{\rho}\times\boldsymbol{\rho}_b) - (\boldsymbol{\rho}\times\boldsymbol{\rho}_b)\mathbf{T}^+\}, \tag{2.2}$$

$$\mathbf{S}^{(2)} = \frac{1}{i\hbar}\mathrm{Sp}_b\{\mathbf{T}(\boldsymbol{\rho}\times\boldsymbol{\rho}_b)\mathbf{K}^+ - \mathbf{K}(\boldsymbol{\rho}\times\boldsymbol{\rho}_b)\mathbf{T}^+\}, \tag{2.3}$$

where $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$ are the removal and arrival terms, $\boldsymbol{\rho}$ and $\boldsymbol{\rho}_b$ are single-particle density matrices of the particle under study and the buffer particle, $\mathbf{T}$ and $\mathbf{K}$ are scattering matrices [5], and Sp_b denotes the trace with respect to all variables of the buffer particle. We shall separate the translational degrees of freedom and employ the momentum representation. Instead of the momentum variables $\mathbf{p}_a$, $\mathbf{p}'_a$ and $\mathbf{p}_b$, $\mathbf{p}'_b$ in the density matrices $\boldsymbol{\rho}(\mathbf{p}_a, \mathbf{p}'_a)$ and $\boldsymbol{\rho}_b(\mathbf{p}_b, \mathbf{p}'_b)$, it is convenient to introduce the Wigner variables

$$\begin{aligned} \mathbf{q} &= \frac{\mathbf{p}_a + \mathbf{p}'_a}{2}, \quad \boldsymbol{\tau} = \mathbf{p}_a - \mathbf{p}'_a, \\ \mathbf{q}_b &= \frac{\mathbf{p}_b + \mathbf{p}'_b}{2}, \quad \boldsymbol{\tau}_b = \mathbf{p}_b - \mathbf{p}'_b. \end{aligned} \tag{2.4}$$

We now take into account the law of conservation of the center-of-mass momentum of the colliding particles:

$$\begin{aligned} \mathbf{T}(\mathbf{p}_a\mathbf{p}_b|\mathbf{p}_{a1}\mathbf{p}_{b1}) &= \delta(\mathbf{p}_{a1} + \mathbf{p}_{b1} - \mathbf{p}_a - \mathbf{p}_b)\mathbf{T}(\mathbf{p}|\mathbf{p}_1), \\ \mathbf{p} &= \mu_b\mathbf{p}_a - \mu_a\mathbf{p}_b, \quad \mathbf{p}_1 = \mu_b\mathbf{p}_{a1} - \mu_a\mathbf{p}_{b1}, \\ \mu_a &= \frac{m_a}{m_a + m_b}, \quad \mu_b = \frac{m_b}{m_a + m_b}, \end{aligned} \tag{2.5}$$

where $\mathbf{p}$ and $\mathbf{p}_1$ are the momenta of the relative motion after and before a collision, respectively. In the notations indicated above, $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$ can be written as

$$\begin{aligned} \mathbf{S}^{(1)}(\mathbf{q}, \tau) &= \frac{i}{\hbar}\widetilde{\mathrm{Sp}}_b\int d\mathbf{q}_b d\boldsymbol{\tau}_b d\boldsymbol{\tau}_1 \delta(\boldsymbol{\tau} - \boldsymbol{\tau}_1 - \boldsymbol{\tau}_b) \\ &\times \{\mathbf{T}(\mathbf{p}|\mathbf{p} - \boldsymbol{\tau}_b)\boldsymbol{\rho}(\mathbf{q} - \boldsymbol{\tau}_b/2, \boldsymbol{\tau}_1)\rho_b(\mathbf{q}_b, \boldsymbol{\tau}_b) \\ &- \boldsymbol{\rho}(\mathbf{q} + \boldsymbol{\tau}_b/2, \boldsymbol{\tau}_1)\rho(\mathbf{q}_b, \boldsymbol{\tau}_b)\mathbf{T}^+(\mathbf{p}' + \boldsymbol{\tau}_b|\mathbf{p}')\}, \end{aligned} \tag{2.6}$$

$$\begin{aligned} \mathbf{p} &= \bar{\mathbf{p}} + \frac{\mu_b\boldsymbol{\tau} + \mu_a\boldsymbol{\tau}_b}{2}, \quad \mathbf{p}' = \bar{\mathbf{p}} - \frac{\mu_b\boldsymbol{\tau} + \mu_a\boldsymbol{\tau}_b}{2}, \\ \bar{\mathbf{p}} &= \mu_b\mathbf{q} - \mu_a\mathbf{q}_b, \end{aligned} \tag{2.7}$$

$$\begin{aligned} \mathbf{S}^{(2)}(\mathbf{q}, \boldsymbol{\tau}) &= \frac{1}{i\hbar}\widetilde{\mathrm{Sp}}_b\int d\mathbf{q}_b d\mathbf{q}_{b1} d\boldsymbol{\tau}_{b1} d\mathbf{q}_1 d\boldsymbol{\tau}_1 \\ &\times \delta(\boldsymbol{\tau} - \boldsymbol{\tau}_1 - \boldsymbol{\tau}_{b1})\delta(\mathbf{q}_1 + \mathbf{q}_{b1} - \mathbf{q} - \mathbf{q}_b) \\ &\times \{\mathbf{T}(\mathbf{p}_1|\mathbf{p}_2)\boldsymbol{\rho}(\mathbf{q}_1, \boldsymbol{\tau}_1)\boldsymbol{\rho}_b(\mathbf{q}_{b1}, \boldsymbol{\tau}_{b1})\mathbf{K}^+(\mathbf{p}'_2|\mathbf{p}'_1) \\ &- \mathbf{K}(\mathbf{p}_1|\mathbf{p}_2)\boldsymbol{\rho}(\mathbf{q}_1, \boldsymbol{\tau}_1)\boldsymbol{\rho}_b(\mathbf{q}_{b1}, \boldsymbol{\tau}_{b1})\mathbf{T}^+(\mathbf{p}'_2|\mathbf{p}'_1)\}, \end{aligned} \tag{2.8}$$

$$\begin{aligned} \mathbf{p}_1 &= \bar{\mathbf{p}} + \frac{\mu_a\boldsymbol{\tau}}{2}, \quad \mathbf{p}_2 = \bar{\mathbf{p}}_1 + \frac{\mu_b\boldsymbol{\tau}_1 - \mu_a\boldsymbol{\tau}_{b1}}{2}, \\ \bar{\mathbf{p}}_1 &= \mu_b\mathbf{q}_1 - \mu_a\mathbf{q}_{b1}, \\ \mathbf{p}'_1 &= \bar{\mathbf{p}} - \frac{\mu_b\boldsymbol{\tau}}{2}, \quad \mathbf{p}'_2 = \bar{\mathbf{p}}_1 - \frac{\mu_b\boldsymbol{\tau}_1 - \mu_a\boldsymbol{\tau}_{b1}}{2}. \end{aligned} \tag{2.9}$$

Here $\widetilde{\mathrm{Sp}}_b$ denotes the trace with respect to the internal degrees of freedom of particle b, and the matrices $\boldsymbol{\rho}(\mathbf{q}, \boldsymbol{\tau})$ and $\boldsymbol{\rho}_b(\mathbf{q}_b, \boldsymbol{\tau}_b)$ are operators with respect to these variables.

We recall that the quantity $\boldsymbol{\tau}$ is the Fourier-conjugate variable with respect to the coordinate $\mathbf{r}$ in the Wigner representation $\boldsymbol{\rho}^W(\mathbf{q}, \mathbf{r})$:

$$\begin{aligned} \boldsymbol{\rho}^W(\mathbf{q}, \mathbf{r}) &= (2\pi\hbar)^{-3}\int\boldsymbol{\rho}(\mathbf{q}, \boldsymbol{\tau})\exp(i\mathbf{r}\boldsymbol{\tau}/\hbar)d\boldsymbol{\tau}, \\ \boldsymbol{\rho}(\mathbf{q}, \boldsymbol{\tau}) &= \int\boldsymbol{\rho}^W(\mathbf{q}, \mathbf{r})\exp(-i\mathbf{r}\boldsymbol{\tau}/\hbar)d\mathbf{r}. \end{aligned} \tag{2.10}$$

In other words, the dependence on the variable $\boldsymbol{\tau}$ of the momentum representation characterizes the spatial nonuniformity in the Wigner representation. Therefore, $\delta(\boldsymbol{\tau} - \boldsymbol{\tau}_1 - \boldsymbol{\tau}_b)$ and $\delta(\boldsymbol{\tau} - \boldsymbol{\tau}_1 - \boldsymbol{\tau}_{b1})$ in the expressions (2.6) and (2.8) indicate that the spatial nonuniformity of the collision integral "consists of" the spatial nonuniformities of the density matrices of the colliding particles. It is evident from the relations (2.6)–(2.9) that the arguments of the density matrices $\mathbf{T}$ and $\mathbf{K}$ and the density matrices $\boldsymbol{\rho}$ and $\boldsymbol{\rho}_b$, generally speaking, depend on $\boldsymbol{\tau}_1$ and $\boldsymbol{\tau}_b$, i.e., on the spatial nonuniformity of both colliding particles. Since the diagonal and off-diagonal (with respect to the internal quantum numbers) components possess different spatial nonuniformities, this question will be examined later, after the collision integral is specified with respect to the internal degrees of freedom and the geometry of the external field. Here we note that ordinarily the expression used for $\mathbf{S}$ corresponds to a spatially uniform distribution of the buffer particles:

$$\boldsymbol{\rho}_b(\mathbf{q}_b, \boldsymbol{\tau}_b) = \boldsymbol{\rho}_b(\mathbf{q}_b)(2\pi\hbar)^3\delta(\boldsymbol{\tau}_b). \tag{2.11}$$

In this case we obtain from Eqs. (2.6)–(2.9)

$$\begin{aligned} \mathbf{S}^{(1)}(\mathbf{q}, \boldsymbol{\tau}) &= \frac{i}{\hbar}(2\pi\hbar)^3\mathrm{Sp}_b\int(d\mathbf{q}_b)\{\mathbf{T}(\mathbf{p}|\mathbf{p})\boldsymbol{\rho}(\mathbf{q}, \boldsymbol{\tau})\boldsymbol{\rho}_b(\mathbf{q}_b) \\ &- \boldsymbol{\rho}(\mathbf{q}, \boldsymbol{\tau})\boldsymbol{\rho}_b(\mathbf{q}_b)\mathbf{T}^+(\mathbf{p}'|\mathbf{p}')\}, \\ \mathbf{S}^{(2)}(\mathbf{q}, \boldsymbol{\tau}) &= \frac{1}{i\hbar}(2\pi\hbar)^3\widetilde{\mathrm{Sp}}_b\int d\mathbf{q}_b d\mathbf{q}_{b1} d\mathbf{q}_1 \\ &\times \delta(\mathbf{q}_1 + \mathbf{q}_{b1} - \mathbf{q} - \mathbf{q}_b) \\ &\times \{\mathbf{T}(\mathbf{p}|\mathbf{p}_1)\boldsymbol{\rho}(\mathbf{q}_1, \boldsymbol{\tau})\boldsymbol{\rho}_b(\mathbf{q}_{b1})\mathbf{K}^+(\mathbf{p}'_1|\mathbf{p}') \\ &- \mathbf{K}(\mathbf{p}|\mathbf{p}_1)\boldsymbol{\rho}(\mathbf{q}_1, \boldsymbol{\tau})\boldsymbol{\rho}_b(\mathbf{q}_{b1})\mathbf{T}^+(\mathbf{p}'_1|\mathbf{p}')\}, \end{aligned} \tag{2.12}$$

$$\begin{aligned} \mathbf{p} &= \bar{\mathbf{p}} + \frac{\mu_b\boldsymbol{\tau}}{2}, \quad \mathbf{p}' = \bar{\mathbf{p}} - \frac{\mu_b\boldsymbol{\tau}}{2}, \\ \mathbf{p}_1 &= \bar{\mathbf{p}}_1 + \frac{\mu_b\boldsymbol{\tau}}{2}, \quad \mathbf{p}'_1 = \bar{\mathbf{p}}_1 - \frac{\mu_b\boldsymbol{\tau}}{2}. \end{aligned} \tag{2.13}$$

Here, therefore, $\mathbf{S}^{(1)}$ is determined by the "forward" scattering amplitudes. The arguments of the matrices $\mathbf{T}(\mathbf{p}|\mathbf{p}_1)$ and $\mathbf{K}(\mathbf{p}'_1|\mathbf{p}')$ in the arrival term $\mathbf{S}^{(2)}$ remain different because of the spatial nonuniformity, but the dif-

ferences of the arguments playing the determining role in small-angle scattering are the same:

$$\mathbf{p}_1' - \mathbf{p}' = \mathbf{p}_1 - \mathbf{p}.$$

According to the formulas (2.12) and (2.13), $\mathbf{S}(\mathbf{q}, \boldsymbol{\tau})$ depends on $\boldsymbol{\tau}$ through $\boldsymbol{\rho}(\mathbf{q}, \boldsymbol{\tau})$ as well as through the arguments of the scattering matrices. Therefore, even in the simple case (2.11) the collision integral in the Wigner representation depends nonlocally on the coordinates $\mathbf{r}$. With a spatially nonuniform $\boldsymbol{\rho}_b$ the nonlocal dependence will be stronger.

The spatial nonuniformity of the buffer particles could be caused by their interaction with an external electromagnetic field. From this standpoint collisions of identical particles, when the field equally influences $\boldsymbol{\rho}$ and $\boldsymbol{\rho}_b$, are of greatest interest. In this case $\mu_a = \mu_b = 1/2$. More complicated situations, where the atoms are different and the field contains several spectral components, whose frequencies are related by definite combination relations, and as a result resonance interaction with several adjoining transitions occurs, are also possible. Such situations are indicated in [2].

The periodic nonuniformity, caused, for example, by a monochromatic plane wave, is interesting for spectroscopic applications. In this case the Wigner representation has the form

$$\begin{aligned}\boldsymbol{\rho}^W(\mathbf{q}, \mathbf{r}) &= \sum_{s_a} \boldsymbol{\rho}_{s_a}(\mathbf{q}) \exp(i s_a \mathbf{k} \cdot \mathbf{r}),\\ \boldsymbol{\rho}^W(\mathbf{q}_b, \mathbf{r}) &= \sum_{s_b} \boldsymbol{\rho}_{s_b}(\mathbf{q}_b) \exp(i s_b \mathbf{k} \cdot \mathbf{r})\end{aligned} \tag{2.14}$$

(s_a and s_b are integers), and in the momentum representation

$$\begin{aligned}\boldsymbol{\rho}(\mathbf{q}, \boldsymbol{\tau}) &= (2\pi h)^3 \sum_{s_a} \rho_{s_a}(\mathbf{q}) \delta(\boldsymbol{\tau} - s_a h \mathbf{k}),\\ \boldsymbol{\rho}_b(\mathbf{q}_b, \boldsymbol{\tau}_b) &= \sum_{s_b} \rho_{s_b}(\mathbf{q}_b) \delta(\boldsymbol{\tau}_b - s_b h \mathbf{k}).\end{aligned} \tag{2.15}$$

Under such conditions we have for the collision integrals

$$\mathbf{S}(\mathbf{q}\boldsymbol{\tau}) = (2\pi h)^3 \sum_{s_a s_b} \delta[\boldsymbol{\tau} - (s_a + s_b) h \mathbf{k}] S_{s_a s_b}(\mathbf{q}), \tag{2.16}$$

$$\begin{aligned}S^{(1)}_{s_a s_b}(\mathbf{q}) = \frac{i}{h}(2\pi h)^3 \widetilde{\mathrm{Sp}}_b \sum_{s_a s_b} \int d\mathbf{q} \{ &\mathbf{T}(\mathbf{p}|\mathbf{p} - s_b h \mathbf{k})\\ &\times \boldsymbol{\rho}_{s_a}(\mathbf{q} - s_a h \mathbf{k}/2) \rho_{s_b}(\mathbf{q}_b)\\ &- \boldsymbol{\rho}_{s_a}(\mathbf{q} + s_b h \mathbf{k}/2) \rho_{s_b}(\mathbf{q}_b) \mathbf{T}(\mathbf{p}' - s_b h \mathbf{k}|\mathbf{p}') \},\end{aligned} \tag{2.17}$$

$$\begin{aligned}S^{(2)}_{s_a s_b}(\mathbf{q}) = \frac{1}{ih}(2\pi h)^2 \widetilde{\mathrm{Sp}}_b \sum_{s_a s_b} \int d\mathbf{q}_b d\mathbf{q}_{b1} d\mathbf{q}_1 &\\ \times \delta(\mathbf{q}_1 + \mathbf{q}_{b1} - \mathbf{q} - \mathbf{q}_b) &\\ \times \{ \mathbf{T}(\mathbf{p}_1|\mathbf{p}_2) \boldsymbol{\rho}_{s_a}(\mathbf{q}_1) \rho_{s_b}(\mathbf{q}_{b1}) \mathbf{K}^+(\mathbf{p}_2'|\mathbf{p}_1') &\\ - \mathbf{K}(\mathbf{p}_1|\mathbf{p}_2) \boldsymbol{\rho}_{s_a}(\mathbf{q}_1) \rho_{s_b}(\mathbf{q}_{b1}) \mathbf{T}^+(\mathbf{p}_2'|\mathbf{p}_1') \}, &\end{aligned} \tag{2.18}$$

$$\begin{aligned}\mathbf{p} &= \bar{\mathbf{p}} - \frac{(\mu_b s_a + s_b) h \mathbf{k}}{2},\\ \mathbf{p}' &= \bar{\mathbf{p}} - \frac{(\mu_b s_a + s_b) h \mathbf{k}}{2},\\ \mathbf{p}_1 &= \bar{\mathbf{p}} + \frac{\mu_a s_b h \mathbf{k}}{2}, \quad \mathbf{p}_1' = \bar{\mathbf{p}} - \frac{\mu_a s_b h \mathbf{k}}{2},\\ \mathbf{p}_2 &= \bar{\mathbf{p}}_1 + \frac{(\mu_b s_a - \mu_a s_b) h \mathbf{k}}{2},\\ \mathbf{p}_2' &= \bar{\mathbf{p}}_1 - \frac{(\mu_b s_a - \mu_a s_b) h \mathbf{k}}{2}.\end{aligned} \tag{2.19}$$

It is evident from the expressions presented that the periodic spatial nonuniformity results in an analog of the recoil effect. This is manifested in the arguments of the scattering amplitudes and the density matrices. In the outgoing term the difference of the right and left arguments of the scattering amplitudes is a multiple of the photon momentum $h\mathbf{k}$.

The possibility of spatial nonuniformity through the collision integral is determined by the same terms as in the case of a dynamic interaction with an external field [3, 4, 6], i.e., the ratio of hk to the width of the structure in the distribution $\boldsymbol{\rho}(\mathbf{q})$. Let this width be Maxwellian $\sqrt{2Tm}$. Then, in the visible region of the spectrum and at room temperature

$$\frac{hk}{\sqrt{2Tm}} \sim 10^{-5},$$

and the recoil effect is negligible. However, if $\boldsymbol{\rho}(\mathbf{q})$ possesses a sharp Bennett structure with the characteristic width $m\Gamma/k$ (Γ is the homogeneous line width), then the ratio

$$\frac{hk}{m\Gamma/k} = \frac{hk^2}{m\Gamma}$$

can be of the order of 1 for sufficiently small values

$$\frac{m\Gamma}{k} \sim 10^{-5}\sqrt{Tm}.$$

Such conditions were realized in an experimental study of the splitting of nonlinear resonances because of the recoil effect (see, for example, [7, 8]).

The effect of a spatial nonuniformity through the scattering matrix is determined by other parameters,

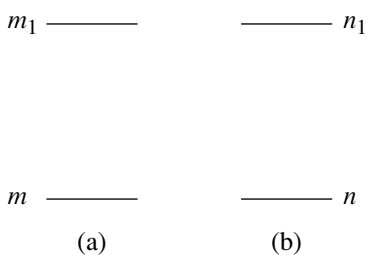

Fig. 1. Scheme of the levels of the colliding atoms (a) (m_1, m) and (b) (n_1, n).

specifically, the product of k and the effective interaction radius ρ_W (the Weisskopf radius):

$$k\rho_W = 2\pi\rho_W/\lambda.$$

For the dipole–dipole interaction ρ_W can be of the order of 10^{-6} cm, so that in the visible region of the spectrum ($\lambda = 0.5 \times 10^{-4}$ cm) we have

$$k\rho_W \approx 0.1,$$

i.e., an appreciable value. Forward scattering can be especially sensitive to this effect.

3. COLLISION INTEGRAL FOR TWO-LEVEL ATOMS IN THE FIELD OF A MONOCHROMATIC TRAVELING PLANE WAVE

To specify further the collision integral for a particular structure of the internal states of atoms, we shall employ the well-known relations between the elements of the matrices **T** and **K** and the scattering amplitudes f (see, for example, [3, 4]):

$$\begin{aligned}\mathbf{T}(\alpha\beta\mathbf{p}_a\mathbf{p}_b|\alpha_1\beta_1\mathbf{p}_{a1}\mathbf{p}_{b1}) &= -(4\pi^2\hbar\mu)^{-1}\\ \times f(\alpha\beta\mathbf{p}|\alpha_1\beta_1\mathbf{p}_1)&\delta(\mathbf{p}_{a1}+\mathbf{p}_{b1}-\mathbf{p}_a-\mathbf{p}_b)\\ \times \exp[i(E_\alpha+E_\beta-E_{\alpha1}-E_{\beta1})t/\hbar],&\end{aligned} \tag{3.1}$$

$$\begin{aligned}\mathbf{K}(\alpha\beta\mathbf{p}_a\mathbf{p}_b|\alpha_1\beta_1\mathbf{p}_{a1}\mathbf{p}_{b1})&\\ = -2\pi i\mathbf{T}(\alpha\beta\mathbf{p}_a\mathbf{p}_b|\alpha_1\beta_1\mathbf{p}_{a1}\mathbf{p}_{b1})&\\ \times\delta_+[(E_{\alpha1}+E_{\beta1}-E_\alpha-E_\beta)+(\mathbf{p}_1^2-\mathbf{p}^2)/2\mu],&\end{aligned} \tag{3.2}$$

$$\begin{aligned}\mathbf{p} &= \mu_b\mathbf{p}_a-\mu_a\mathbf{p}_b,\quad \mathbf{p}_1 = \mu_b\mathbf{p}_{a1}-\mu_a\mathbf{p}_{b1},\\ \mu &= \frac{m_am_b}{m_a+m_b}.\end{aligned} \tag{3.3}$$

Here α, α_1 and β, β_1 are sets of quantum numbers of the internal motions of the particles a and b, and E_α and E_β are the energies of the states.

Let the external electromagnetic field, in the form of a traveling electromagnetic plane wave, which depends on time and coordinates as $\exp[-i(\omega t - \mathbf{k}\cdot\mathbf{r})]$, where ω and $\mathbf{k}$ are the frequency and wave vector, produce the spatial nonuniformity of the gas. In such a field, as can be easily established, the elements of the density matrices, which are diagonal with respect to the principal quantum numbers, do not depend on $\mathbf{r}$, and the off-diagonal elements are proportional to $\exp(\pm i\mathbf{k}\cdot\mathbf{r})$. Therefore the numbers s_a and s_b in Eq. (2.15) assume the values 0 and ±1.

We shall consider resonant dipole–dipole interaction of the colliding atoms. In this case the scattering occurs only if the interacting atoms are in states with different energies, i.e., the scattering amplitudes $f(\alpha\beta\mathbf{p}|\alpha_1\beta_1\mathbf{p}_1)$ with $E_\alpha \neq E_\beta$ and $E_{\alpha1} \neq E_{\beta1}$ are different from zero [1]. At the same time, the condition $E_\alpha + E_\beta = E_{\alpha1} + E_{\beta1}$ is satisfied. We shall use the following notations: m and m_1 denote the collection of quantum numbers, respectively, of the ground and excited states of the particles a, n, and n_1, and similarly for particle b (Fig. 1). Having in mind the case of identical colliding particles, we shall drop the index b in ρ_b.

For the conditions listed above we obtain from Eq. (2.17) expressions for the off-diagonal and diagonal (with respect to the principal quantum numbers) elements of the outgoing term of the collision integral:

$$\begin{aligned}S_1^{(1)}(m_1m\mathbf{q}) &= \sum_{s_a+s_b=1} S_{s_as_b}^{(1)}(m_1m\mathbf{q})(2\pi\hbar/i\mu)\widetilde{\widetilde{\mathrm{Sp}}}_b\\ \times\int d\mathbf{q}_b\{&f(m_1n\bar{\mathbf{p}}+\hbar\mathbf{k}/4|m_1'n'\bar{\mathbf{p}}+\hbar\mathbf{k}/4)\boldsymbol{\rho}_1\\ \times(m_1'm\mathbf{q})&\boldsymbol{\rho}_0(n'n\mathbf{q}_b)-\boldsymbol{\rho}_1(m_1m'\mathbf{q})\boldsymbol{\rho}_0(n_1n_1'\mathbf{q}_b)\\ \times f^+&(m'n_1'\bar{\mathbf{p}}-\hbar\mathbf{k}/4|mn_1\bar{\mathbf{p}}-\hbar\mathbf{k}/4)\\ +f&(m_1n\bar{\mathbf{p}}+\hbar\mathbf{k}/2|m'n_1\bar{\mathbf{p}}-\hbar\mathbf{k}/2)\\ \times\boldsymbol{\rho}_0&(m'm\mathbf{q}-\hbar\mathbf{k}/2)\boldsymbol{\rho}_1(n_1n\mathbf{q}_b)\\ -\boldsymbol{\rho}_0&(m_1m_1'\mathbf{q}+\hbar\mathbf{k}/2)\boldsymbol{\rho}_1(n_1n\mathbf{q}_b)\\ \times f^+&(m_1'n\bar{\mathbf{p}}+\hbar\mathbf{k}/2|mn_1\bar{\mathbf{p}}-\hbar\mathbf{k}/2)\},\end{aligned} \tag{3.4}$$

$$\begin{aligned}S_0^{(1)}(m_1m_1'\mathbf{q}) &= \sum_{s_a+s_b=0} S_{s_as_b}^{(1)}(m_1m_1'\mathbf{q})\frac{2\pi\hbar}{i\mu}\widetilde{\widetilde{\mathrm{Sp}}}_b\\ \times\int d\mathbf{q}_b\{&f(m_1n\bar{\mathbf{p}}|m_1''n'\bar{\mathbf{p}})\rho_0(m_1''m_1'\mathbf{q})\rho_0(n'n\mathbf{q}_b)\\ -\rho_0&(m_1m_1''\mathbf{q}_b)\rho_0(n'n\mathbf{q}_b)f^+(m_1''n\bar{\mathbf{p}}|m_1'n'\bar{\mathbf{p}})\\ +f&(m_1n\bar{\mathbf{p}}+\hbar\mathbf{k}/4|mn_1\bar{\mathbf{p}}-3\hbar\mathbf{k}/4)\\ \times\boldsymbol{\rho}_{-1}&(mm_1'\mathbf{q}-\hbar\mathbf{k}/2)\boldsymbol{\rho}_1(n_1n\mathbf{q}_b)\\ -\boldsymbol{\rho}_1&(m_1m\mathbf{q}-\hbar\mathbf{k}/2)\boldsymbol{\rho}_{-1}(nn_1\mathbf{q}_b)\\ \times f^+&(mn_1\bar{\mathbf{p}}-3\hbar\mathbf{k}/4|m_1'n\bar{\mathbf{p}}-\hbar\mathbf{k}/4)\}.\end{aligned} \tag{3.5}$$

Here $\boldsymbol{\rho}_1(m_1 m\mathbf{q})$ and so on are matrices relative to the magnetic quantum numbers. The symbol $\overline{\overline{\overline{\mathrm{Sp}}}}_b$ denotes a trace with respect to these numbers of the atom *b*.

The relations (3.4) and (3.5) each contain four terms due to elastic scattering with no exchange of excitation (the first two terms in Eqs. (3.4) and (3.5)) and with exchange excitation (third and fourth terms). The zero-exchange terms contain the "forward" scattering amplitudes, and the arguments of the scattering amplitudes in the diagonal and off-diagonal terms differ by $\pm h\mathbf{k}/4$. In the "exchange" terms the difference of the left- and right-hand arguments of the scattering amplitude is equal to the photon momentum $h\mathbf{k}$, i.e., scattering by a nonzero angle is important.

The products of the zeroth harmonics of the density matrices, corresponding to the populations, and the first harmonics of the off-diagonal elements appear in $S_1^{(1)}(m_1 m\mathbf{q})$. The momentum arguments of the zeroth harmonics in the "exchange" terms (the last four rows in Eq. (3.4)) shift by $\pm h\mathbf{k}/2$, i.e., the analog of the recoil effect occurs.

The diagonal argument $S_0^{(1)}(m_1 m_1'\mathbf{q})$ contains products of the zeroth harmonics ("zero exchange" scattering) and products of the harmonics ± 1 (exchange of excitation). The latter means that in a collision of polarized particles the optical oscillations in them are quenched and a new collision channel arises for transitions of populations (polarization transitions of the particles). Thus, this result, which Kazantsev [1] obtained for the case of levels with angular momenta 0 and 1, is valid for arbitrary values of the angular momenta. It will be shown below that the polarization mechanism for transitions of the particles leads to unique effects in the Bennett structure of the velocity distribution (Section 4).

It follows from the expression (2.8) that under the resonance scattering conditions considered the arrival term for the off-diagonal element of the collision integral is strictly zero:

$$S^{(2)}(m_1 m\mathbf{q}) = 0. \tag{3.6}$$

For the diagonal element of the collision integral, however, the arrival term is given by the formula

$$\begin{aligned}
S_0^{(2)}(m_1 m_1'\mathbf{q}) &= \sum_{s_a + s_b = 0} S_{s_a a_b}^{(2)}(m_1 m_1'\mathbf{q}) = 2\mu^{-1}\overline{\overline{\overline{\mathrm{Sp}}}}_b \\
&\times \int d\mathbf{q}_b d\mathbf{q}_{b1} d\mathbf{q}_1 \delta(\mathbf{q}_1 + \mathbf{q}_{b1} - \mathbf{q} - \mathbf{q}_b) \\
&\times \{ f(m_1 n\bar{\mathbf{p}} | m_1'' n'\bar{\mathbf{p}}_1)\boldsymbol{\rho}_0(m_1'' m_1'''\mathbf{q}_1)\boldsymbol{\rho}_0(n' n''\mathbf{q}_{b1}) \\
&\times f^+(m_1''' n''\bar{\mathbf{p}}_1 | m_1' n\bar{\mathbf{p}})\delta(\bar{\mathbf{p}}_1^2 - \bar{\mathbf{p}}^2) \\
&+ f(m_1 n\bar{\mathbf{p}} | m n_1\bar{\mathbf{p}}_1)\boldsymbol{\rho}_0(m m'\mathbf{q}_1)\boldsymbol{\rho}_0(n_1 n_1'\mathbf{q}_{b1}) \\
&+ f^+(m' n_1'\bar{\mathbf{p}}_1 | m_1' n\bar{\mathbf{p}})\delta(\bar{\mathbf{p}}_1^2 - \bar{\mathbf{p}}^2) \\
&+ f(m_1 n\bar{\mathbf{p}} | m_1'' n'\bar{\mathbf{p}}_1 + h\mathbf{k}/2)\boldsymbol{\rho}_1(m_1'' m\mathbf{q}_1)\boldsymbol{\rho}_{-1}(n' n_1\mathbf{q}_{b1}) \\
&\times f^+(m n_1\bar{\mathbf{p}}_1 - h\mathbf{k}/2 | m_1' n\bar{\mathbf{p}}) \\
&\times (\delta_-[(\bar{\mathbf{p}}_1 - h\mathbf{k})^2 - \bar{\mathbf{p}}^2] + \delta_+[(\bar{\mathbf{p}}_1 + h\mathbf{k})^2 - \bar{\mathbf{p}}^2]) \\
&+ f(m_1 n\bar{\mathbf{p}} | m n_1\bar{\mathbf{p}}_1 - h\mathbf{k}/2)\boldsymbol{\rho}_{-1}(m m_1''\mathbf{q}_1)\boldsymbol{\rho}_1(n_1 n'\mathbf{q}_{b1}) \\
&\times f^+(m_1'' n'\bar{\mathbf{p}}_1 + h\mathbf{k}/2 | m_1' n\bar{\mathbf{p}}) \\
&\times (\delta_-[(\bar{\mathbf{p}}_1 + hk/2)^2 - \bar{\mathbf{p}}^2] + \delta_+[(\bar{\mathbf{p}}_1 + hk/2)^2 - \bar{\mathbf{p}}^2]) \}.
\end{aligned} \tag{3.7}$$

Here the first two terms in the integrand give the contribution of the scattering channels without and with exchange of the interaction. They have the standard form: the products of the scattering amplitudes give the cross sections without exchange (first term) and with exchange (second term) of the excitation. The third and fourth terms are due to the interference of the indicated channels and contain the product of the scattering amplitudes without and with exchange of the excitation. These terms describe the polarization transitions. The momentum arguments of the scattering amplitudes in them have additional shifts by the amount $h\mathbf{k}/2$; this is due to the vanishing of the polarizations in a collision just as happens in the case of photon absorption or emission. In addition, the terms under discussion contain the sums of δ_- and δ_+ functions, which do not reduce to a δ function because of the recoil effect.

Thus, even the simplest spatial nonuniformity of the distribution of randomly colliding polarized particles, which is introduced by a traveling monochromatic plane wave, can result in definite phenomena which are close to the phenomena which an external field gives because of the dynamical interaction of particles with the field. Under the conditions indicated at the end of Section 2, these effects could be substantial. For a standing electromagnetic wave expansions of the type (2.14)–(2.16) for the density matrices and the collision integral will contain not two but an infinite number of harmonics. High-order harmonics will lead, respectively, to large shifts by the amounts $s_a h\mathbf{k}$ and $s_b h\mathbf{k}$.

The formulas (3.5) and (3.7) give the collision integral for excited particles. The collision integral $S(mm'\mathbf{q})$ for particles in the ground state is obtained from these expressions by interchanging the indices and the sign of $\mathbf{k}$:

$$\begin{aligned}
&m \longleftrightarrow m_1, \quad m' \longleftrightarrow m_1', \quad m'' \longleftrightarrow m_1'', \\
&n \longleftrightarrow n_1, \quad n' \longleftrightarrow n_1', \quad \mathbf{k} \longleftrightarrow -\mathbf{k}.
\end{aligned} \tag{3.8}$$

We note that the expressions (3.4), (3.5), and (3.7) implicitly contain effects which are nonlinear in the field strength: the products $\boldsymbol{\rho}_1\boldsymbol{\rho}_{-1}$ and the populations of the excited states depend on the radiation intensity.

4. POLARIZATION TRANSITIONS IN A SYSTEM OF TWO LEVELS WITH ANGULAR MOMENTA 0 AND 1

It is well known that for large Doppler broadening the resonance interaction of an atom with a monochromatic plane wave results in the formation of a sharp structure in the distribution of the atoms over the projection of the velocity on the wave vector of the field. This Bennett structure has been studied in detail experimentally and theoretically in an enormous number of works and is the main physical example in laser nonlinear spectroscopy of gases. However, the theoretical studies neglected polarization transitions, and therefore it is interesting to clarify the influence of such transitions on the Bennett structure and nonlinear resonances. This question is analyzed below for the example of dipole–dipole interaction of two-level atoms with angular momenta 0 and 1 of the ground and excited states. The collision integral for this case was found in [1], and it will be used below.

Before proceeding to the analysis, we shall make several preliminary remarks. One simplification made in [1] is that the change in the velocity of the atoms in collisions was neglected in the exchange and zero-exchange channels. The estimates show that the real changes in the direction of the velocity are 10^{-2}–10^{-3} rad. The possibility of a manifestation of such a change in the velocity depends on the ratio of the Doppler and homogeneous widths. The Kazantsev model, employed in [1], will be used below. In the absence of any changes in the velocity the outgoing and ingoing terms in the collision integrals combine into a single expression, and in [1] they are presented in such a combined form. In [1] the features of the collision integral which are associated with the spatial nonuniformity and which are analyzed in Sections 2 and 3 are also neglected in [1]. However, the recoil effect is more subtle than a change in the velocity, and neglecting the latter makes it natural to neglect the former.

We note two formal circumstances. Instead of using the Wigner momentum, Kazantsev works with the velocity $\mathbf{v} \equiv \mathbf{q}/m$. The specific values of the angular momenta (0 and 1) permit making certain simplifications. Specifically, in describing the angular variables of the optical electron Kazantsev uses not the spherical functions $Y_{JM}(\theta\varphi)$ but rather their linear combinations $Y_{JM} \pm Y^*_{JM}$ [9]. Finally (for $J = 1$), x_j/r are used instead of the spherical angles θ and φ, and the indices j of the Cartesian coordinates x_j are used instead of the magnetic numbers.

In such a Cartesian representation the density matrix of the excited state is a tensor of rank 2, $\rho_{ij}(\mathbf{v})$, which is expanded in [1] in terms of irreducible tensors:

$$\rho_{ij}(\mathbf{v}) = \frac{1}{3}\delta_{ij}\eta(\mathbf{v}) + i\varepsilon_{ijl}\eta_l(\mathbf{v}) + \eta_{ij}(\mathbf{v}), \qquad \eta_{ij} = \eta_{ji}, \quad \eta_{ii} = 0, \tag{4.1}$$

where $\eta(\mathbf{v})$, $\eta_l(\mathbf{v})$, and $\eta_{ij}(\mathbf{v})$ are real, and ε_{ijl} is the completely antisymmetric unit tensor. The nondiagonal element $\rho(m_1 m\mathbf{v})$ in the Cartesian representation is a vector whose components we shall denote by $\rho_i(\mathbf{v})$. Finally, the element of the ground-state density matrix ($J = 0$) is a scalar, denoted below as $\rho(\mathbf{v})$.

The Kazantsev collision integral has a very complicated form, due in part to the fact that the anisotropy of the collisions (wind effect) is taken into account. According to a small number of works, where calculations are made for specific potentials [1, 10, 11], the spectroscopic manifestations of the wind effect are comparatively small; the line contour changes by not more than several percent. Consequently, to distinguish the role of the polarization transitions and to determine their effect on the Bennett structure and nonlinear resonances in a simple and clear form, the wind effect is neglected in what follows.

The kinetic equations for $\rho_i(\mathbf{v})$ and the irreducible tensors (4.1) with the Kazantsev collision integrals [1] for an interaction with a traveling monochromatic plane wave have the form

$$[\Gamma - i(\Omega - \mathbf{kv})]\rho_i = i\tilde{\nu}(\rho - \eta)\langle\rho_i\rangle + i\bar{\nu}(\langle\rho_i\rangle\rho - \rho_{ij}\langle\rho_j\rangle) + i(G_i\rho - \rho_{ij}G_j), \tag{4.2}$$

$$(\gamma_0 + \nu_0\langle\rho\rangle)\eta = \tilde{\nu}_0\rho\langle\eta\rangle - \bar{\nu}_0 2\mathrm{Re}(i\rho_k\langle\rho^*_k\rangle) - 2\mathrm{Re}(i\rho_k G^*_k), \tag{4.3}$$

$$(\gamma_0 + \nu_1\langle\rho\rangle)\eta_i = \tilde{\nu}_1\rho\langle\eta_i\rangle - \bar{\nu}_1\mathrm{Re}((\varepsilon_{ilk}\rho_k\langle\rho^*_k\rangle) + \mathrm{Re}(\varepsilon_{ilk}\rho_k G^*_l),) \tag{4.4}$$

$$(\gamma_0 + \nu_2\langle\rho\rangle)\eta_{ij} = \tilde{\nu}_2\rho\langle\eta_{ij}\rangle - \bar{\nu}_2\mathrm{Re}\left[i\left(\rho_i\langle\rho^*_j\rangle + \rho_j\langle\rho_i\rangle - \frac{2}{3}\delta_{ij}\rho_k\langle\rho^*_k\rangle\right)\right] - \mathrm{Re}\left[i\left(\rho_i G^*_j + \rho_j G^*_j - \frac{2}{3}\delta_{ij}\rho_k G^*_k\right)\right], \tag{4.5}$$

$$\eta + \rho = W(\mathbf{v}), \tag{4.6}$$

$$\Omega = \omega - \omega_{m_1 m}, \quad G_i = \frac{dE_i}{2h}, \quad \Gamma = \gamma_0 + \nu. \tag{4.7}$$

Here the angular brackets denote integration over the velocity, d is the reduced dipole moment matrix element for the transition $m - m_1$, E_i is the Cartesian component of the field intensity, and γ_0 is the probability of a spontaneous decay of the excited state. In [1] the change in the velocity in a collision is neglected; this explains the simple form of Eq. (4.6) for the ground-state population ρ [$W(\mathbf{v})$ is a Maxwellian distribution]. The set of collision frequencies characterizes the relaxation of the population (η), orientation (η_l), and alignment (η_{ij}) of the excited state as well as the polarization (ρ_i).

Kazantsev found the values of these frequencies numerically. They are all proportional to the parameter

$$\gamma = \gamma_0(\lambda/2\pi)^3 N \tag{4.8}$$

(N is the density of atoms) and are given by the following formulas:

$$\begin{gathered}\nu = 5.674\gamma, \quad \tilde{\nu} = -1.168\gamma, \quad \bar{\nu} = -0.72\gamma,\\ \nu_0 = 5.30\gamma, \quad \tilde{\nu}_0 = \nu_0, \quad \bar{\nu}_0 = 0.783\gamma,\\ \nu_1 = 7.147\gamma, \quad \tilde{\nu}_1 = -1.473\gamma, \quad \bar{\nu}_1 = -0.944\gamma,\\ \nu_2 = 7.623\gamma, \quad \tilde{\nu}_2 = 0.273\gamma, \quad \bar{\nu}_2 = 0.095\gamma.\end{gathered} \tag{4.9}$$

Besides the equality $\tilde{\nu}_0 = \nu_0$, the relations

$$\nu = \nu_1 + \tilde{\nu}_1, \quad 2\bar{\nu}_1 = \bar{\nu} + \tilde{\nu}, \tag{4.10}$$

also hold strictly, so that only 9 of the 12 parameters in Eqs. (4.9) are independent. The dependence of ν_κ, $\tilde{\nu}_\kappa$, and $\bar{\nu}_\kappa$ on κ ($\kappa = 0, 1, 2$) means that the model of nondegenerate states in this case will be a poor approximation. The terms on the right-hand sides of Eqs. (4.2)–(4.5) with the frequencies $\tilde{\nu}$ and $\tilde{\nu}_\kappa$ are due to transfer of excitation in collisions, and the terms with G_i are due to the interaction with a field. We shall be interested primarily in the terms with the frequencies $\bar{\nu}_\kappa$, describing the polarization transitions of the particles, orientation, and alignment.

The analogy between the polarization and field terms is interesting: the former terms are obtained from the latter terms by making the substitution

$$G_i \longrightarrow \bar{\nu}_\kappa \langle \rho_i \rangle. \tag{4.11}$$

Therefore, the polarization transitions are equivalent to induced transitions in an effective field. The analogy is strengthened by the fact that in weak fields $\rho_i \propto iG_i$, i.e., the effective field is proportional to the amplitude of the external field. The big difference lies in the factor i (phase shift by $\pi/2$), as a result of which, as will be shown below, the nonlinear resonances possess a nontrivial form. The indicated analogy sheds light on the existence of the recoil effect in the collision integral, which we discussed in Section 3.

We call attention to an important fact: the integrals of the polarization terms over $\mathbf{v}$, as one can see from Eqs. (4.3)–(4.5), are strictly zero. Consequently,

$$\begin{gathered}\gamma_0 \langle \eta \rangle = -2\mathrm{Re}(i\langle \rho \rangle G_k^*),\\ [\gamma_0 + (\nu_1 - \tilde{\nu}_1)\langle \rho \rangle]\langle \eta_l \rangle = \mathrm{Re}(\varepsilon_{ilk}\langle \rho_k \rangle G_l^*),\\ [\gamma_0 + (\nu_2 - \tilde{\nu}_2)\langle \rho \rangle]\langle \eta_{ij} \rangle\\ = -\mathrm{Re}\left[i\left(\langle \rho_i \rangle G_j^* + \langle \rho_j \rangle G_i^* - \frac{2}{3}\delta_{ij}\langle \rho_k \rangle G_k^* \right)\right].\end{gathered} \tag{4.12}$$

Thus, the "collisions of coherences" (or polarizations) do not influence the velocity-integrated populations $\langle \eta \rangle$, orientation $\langle \eta_l \rangle$, and alignment $\langle \eta_{ij} \rangle$; they change only the velocity distribution, creating in them sign-alternating elements of the structure on the excited level. Changes in $\langle \eta \rangle$, $\langle \eta_l \rangle$, and $\langle \eta_{ij} \rangle$ are generated only by transitions induced directly by the field. By virtue of Eq. (4.6) the same sign-alternating structure exists in the population $\rho(\mathbf{v})$ of the lower level.

The complete Kazantsev collision integral contains terms in addition to the terms written out in Eqs. (4.2)–(4.5), which are due to the wind effect (specifically, the parameter Γ in [1] is of a tensor character and depends on the velocity; the terms coupling the equations for the population and the alignment are also dropped). These terms are dropped for the reasons presented above.

We shall solve the system of equations (4.2)–(4.6) by the method of successive approximations in the amplitude of the field. In the zeroth approximation all atoms are in the ground state and possess a Maxwellian velocity distribution:

$$\begin{gathered}\eta = \eta_l = \eta_{ij} = 0,\\ \rho(\mathbf{v}) = W(\mathbf{v}) = (\pi^{1/2} v_T)^{-3} \exp[(\mathbf{v}/v_T)^2],\\ v_T = \sqrt{2T/m}.\end{gathered} \tag{4.13}$$

In an approximation linear in G_i, the equation for $\rho_i(\mathbf{v})$

$$\begin{gathered}[\Gamma - i(\Omega - \mathbf{kv})]\rho_i(\mathbf{v}) = i\nu_q \langle \rho_i \rangle + iG_i W(\mathbf{v}),\\ \nu_q = \tilde{\nu} + \bar{\nu} = 2\bar{\nu}_1 = -1.888\gamma,\end{gathered} \tag{4.14}$$

has the form of the standard model of strong collisions with a purely imaginary arrival frequency $i\nu_q$. The solution of Eq. (4.14) has the form (see [3, 4, 12])

$$\begin{gathered}\rho_i(\mathbf{v}) = iG_i L(\Omega, \mathbf{v}),\\ L(\Omega, \mathbf{v}) = \frac{W(\mathbf{v})}{[\Gamma - i(\Omega - \mathbf{kv})](1 - i\nu_q w)},\end{gathered} \tag{4.15}$$

$$\begin{gathered}\langle \rho_i \rangle = iG_i \langle L \rangle, \quad \langle L \rangle = \frac{w}{1 - i\nu_q w},\\ w = \int \frac{W(\mathbf{v})d\mathbf{v}}{\Gamma - i(\Omega - \mathbf{kv})}.\end{gathered} \tag{4.16}$$

The work performed by the field $P(\Omega)$ per unit time (or the absorbed power) is given by the general formula [3, 4]

$$P(\Omega) = -2h\omega \mathrm{Re}(i\langle \rho_i \rangle G_i^*). \tag{4.17}$$

In the linear approximation (4.16), we obtain

$$P(\Omega) = 2h\omega |G_i|^2 \mathrm{Re}\langle L \rangle. \tag{4.18}$$

Therefore, the quantity $\mathrm{Re}\langle L \rangle$ describes the frequency dependence of the power or the contour of the spectral absorption line. The formula (4.16) determines the so-called Rautian–Sobel'man contour [12]. If the arrival frequency possessed a real part, the expression (4.16) would contain (in the limit of small Doppler broaden-

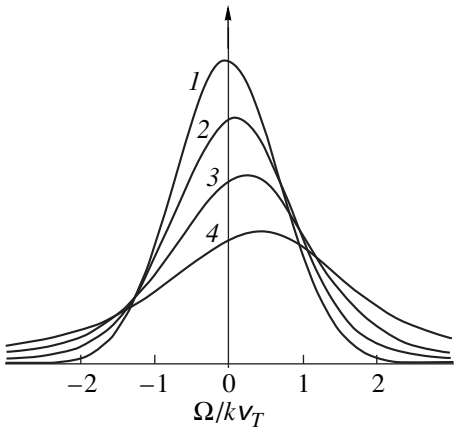

Fig. 2. Plot of the function $\pi^{-1/2}kv_T\mathrm{Re}\langle L\rangle$ (line contour). The curves *1*, *2*, *3*, and *4* correspond to $\Gamma/kv_T = 0$, 0.2, 0.5, and 1.0, respectively.

ing) Dicke line narrowing. Since in our case the arrival frequency is purely imaginary, the contour (4.16) broadens as Γ increases and undergoes a shift. In limiting cases the following approximate expressions follow from the formula (4.16):

$$\langle L\rangle \approx \frac{\pi^{1/2}}{kv_T}\left[1 - \frac{2\Gamma}{\pi^{1/2}kv_T} - \frac{(\Omega + 2\nu_q)^2}{(kv_T)^2}\right], \qquad (4.19)$$
$$\Gamma, |\Omega| \ll kv_T,$$

$$\langle L\rangle \approx \left[\Gamma - i(\Omega + \nu_q) + \frac{(kv_T)^2/2}{\Gamma - i\Omega}\right]^{-1}, \qquad (4.20)$$
$$\Gamma \gg kv_T.$$

It is evident from the expressions presented that for low collision frequencies the maximum of the line occurs at the point $\Omega_1 = -2\nu_q$. The shift occurs in the short-wavelength direction, since $\nu_q < 0$. The value of $\mathrm{Re}\langle L\rangle$ at the maximum decreases with increasing Γ. This decrease is monotonic, and for $\Gamma \gg kv_T$ the contour becomes Lorentzian with a maximum at the point $\Omega_2 = -\nu_q$. The contour is asymmetric in accordance with general model conclusions [12]. Figure 2 illustrates the contour properties listed above. The relation between Γ and ν_q in the calculations were chosen according to the data (4.9), i.e.,

$$\Gamma/\nu_q = -3.$$

Thus, Dicke narrowing does not occur for resonant dipole-dipole interactions of identical colliding particles. This factor is due to the absence of the arrival term $S^{(2)}(m_1 m\mathbf{q})$ (see Eq. (3.6)). Since the forward scattering amplitudes appear in the outgoing term $S^{(1)}(m_1 m\mathbf{q})$, the conclusion drawn remains in force when the change in velocity in a collision is taken into account. The imaginary nature of the arrival frequency (in terms of the strong-collisions model) is evidently due to the phase shift by $\pi/2$ in the exchange channel as compared with the zero-exchange channel; this is characteristic for resonance processes.

The quadratic approximation is found from Eqs. (4.3)–(4.5), if in these equations ρ_i and $\langle\rho_i\rangle$ are taken in the first approximation (4.14) and (4.15) and ρ and $\langle\rho\rangle$ in the zeroth approximation (4.12). The polarization tensor $G_iG_j^*$, which it is desirable to decompose in terms of irreducible tensors, appears in this approximation:

$$G_iG_j^* = \frac{1}{3}\delta_{ij}J + i\varepsilon_{ijl}J_l + J_{ij},$$
$$J_{ij} = J_{ji}, \quad J_{ii} = 0,$$
$$J = G_kG_k^*, \quad J_i = \frac{i}{2}\varepsilon_{ilk}G_kG_l^*, \qquad (4.21)$$
$$J_{ij} = \frac{1}{2}\left(G_iG_j^* + G_jG_i^* - \frac{2}{3}\delta_{ij}G_kG_k^*\right).$$

The equations for the quantities which are quadratic in the field have the form

$$(\gamma_0 + \nu_0)\eta = \tilde{\nu}_0 W(\mathbf{v})\langle\eta\rangle - 2J\mathrm{Re}[L(i\bar{\nu}_0\langle L^*\rangle - 1)],$$
$$(\gamma_0 + \nu_1)\eta_i = \tilde{\nu}_1 W(\mathbf{v})\langle\eta_i\rangle - 2J_i\mathrm{Re}[L(i\bar{\nu}_1\langle L^*\rangle - 1)], \qquad (4.22)$$
$$(\gamma_0 + \nu_2)\eta_{ij} = \tilde{\nu}_2 W(\mathbf{v})\langle\eta_{ij}\rangle - 2J_{ij}\mathrm{Re}[L(i\bar{\nu}_2\langle L^*\rangle - 1)].$$

The quantities $\langle\eta\rangle$, $\langle\eta_i\rangle$, and $\langle\eta_{ij}\rangle$ are found from Eqs. (4.22) by integrating over $\mathbf{v}$. Substituting the expressions obtained into Eq. (4.22) gives

$$\eta(\mathbf{v}) = \frac{2J}{\gamma_0 + \nu_0}$$
$$\times\left\{\frac{\nu_0}{\gamma_0}W(\mathbf{v})\mathrm{Re}\langle L\rangle + \mathrm{Re}L - \mathrm{Re}(i\bar{\nu}_0\langle L^*\rangle L)\right\},$$
$$\eta_i(\mathbf{v}) = \frac{2J_i}{\gamma_0 + \nu_1}$$
$$\times\left\{\frac{\tilde{\nu}_1 W(\mathbf{v})}{\gamma_0 + \nu_1 - \tilde{\nu}_1}\mathrm{Re}\langle L\rangle + \mathrm{Re}L - \mathrm{Re}(i\bar{\nu}_1\langle L^*\rangle L)\right\}, \qquad (4.23)$$
$$\eta_{ij}(\mathbf{v}) = \frac{2J_{ij}}{\gamma_0 + \nu_2}$$
$$\times\left\{\frac{\tilde{\nu}_2 W(\mathbf{v})}{\gamma_0 + \nu_2 - \tilde{\nu}_2}\mathrm{Re}\langle L\rangle + \mathrm{Re}L - \mathrm{Re}(i\bar{\nu}_2\langle L^*\rangle L)\right\}.$$

In this approximation η, η_i, and η_{ij} are proportional to, respectively, J, J_i, and J_{ij} and consist of identical structural elements which differ only by the weight

with which they appear in the expression (4.23). The term ReL determines the Bennett structure of an approximately Lorentzian form with characteristic width Γ and lying near the velocity $\mathbf{k}\cdot\mathbf{v} = \Omega$. The first terms in the braces describe the so-called collisional band of uniform saturation with a Maxwellian shape, and its frequency dependence is identical to the line contour (4.16) obtained in the first approximation. These two elements are well known [3, 4]. The new elements which are due to the polarization transitions are described by the terms $\mathrm{Re}(i\bar{\nu}_\kappa\langle L^*\rangle L)$. In accordance with the general results [the relations (4.12)], they do not change the velocity integrals of the quantities and only influence the velocity distribution. We shall write out the explicit expressions:

$$\mathrm{Re}L = \frac{W(\mathbf{v})}{|1-i\nu_q w|^2} \times \frac{(1+\nu_q w'')\Gamma - (\Omega-\mathbf{k}\cdot\mathbf{v})w'\nu_q}{\Gamma^2+(\Omega-\mathbf{k}\cdot\mathbf{v})^2}, \tag{4.24}$$

$$\bar{\nu}_\kappa \mathrm{Re}(iL\langle L^*\rangle) = \frac{W(\mathbf{v})}{|1-i\nu_q w|^2} \times \frac{w''\Gamma - (\Omega-\mathbf{k}\cdot\mathbf{v})w'}{\Gamma^2+(\Omega-\mathbf{k}\cdot\mathbf{v})^2}\bar{\nu}_\kappa, \tag{4.25}$$

$$w(\Omega) = w'(\Omega) + iw''(\Omega), \quad w'(-\Omega) = w'(\Omega), \quad w''(-\Omega) = -w''(\Omega). \tag{4.26}$$

It is evident that the main term in the expression (4.25) is the term that is antisymmetric as a function of $\Omega - \mathbf{k}\cdot\mathbf{v}$. The same antisymmetric term is present also in Eq. (4.24). The ratio of the antisymmetric terms in Eqs. (4.25) and (4.24) is $\bar{\nu}_\kappa/\nu_q$, which, according to Eq. (4.9), is –0.41, 0.50, and –0.050 for $\kappa = 0$, 1, and 2, respectively.

Thus, the polarization transitions and the imaginary nature of the arrival frequency ($i\nu_q$) form an antisymmetric component in the Bennett structure, and because of the variation in the signs and the absolute values of $\bar{\nu}_\kappa$ its contribution is different for different irreducible tensors.

To calculate $P(\Omega)$ to within $|G|^4$, i.e., to determine the first nonlinear corrections, it is necessary to find $\rho_i(\mathbf{v})$ in the cubic approximation. We shall not do this: just as in other related phenomena of nonlinear spectroscopy, $\langle\rho_i\rangle$ will not contain sharp nonlinear resonances, even though a Bennett structure is present. To determine the latter it is necessary to use one of the variants of the probe-field method, which "probes" the velocity distribution produced by a strong field. Let us consider the simplest method of a counterpropagating probe wave with the same frequency ω [7]. In this case the nonlinear interference effects are negligible and only the change in the velocity distribution need be taken into account [3, 4]. The equation for the off-diagonal element $\rho_{i\mu}$, induced by the counterpropagating wave (wave vector $-\mathbf{k}$, amplitude $E_{i\mu}$), has the form

$$\begin{aligned}[\Gamma - i(\Omega+\mathbf{k}\cdot\mathbf{v})]\rho_{i\mu} - i\nu_q W(\mathbf{v})\langle\rho_{i\mu}\rangle - iG_{i\mu}W(\mathbf{v}) \\ = -2i\nu_q\eta(\mathbf{v})\langle\rho_{i\mu}\rangle - i\bar{\nu}[\eta(\mathbf{v})\delta_{ij} - \rho_{ij}(\mathbf{v})]\langle\rho_{j\mu}\rangle \\ - i[G_{i\mu}\eta(\mathbf{v}) + \rho_{ij}(\mathbf{v})G_{j\mu}], \\ G_{i\mu} = \frac{dE_{i\mu}}{2h}.\end{aligned} \tag{4.27}$$

Terms where η and ρ_{ij} must be expressed using Eqs. (4.23) of the quadratic approximation and $\langle\rho_{i\mu}\rangle$ from the linear approximation are collected on the right-hand side. In calculating the work performed by the probe field

$$P_\mu(\Omega) = -2h\omega\mathrm{Re}(i\langle\rho_{i\mu}G^*_{i\mu}\rangle) \tag{4.28}$$

a polarization tensor of the probe field $G_{i\mu}G^*_{i\mu}$ will arise. We shall expand it in terms of irreducible tensors as

$$G_{i\mu}G^*_{i\mu} = \frac{1}{3}\delta_{ij}J_\mu + i\varepsilon_{ijl}J_{l\mu} + J_{ij\mu}. \tag{4.29}$$

After simple transformations, we arrive at the relations

$$\begin{aligned}\frac{P_\mu(\Omega)}{2h\omega} = \mathrm{Re}\langle L\rangle J_\mu - \mathrm{Re}(1+i\nu_q w)^{-1} \\ \times\left\{\frac{4}{3}\left[1 - i\left(\nu_q + \frac{\tilde{\nu}}{2}\right)\langle L\rangle\right](a_0+b_0+c_0)\langle L\rangle JJ_\mu\right. \\ + (1-i\bar{\nu}\langle L\rangle) \\ \left.\times[(a_1+b_1+c_1)2J_lJ_{l\mu} + (a_2+b_2+c_2)J_{ij}J_{ij\mu}]\right\},\end{aligned} \tag{4.30}$$

$$\begin{aligned}a_\kappa = \frac{2}{\gamma_0+\nu_\kappa}\frac{\tilde{\nu}_\kappa a(\Omega)}{\gamma_0+\nu_k-\tilde{\nu}_k}, \\ b_\kappa = \frac{2b(\Omega)}{\gamma_0+\nu_k}, \quad c_\kappa = -\frac{2\bar{\nu}_\kappa c(\Omega)}{\gamma_0+\nu_k}, \\ a(\Omega) = \langle L\rangle w, \quad b(\Omega) = \left\langle\frac{\mathrm{Re}L}{\Gamma - i(\Omega+\mathbf{k}\cdot\mathbf{v})}\right\rangle, \\ c(\Omega) = \left\langle\frac{\mathrm{Re}(i\langle L^*\rangle L)}{\Gamma - i(\Omega+\mathbf{k}\cdot\mathbf{v})}\right\rangle.\end{aligned} \tag{4.31}$$

According to Eq. (4.9), the quantities $\gamma_0 + \nu_\kappa$ do not vary much; their difference from the average value does not exceed 20%. The frequencies $\tilde{\nu}_k$ differ in sign, and they differ in absolute value by an order of magnitude. This also true of the collection of frequencies $\bar{\nu}_k$. The quantities a_κ, b_κ, and c_κ possess spectral contours which are independent of κ. The terms a_κ describe a structureless band with an almost symmetric Doppler width, if

Table

	(↑↑)	(↑→)	(↑+)	(++)	(+–)
JJ_μ	0	1	1	1	1
$J_lJ_{l\mu}$	1	0	0	1/4	–1/4
$J_{ij}J_{ji\mu}$	2/3	–1/3	1/6	1/6	1/6

$\Gamma \ll kv_T$. The functions $b(\Omega)$ and $c(\Omega)$ can be presented as

$$b(\Omega) = \left[\frac{\langle L\rangle}{\Gamma - i\Omega} + (1 - i\nu_q w)^{-1}\frac{w''}{\Omega}\right], \qquad (4.32)$$

$$c(\Omega) = \frac{i}{2}|1 - i\nu_q w|^2\left(\frac{|w|^2}{\Gamma - i\Omega} - \frac{ww''}{\Omega}\right). \qquad (4.33)$$

For $\Gamma \ll kv_T$ $\mathrm{Re}\,b(\Omega)$ and $\mathrm{Re}\,c(\Omega)$ make the main contribution to the nonlinear component $P_\mu(\Omega)$. It is evident from the expressions (4.32) and (4.33) that $b(\Omega)$ will determine the symmetric part of the nonlinear resonance of the form $\Gamma/(\Gamma^2 + \Omega^2)$, and $c(\Omega)$ will determine the antisymmetric part $\Omega/(\Gamma^2 + \Omega^2)$. The ratio $|c(\Omega)/b(\Omega)|$ is $\bar{\nu}_\kappa/kv_T$ in order of magnitude.

We shall now consider the polarization properties of the nonlinear resonances. We choose for the *z*-axis the direction of the wave vector. We shall consider the linear and circular polarizations, which are usually used, and we shall introduce the following symbolic designations: we denote the linear polarizations along the *x* and *y* axes by ↑ and ⟶, respectively, and we denote by "+" and "–" the clockwise and counterclockwise circular polarizations. We take JJ_μ as 1. The numerical values of $J_lJ_{l\mu}$ and $J_{ij}J_{ij\mu}$ for the combinations of the indicated polarizations in the strong and counterpropagating waves are presented in the table. Combining the signals corresponding to different columns in the table, we can identify the contributions of various irreducible tensors. For example, the sum of the signals (↑↑) + 2(↑⟶) will contain only a scalar part, the difference (++) – (+–) will be proportional only to the orientation, the difference (↑↑) – (↑⟶) will be proportional only to the alignment, and so on.

We studied above the simplest nonlinear resonance of a counterpropagating probe wave. Similar asymmetric elements of nonlinear resonances also exist in the absorption spectrum of the probe field, resonant to adjoining transitions (e.g., $m_1 - l$, $m - g$).

5. DISCUSSION

One of the main ideas of nonlinear saturation spectroscopy consists in describing the relaxation and nonlinear effects using additive, independent terms in the quantum kinetic equation—the field, dynamic term $i[V, \rho]$ (V is the atom–field interaction Hamiltonian) and the collision integral S, where the thermostat is assumed to be in an equilibrium state [3, 4, 13]. This idea has been used as the basis for solving an enormous number of specific problems and for interpreting many phenomena; it has become a self-evident, common truth. However, this important idea has turned out to be unsound for self-broadening or, more generally, for the interaction of a buffer gas ("thermostat") with sufficiently powerful radiation. Three types of specific nonlinear effects associated with the collision integral were revealed in Sections 2–4. First, there is the obvious dependence of the relaxation characteristics on the intensity, frequency, and polarization of the field. The second factor is more subtle: specific polarization transitions that do not exist in the absence of a field. These transitions are of an interference nature, since they are due to the interference of exchange and zero-exchange scattering channels. They are unusual also in the kinetic respect, since the number of polarization transitions per unit time is proportional to the product of not the concentration of the colliding particles but rather the optical coherences induced by the field. The polarization transitions, by virtue of their interference origin, produce sign-alternating elements in the fine structure of the velocity distribution with a zero integral and, correspondingly, asymmetric elements in the nonlinear resonances of the probe field.

The third mechanism for coupling of the collision integral with intense radiation is the field-induced spatial nonuniformity of the buffer and experimental gas. This factor, similarly to the dynamic interaction of an atom and a field, engenders a unique recoil effect, which is manifested through the scattering amplitude and the density matrix of the colliding atoms.

ACKNOWLEDGMENTS

It is my pleasant duty to thank A.P. Kol'chenko, P.L. Chapovskiĭ, and A.M. Shalagin for a discussion of the questions analyzed in this paper.

This work was supported by the State Scientific and Technical Program "Fundamental Spectroscopy" (grant no. 08.02.31) and the Federal target program "Integration" (grant no. 274).

REFERENCES

1. A. P. Kazantsev, Zh. Éksp. Teor. Fiz. **51**, 1751 (1966) [Sov. Phys. JETP **24**, 1183 (1966)].
2. S. G. Rautian, Zh. Éksp. Teor. Fiz. **103**, 785 (1993) [Sov. Phys. JETP **76**, 386 (1993)].
3. S. G. Rautian, G. I. Smirnov, and A. M. Shalagin, *Nonlinear Resonances in Atomic and Molecular Spectrum* (Nauka, Novosibirsk, 1979).
4. S. G. Rautian and A. M. Shalagin, *Kinetic Problems of Nonlinear Spectroscopy* (North-Holland, Amsterdam, 1991).

5. R. G. Newton, *Scattering Theory of Waves and Particles* (McGraw-Hill, New York, 1966; Mir, Moscow, 1969).
6. A. P. Kol'chenko, S. G. Rautian, and R. I. Sokolovskiĭ, Zh. Éksp. Teor. Fiz. **55**, 1864 (1968) [Sov. Phys. JETP **28**, 986 (1968)].
7. V. S. Letokhov and V. P. Chebotaev, *Nonlinear Laser Spectroscopy* (Springer-Verlag, Berlin, 1977).
8. J. L. Hall, C. J. Borde, and J. Uehara, Phys. Rev. Lett. **37**, 1339 (1976).
9. D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskii, *Quantum Theory of Angular Momentum* (Nauka, Leningrad, 1975; World Scientific, Singapore, 1988).
10. Yu. A. Vdovin and V. M. Galitskiĭ, Zh. Éksp. Teor. Fiz. **52**, 1345 (1967) [Sov. Phys. JETP **25**, 894 (1967)].
11. V. K. Matskevich, I. E. Evseev, and V. M. Ermachenko, Opt. Spektrosk. **45**, 17 (1978) [Opt. Spectrosc. **45**, 8 (1978)].
12. S. G. Rautian and I. I. Sobel'man, Usp. Fiz. Nauk **90**, 209 (1966) [Sov. Phys. Usp. **9**, 701 (1967)].
13. S. G. Rautian, Zh. Éksp. Teor. Fiz. **51**, 1176 (1966) [Sov. Phys. JETP **24**, 788 (1966)].

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 4, 2000, pp. 724–735.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 4, 2000, pp. 835–848.

PLASMA, GASES

Step Propagation of a Streamer in an Electronegative Gas

N. L. Aleksandrov[a, *] and E. M. Bazelyan[b]

[a]*Moscow Institute of Physics and Technology, Dolgoprudnyi, Moscow oblast, 141700 Russia*
[b]*Krzhizhanovskii Power Engineering Institute, Moscow, 117927 Russia*
**e-mail: alek@alef.crec.mipt.ru*

Received March 21, 2000

Abstract—It is shown on the basis of numerical simulation and an experimental investigation that a streamer can propagate in a step manner in an electronegative gas. The experiments and most calculations were performed for air under close to normal conditions. The step motion is associated with the appearance of a secondary ionization wave near the electrode and propagation of this wave along the channel of the streamer; this wave maintains the channel in a conducting state and allows for the propagation of the streamer in a nonuniform external field over distances which are inaccessible under ordinary conditions of a streamer discharge. Simulation in heated air, oxygen, and SF_6 demonstrated that the phenomenon studied is common for various gases and that the special features of its manifestation remain in a wide range of decay rates of the streamer channel. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

A streamer is the structurally simplest component of a long spark. A strong electric field wave, on whose front intense ionization produces plasma, creates a streamer channel. The small specific energy release leaves the gas in the channel essentially cold, as a result of which the electrons produced vanish quite rapidly in electron-ion recombination reactions and they also vanish in an electronegative medium (largely because of attachment to neutral particles). The loss of electrons is essentially not made up by ionization, which in a cold gas requires a strong electric field, which, as a rule, is not attainable in the channel behind the wave front (behind the streamer head). The voltage drop on the channel, increasing in time and with distance, decreases the potential and the field of the streamer head, as a result of which streamer growth slows down and completely stops in the gaps with a sharply nonuniform electric field.

In the streamer process explained above there seems to be no place for the step formation that is typical for many manifestations of a long spark. The step development of a spark discharge in long gaps with sharply nonuniform electric fields is well known from experiments [1–3]. Most such data refer to the so-called leader process. The leader mechanism has been studied in greatest detail for air, for which extensive experimental data have been accumulated for lightning and multimeter laboratory sparks. The step development of a leader depends on the conditions, but it is always associated with the complicated structure of the leader, first and foremost, the processes occurring in the streamer zone in front of the head of the leader channel. This zone is filled with numerous streamers, successively starting from the head with frequency 10^{10} s^{-1} [3]. The streamers together form a current, which, feeding the leader, ultimately heats the channel of the leader up to arc temperatures. The space-charge cover favors heating. Streamers inject this charge into the streamer zone, and as the leader channel grows the charge surrounds it. The reverse field generated by such a charged cover, by lowering the radial field on the outer surface of the channel, whose potential is ~1–10 MV and higher (the latter is characteristic for lightning), prevents ionization expansion. This maintains the radius of the channel at relatively small values and gives rise to a high density of energy release, which fundamentally distinguishes a leader from a streamer [3]. The lifetime of the leader is essentially infinite, and it is capable of growing to tens of kilometers; ultimately, electric fields of the order of the fields in arcs are sufficient to maintain the leader channel in a conducting state.

A typical example of a step spark process is a step leader of a negative long spark or lightning. The phenomenon recorded in streak photographs as a series of successive flashes of a channel continuously increasing in length is actually due to the interaction of the main channel, growing extremely slowly from the cathode, and volume positive leaders moving toward it [3, 4]. Each next volume leader starts from the interior of the gap, using the local intensification of the field at the ends of the negative anode-directed streamers of the main leader, which have still not lost their conductivity. The high-conductivity streamer ends become polarized in an external field approximately in the same way as small metal rods placed in the same location (the experiment showed that metal and plasma are completely "interchangeable" here). Cathode-directed streamers can be initiated from them, just as from metal electrodes. Then, if the field is sufficiently strong, a positive leader can be initiated. This leader is named a volume leader

after the location where it appears.[1] When such a volume leader reaches the main leader its channels merge and, at the end of the wave charge-exchange process, acquire a common potential. The rapid release of energy accompanying charge transfer intensifies the radiation and produces a bright flash, demonstrating an instantaneous, step increase in the length of the main leader by the length of the volume leader which has grown (the latter emits much more weakly as it grows; it is virtually indistinguishable in lightning, though it is clearly recorded in the laboratory, where image converters with light amplification are used [4]).

The present work studies a streamer, which in contrast to a leader does not have interacting structural components that can produce a quasi-step structure. Here one can talk only about growth with a truly step nature. To observe the process the potential of the streamer head must vary nonmonotonically, growing repeatedly after each interval of decrease. If the artificial situation with a step voltage pulse on a gap is neglected completely, this is possible only with initiation of a secondary ionization wave. It should restore the conductivity of the plasma in the existing channel and consequently increase the potential of the head, determining the velocity of the streamer.

The propagation of a secondary wave, called a secondary streamer, was observed after the streamer covered the discharge gap (see, for example, [5–7]). At the moment of overlapping the head of the positive streamer acquires the cathode potential and the entire applied voltage "falls" across the channel. After the redistribution of the surface charge of the channel is completed, the current along it stabilizes and the field E becomes inversely proportional to the streamer conductivity per unit length. Consequently, the strongest field occurs near the start location, where the plasma decayed in the longest time. If this field exceeds the threshold, for which the ionization rate becomes greater than the electron loss rate, a secondary ionization wave propagating from the channel base to the head forms [6, 7]. It has been reported that the secondary streamer has also been observed before the gap is covered [5, 8].

Numerous numerical simulations (see, e.g., the citations in [3, 9]) as well as analytical works (see, e.g., [10, 11]) examine regimes where a streamer propagates either with a constant velocity (ordinarily in a uniform external field) or it possesses two distinct phases: acceleration and subsequent deceleration down to complete stoppage. (If the streamer is simulated in a short time interval, then there may not be enough time for the second phase to occur.) At the same time, experimental evidence [12] shows that, as the voltage across the discharge gap continuously increases, the streamer can move in a step manner with its velocity slowing to zero, after which the streamer once again accelerates. Observations of this phenomenon are very incomplete, and there is no theoretical explanation for it. To detect a step streamer the leading edges of the voltage pulses must not be too short (longer than 10^{-7}–10^{-6} s), the electrode radii must be small (less than 1 cm), and there should be no delay in the appearance of the first electron in the region of the strong field after a voltage is applied to the discharge gap. Since the repeated "revival" of the streamer after stoppage necessarily presupposes restoration of conductivity in its channel by a second wave of ionization, in a step streamer process the latter wave should arise regularly because of the voltage increasing across the gap.

In this paper, we present the results of an experimental investigation of the step development of a positive streamer in air in a sharply nonuniform electric field (the conditions are similar to those described in [12]) and a numerical simulation is performed for the experimental conditions that makes it possible to determine the reasons for the step nature. It is shown that a streamer is capable of preserving a galvanic coupling between its head and the start location for times much longer than the lifetime of a cold streamer plasma, as a result of which the streamer can propagate over distances that are unattainable for previously known mechanisms of streamer propagation. The results of the simulation of the step development of a streamer transfer to other gases (heated air, oxygen, SF_6), which makes it possible to determine the special features of the phenomenon under study in gases with substantially different electron loss rates.

2. DESCRIPTION OF THE THEORETICAL MODEL AND THE EXPERIMENTAL CONDITIONS

Atmospheric-pressure air at temperature T = 300–600 K as well as oxygen and SF_6 were chosen as the gases. The purpose of the simulation was to determine the qualitative aspects of the process. Consequently, a very simple one-and-half dimensional (1.5D) model of a streamer was used [3]. In this model the radius of the streamer channel is assumed to be constant and given a priori, the plasma parameters are averaged over the cross section of the channel, and the longitudinal electric field on the axis is assigned to the entire cross section, while the radial field inside the channel is neglected. All radial processes, including diffusion, are neglected. As a result, the particle densities and the longitudinal electric field depend only on the axial coordinate x and the time. The excess electric charge is assumed to lie only on the surface of the channel and in the streamer head; this makes it possible to calculate the longitudinal electric field using simple analytical formulas relating the excess electric charge in the computed section of the channel and the electric field cre-

[1] The much slower negative leader from the opposite plasma end does not play any special role in the organization of a step, though it is important for maintaining charge balance [4].

ated by this charge, instead of solving the Poisson equation numerically.

The 1.5D model is used because the main events to be investigated occur for channel lengths of tens of centimeters, while the recorded lengths of streamers, reproducible in two-dimensional models, are at least an order of magnitude shorter [3]. As is well known, the main drawback of the 1.5D model is that the streamer radius r_s, on which the maximum field on the front of the ionization wave and the initial density of the plasma produced strongly depend [3, 9], is set a priori. Nonetheless, the model is used quite extensively in numerical experiments when qualitative relations in the streamer process must be given. This is due to the fact that many important streamer parameters, including the velocity and conductivity per unit length of the plasma channel produced, depend not on the radius r_s of the streamer channel but rather on the product $r_s^n E_{\max}^m$, and since the maximum field at the wave front $E_{\max} \propto r_s^{-1}$, the quantitative uncertainty of the results is not so large for $m \approx n$. (For a streamer in air, ordinarily, m = 1–1.5 and n = 1–2.)

The equations of the model, the kinetic scheme for air, and the computational method are described in detail in [9, 13]. The numerically solved equations are the Poisson equation, the equation of balance of charged particles (electrons and positive and negative ions), and neutral active particles (excited particles and atoms), which can influence the electron balance in the streamer channel. In the simplest form, for a single type of active particle (excited molecules), positive and negative ions, this system of equations has the form

$$\Delta\varphi = e\varepsilon_0^{-1}(n_e + n_n - n_p), \tag{1}$$

$$\frac{\partial n_e}{\partial t} + \operatorname{div}(n_e \mathbf{V}_e) = (k_i N + k_i^* n^*) n_e - (k_a' + k_a'' N) N_a n_e + (k_d N + k_d^* n^*) n_n - \beta_{ei} n_p n_e + S_f, \tag{2}$$

$$\frac{\partial n_p}{\partial t} + \operatorname{div}(n_p \mathbf{V}_p) = (k_i N + k_i^* n^*) n_e - \beta_{ei} n_p n_e - \beta_{ii} n_p n_n + S_f, \tag{3}$$

$$\frac{\partial n_n}{\partial t} + \operatorname{div}(n_n \mathbf{V}_n) = (k_a' + k_a'' N) N_a n_e - (k_d N + k_d^* n^*) n_n - \beta_{ii} n_p n_n, \tag{4}$$

$$\frac{\partial n^*}{\partial t} = k^* N n_e - k_i^* n^* n_e - k_q^* N n^*, \tag{5}$$

where N_a, n_e, n_p, n_n, and n^* are, respectively, the densities of the neutral molecules of the gas as a whole, all electronegative components, electrons, positive and negative ions, and excited particles; $\mathbf{V}_k$ is the drift velocity of charged particles of the type k; k_i and k_i^* are the rate constants for electron-impact ionization of unexcited and excited molecules; k_a' and k_a'' are the rate constants for dissociative and three-body attachment of electrons to molecules; k_d and k_d^* are the rate constants for electron detachment from negative ions colliding with unexcited and excited molecules; k^* and k_q^* are the rate constants for excitation and quenching of electronically excited molecules; β_{ei} and β_{ii} are electron–ion and ion–ion recombination coefficients; S_f is the source of electrons produced as a result of photoionization; and, φ is the electric field potential.

The actual complexity of the system being solved depends on the composition of the gas, which largely determines the mechanisms of the decay of the plasma in the streamer channel. In air the rate of plasma decay at the early stage is determined by dissociative recombination of electrons with complex positive ions, while at the latter stage it is determined by three-particle attachment of electrons to O_2 molecules. The negative ions formed in the process can lose electrons in collisions with O atoms and other active particles produced in the streamer head. Consequently, when describing the dynamics of streamer development in air at times comparable to or longer than the lifetime of the streamer plasma, it is necessary to take account of the equations of balance of complex ions (their formation substantially accelerates electron-ion recombination) and active particles, in collisions with which negative ions are efficiently destroyed. As a result, about 50 components and more than 200 ion-molecular processes are taken into account. The kinetics of the decay of a streamer channel in SF_6 is much simpler. Here the main electron loss channel is electron attachment to molecules, and electron detachment from negative ions can be neglected, in part because of the high binding energy of the outer electron in negative ions and in part because of the low density of active particles in the streamer channel. Consequently, all three charged components and four reactions (impact ionization, electron attachment to molecules, and electron–ion and ion–ion recombination) can be circumvented when describing the decay of a streamer plasma in SF_6.

The system of Eqs. (1)–(5) is closed and self-consistently describes processes near the streamer head, where most ionization occurs, and in the streamer channel, where the plasma decays. The system of equations under study was solved numerically using finite differences with an adaptive grid. The computational step was compressed in the direction of the axis near the anode (in a region of the order of the anode radius) and in the streamer head, and it was the same in all other regions of the streamer.

The kinetic scheme employed in this work for cold air is described in [9, 13]; the scheme for hot air is described in [14]. The processes in oxygen were taken

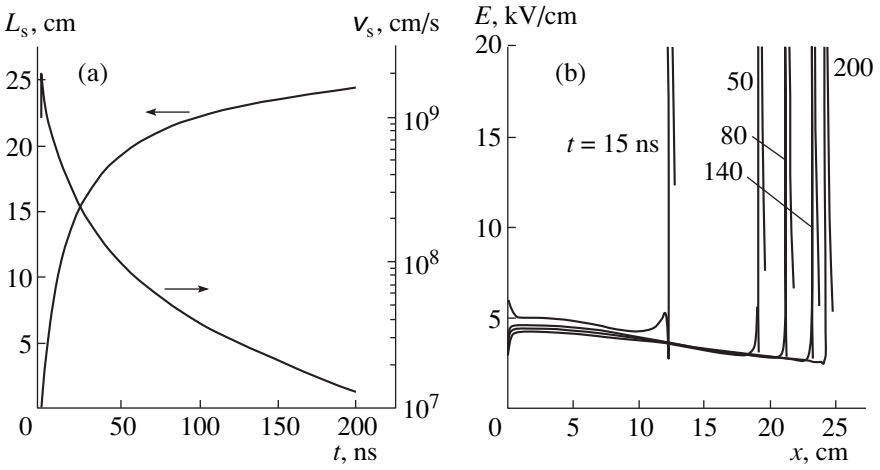

Fig. 1. Streamer characteristics in air with a rectangular voltage pulse: (a) time variation of the streamer velocity v_s and streamer length L_s; (b) spatial distribution of the electric field for various moments in time.

by analogy with the processes in air (with the exception of nitrogen molecules), and the kinetic scheme for SF_6 is taken from [15]. The general approach for determining the rate constants of various processes consist of the following. The rate constants of the electron processes (and the electron drift velocity), in the absence of experimental data, were calculated by solving Boltzmann equation numerically in the two-term approximation. The rate constants of all other processes, as a rule, were extracted from the experimental data.

The description of the experimental setup and the results of certain measurements are presented in [12]. The experiments were performed in atmospheric-pressure air at a temperature of about 300 K in a rod-plane gap up to 130 cm long. The rod anode, 1 cm in diameter, terminated with a sharp conical head, about 5 cm long. Voltage pulses with an exponentially rising front were applied to the gap. The initial, nearly linear section, where the rate of growth of the voltage changed by less than a factor of 2, was used for the measurements. The average slope of the working section of the front reached 3500 kV/μs. To prevent breakdown of the gap, in most experiments the voltage across the gap was cut off with a spark discharger 0.3–4 μs after the pulse generator was actuated.

Continuous streak photographs of streamers were recorded, using an image converter, to measure the velocity of the streamers. Since each streamer burst consisted of a large number of branching streamers developing from a needle-shaped anode, measures were taken to prevent overlapping of many streamers, which smeared the picture, during the recording process. A narrow slit (1–3 cm) oriented from one electrode to another was used for this. In a number of experiments the streamer burst was confined between dielectric barriers, which sharply decreased the number of branchings and streamers developing in parallel. All this simplified the interpretation of the streak pictures and increased their information content.

3. COMPUTATIONAL AND EXPERIMENTAL RESULTS

Most publications describe the results of numerical simulation of streamers with a constant voltage across the discharge gap. In reality this is possible only if an extremely steep voltage pulse with an approximately 1 ns leading edge is applied or if, because of the statistical delay in the appearance of the first electron in the strong-field region or because of the very large radius of the anode, the streamer starts during the pulse amplitude. An example of typical computational data for such a case is displayed in Fig. 1, which reproduces the propagation of a streamer from a 0.5 cm in radius spherical anode with a 100 kV rectangular voltage pulse. The medium is air at pressure p = 1 atm and temperature T = 300 K. The radius of the streamer channel was assumed to be r_s = 0.05 cm.

The streamer enters an external field of less than 1 kV/cm, having propagated only 7 cm into the gap in 7 ns. Nonetheless, it continues to grow actively for at least another 100 ns. In this time the velocity v_s decreases almost by two orders of magnitude down to 10^7 cm/s, which results in virtually complete stopping of the streamer at a distance of about 25 cm from the anode. Thus, with the exception of a short section of about 1 cm near the anode, the entire further development of the streamer proceeds in the deceleration regime. The potential and field at the streamer head decrease, and the electron density n_e in the plasma produced decreases with them. As far as the old sections of the channel closer to the anode are concerned, there is enough time for the initially high values of n_e to decrease appreciably as a result of electron attachment to O_2 molecules and dissociative electron–ion recombi-

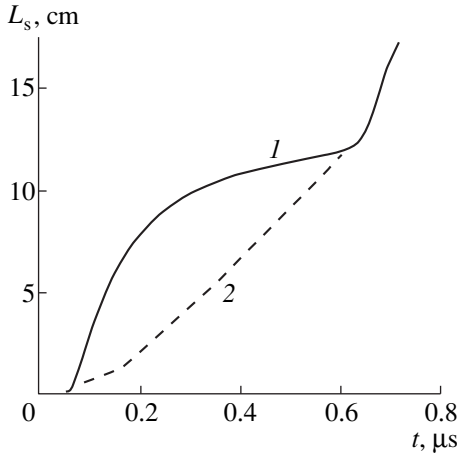

Fig. 2. Time variation of the streamer length (*1*) and position of the secondary wave of ionization (*2*) in air at linearly rising voltage pulse with slope 500 kV/μs.

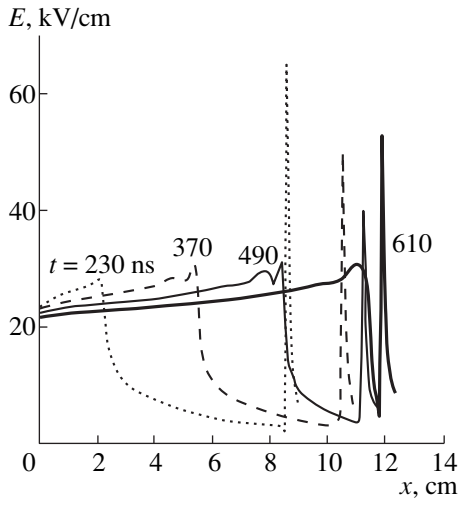

Fig. 3. Evolution of the electric field distribution in a streamer for air at linearly rising voltage pulse with slope 500 kV/μs.

nation. As a result, the electron density becomes more or less uniform along the entire streamer channel. The longitudinal electric field E_c, which in the channel does not rise above 5 kV/cm (Fig. 1b), also varies negligibly in space and time.

Streamer propagation on the front of the voltage pulse is of a completely different character. It is evident from the computational results presented in Fig. 2 and describing the process for an anode of the same radius at a linearly rising pulse with slope 500 kV/μs that the streamer, starting at a voltage of about 25 kV, subsequently moved with acceleration up to 85 ns. This was sufficient for the previously produced sections of the channel to lose conductivity appreciably, hindering transfer of the rising voltage from the anode to the streamer head. The length of the acceleration section hardly reached 2 cm. Subsequently, the streamer velocity decreased and at $t \approx 400$ ns approached the threshold value 10^7 cm/s, for which streamer motion in air can still be observed in experiments. The growth of the streamer channel virtually ceased at a distance of 10–11 cm (this is two times less than with a 100 kV rectangular pulse), though the voltage across the gap by this time exceeded 200 kV.

Such "wasteful" expenditures of voltage are due to the change in the mechanism leading to the formation of current in the streamer channel. If the voltage across the gap remains unchanged, the current in the channel is determined by delivery of charge to the newly produced sections of the streamer, which must be charged up to the head potential φ_h:

$$i = C_c v_s \varphi_h$$

where C_c is the capacitance per unit length of the channel. Since the streamer velocity v_s is proportional to the ionization rate $\nu_i \propto E_h^m$, where $m > 1$ (E_h is the electric field in the streamer head) [3], the dependence of the current on the potential φ_h is very strong: the current decreases rapidly as the voltage drop across the decaying channel increases. Such a streamer propagation regime cannot lead to growth of the longitudinal field in the channel.

Conversely, when the voltage across the gap increases, another term which does not depend on the rate of growth of the streamer is added to the expression for the current:

$$i \approx C_c v_s \varphi_h + C\frac{\partial \varphi_{av}}{\partial t},$$

where C is the total capacitance of the streamer and φ_{av} is the average potential along the streamer channel, varying from the applied voltage $U(t)$ in the channel base to φ_h at the head. This second current component additionally charges the already produced channel as the voltage across the gap increases. If the derivative $\partial\varphi_{av}/\partial t$ is large and the second term is larger than the first term, the total current does not change much during deceleration and even with complete stoppage of the streamer. In a decaying channel, where the conductivity per unit length decreases, this current can increase the longitudinal electric field

$$E_c = i(\pi r^2 e n_e \mu_e)^{-1},$$

where μ_e is the electron mobility. The field increases as the conductivity decreases at a given section of the channel. Consequently, the high-field region at first forms near the anode, where the channel lifetime is longer, and then it gradually advances in the direction of the streamer head. This is manifested as the motion of a secondary wave of intensification of the field, accompanied by ionization, along the channel (Figs. 2 and 3). Ionization limits the amplitude of the field wave. The field in the channel cannot rise much above

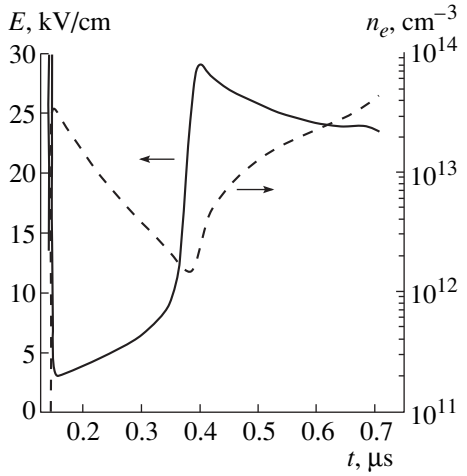

Fig. 4. Time variation of the electron density and electric field on the streamer axis at a point 5.9 cm from the anode. The calculations were performed for air at a linearly rising voltage pulse with slope 500 kV/μs.

the threshold value, because this would increase the rate of ionization very sharply and it would increase the electron density at this location of the channel and the plasma conductivity, which in turn would weaken the field. It is evident from Figs. 3 and 4 that the field behind the wave front is maintained somewhat lower than the threshold field $E_i \approx 30$ kV/cm for unperturbed air. This is as it should be, because the composition of the gas mixture in the streamer channel is altered by the production of active negative-ion annihilators, first and foremost, O atoms [9, 13]. On account of electron detachment from negative ions the positive balance of electrons in a channel is maintained by the field $E_{th} \approx 20$–23 kV/cm (Figs. 3 and 4).

The secondary wave moves along the streamer channel with almost constant velocity, about 2.5×10^7 cm/s in the computational example (Figs. 3 and 4). With respect to the velocity and the effect produced, such a wave can be interpreted as a secondary streamer.

A secondary streamer can be detected experimentally most clearly by using a step voltage pulse. After the section with a linear rise the voltage stabilizes in time Δt, and then it continues to grow at a comparable rate (Fig. 5a). In the time of the pause $\Delta t \approx 100$ ns the primary steamers stop, and the start of a secondary streamer from the anode can be clearly seen in the darkened, weakly emitting gap. It moves as the voltage continues to rise. Everything described above is well reproduced in the calculation using the 1.5D model, the results of which for conditions close to the experimental conditions are presented in Fig. 5b.

The appearance of a secondary streamer does not stimulate the almost stationary primary streamer. Both the calculations (Figs. 2 and 5b) and the experimental results (Fig. 5a) indicate this. The reaction appears only after the second wave of ionization restores conductivity along the entire channel produced. The head potential of the primary streamer increases rapidly, and it renews its motion along the gap primarily with a rapidly increasing velocity (Fig. 2), and then, just as at the initial stage of its development, with deceleration until it once again stops. Then the process repeats, if, of course, voltage growth has not stopped. Step development of a streamer, which is easily observed experimentally with not too high rate of growth of the voltage, obtains. The streak picture in Fig. 6 demonstrates this process in the needle–plane gap with a voltage pulse slope of about 100 kV/μs. The specially chosen slow sweep is insufficient to obtain details of the propagation of each individual streamer step, but a series of successive bursts of the advancing channel is recorded clearly.

A qualitatively similar picture is also obtained by 1.5D simulation (Fig. 7) for close conditions. The streamer steps follow one another with an interval of the order of 1 μs. The motion of a secondary wave of

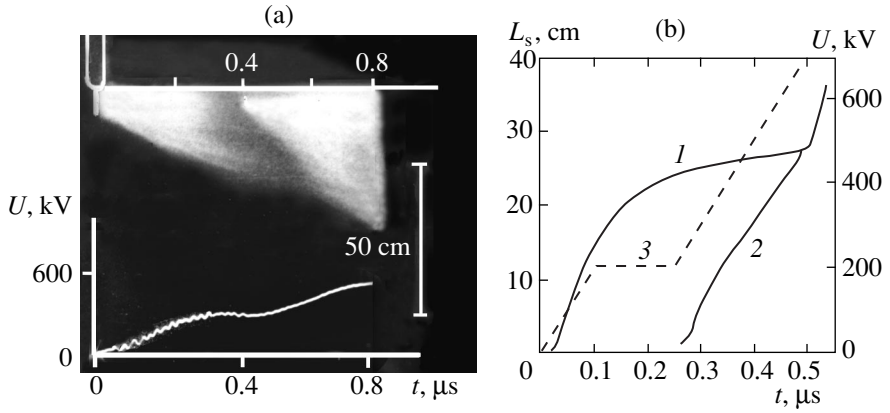

Fig. 5. Appearance of a secondary wave of ionization from the anode under the action of a voltage pulse with a stepped leading edge; (a) streak picture of a discharge and synchronous oscillogram of the voltage pulse; (b) time variation of the computed streamer length (*1*) and position of the secondary wave of ionization (*2*) as well as the voltage pulse shape (*3*) for which the calculation was performed.

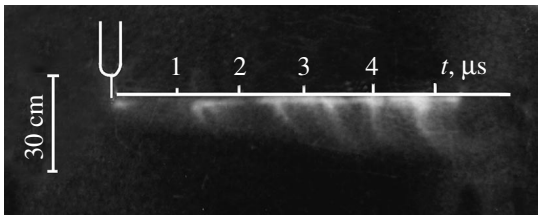

Fig. 6. Streak picture of the step development of a cathode-directed streamer in a 1 m long air gap at the leading edge of a voltage pulse with an average slope of about 100 kV/μs.

ionization is not interrupted, and its velocity changes little. The primary streamer runs away from the secondary wave after each successive restoration of the electrical conductivity of the plasma along the entire length of the already produced channel, and then in the next succession gradually slows down and stops in expectation of a new "restoration." Each such operation is accompanied by rapid, approximately two-fold, increase of plasma density in the existing channel (Fig. 7c), which gives the burst along the entire length of the channel observed in the photographs. The monotonic growth of the electron density between bursts results in background emission from the streamer, which is also noted in the streak pictures (Fig. 6). The process can continue indefinitely, but it is accompanied by a large loss of voltage because the field behind the streamer head is appreciably lower than $E_{th} \approx 20$ kV/cm only within the last step; then, right up to the anode, it remains high, to maintain the plasma in a conducting state.

The spatial length ΔL of a step depends directly on the voltage rise rate. According to the calculations, the streamer first stopped when its channel reached the length $\Delta L \approx 3.7$ cm, when the voltage increased with rate $A_i = 100$ kV/μs, and about 10 cm for $A_i = 330$ kV/μs. As a result, the rate of increase of the streamer channel averaged over many bursts increases, as is observed experimentally. If the streamer is indeed long ($L_s \gg \Delta L$), then with some underestimation the following estimates hold:

$$L_s(t) \approx \frac{U(t)}{E_{th}}, \quad v_s^{av} \approx \frac{A_i}{E_{th}}.$$

The nearly linear relation between the average velocity of the streamer and the slope of the leading edge of the voltage pulse is confirmed by measurements performed in [12].

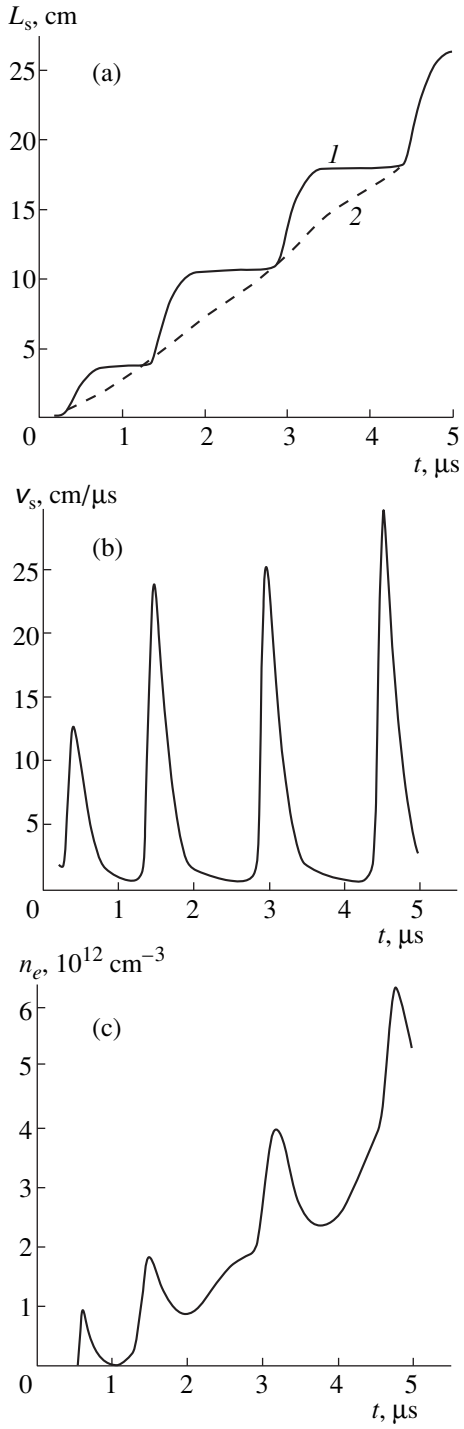

Fig. 7. Computed streamer characteristics in air at linearly rising voltage pulse with slope 100 kV/μs: (a) streamer length (*1*) and position of the front of the secondary wave of ionization (*2*); (b) streamer velocity; (c) electron density in the channel at a point 2.8 cm from the anode.

4. FEATURES OF THE STREAMER PROCESS WITH A HIGH VOLTAGE RISE RATE. RADIAL IONIZATION WAVE

The need to examine these features is obvious from the experimental data. The streak picture displayed in Fig. 8 for a pulse with slope $A_i \approx 3400$ kV/μs shows no traces of a secondary wave, though the streamer has advanced from the needle anode by 75 cm during the detection time. For half the slope the secondary wave is appreciable, but its velocity was approximately two times less than the velocity of the primary streamer.

The latter decreases little in time, so that there is not enough time for the secondary wave to overtake the streamer head during the detection time. The fact that for large values of A_i the primary streamer does not stop, but only slows down somewhat, also follows from the 1.5D simulation results (Fig. 9). The calculation with $A_i = 3400$ kV/μs gives a picture that resembles the streak picture in Fig. 8. The velocity of the primary streamer decreased here only to 8×10^7 cm/s, after which it once again started to increase. This is still not the usual deceleration of a streamer, but it is no longer a step process.

Even though the results outwardly seem similar, the model obviously differs from reality—the model continues to reproduce the secondary wave, while a secondary wave is not observed experimentally. This means that there is a reason why the loss of conductivity per unit length is retarded or completely precluded:

$$\gamma = eN_e\mu_e, \quad N_e = 2\pi\int_0^{r_s} rn_e(r)dr.$$

To maintain the conductivity per unit length at a prescribed level the total number of electrons N_e per unit length of the channel must be maintained at a prescribed level; a decrease of the electron density $n_e(r)$ at a single location in the cross section of the channel is of no significance. Consequently, not only the secondary longitudinal wave of ionization but also the transverse radial wave can maintain the conductivity per unit length. As numerical simulation showed [13, 16], the latter always exists in a streamer process. Actually, it determines the stationary streamer radius, whose upper limit can be estimated by the relation

$$r_s(x) \approx \frac{U(x)}{E_i \ln[L_s/r_s(x)]},$$

where E_i is the threshold field maintaining the ionization, $U(x)$ is the potential of the channel at the point x, and L_s is the length of the streamer. In order for the radial wave to compensate the decrease in the conductivity per unit length of the channel, it must develop with a completely definite rate during the entire streamer propagation time. The initial radius of the channel section under consideration, which was formed with a much lower voltage, is of no special significance, because (if the radial expansion process is intense) it will still subsequently be exceeded manifold. The initial values of the expansion rate of the channel and the density of the plasma reproduced depend on the initial radius, but the influence disappears with time.

This circumstance makes it possible to avoid two-dimensional calculations, which have not yet been performed, and to stay within a one-dimensional model, controlling the course of the process leading to expansion in the radial direction for a fixed location in the channel. To investigate the properties of the radial ionization wave the potential $U_x(t)$ was introduced for a given location of the streamer channel as an initial parameter in such a model. The kinetic scheme of the radial-wave model was the same as for a longitudinal wave of ionization. To a first approximation it was assumed that at the moment a new section of the channel is produced the streamer head imparts to it a potential $U_0 = 50$ kV, which then varies in time similarly to the voltage on the gap, in our case linearly

$$U_x(t) = U_0 + A_x t.$$

It was also assumed that the charge per unit length for a channel of length L_s and radius r_s is determined uniquely by its average capacitance per unit length and the potential:

$$\tau_x(t) = C_t U_x(t).$$

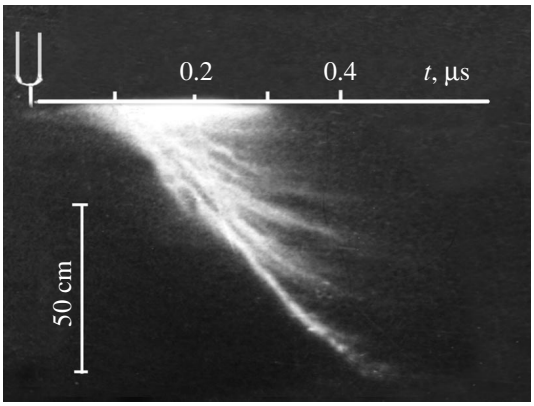

Fig. 8. Streak picture of a streamer burst obtained by using a slit in air with linearly growing voltage pulse with slope 3400 kV/μs.

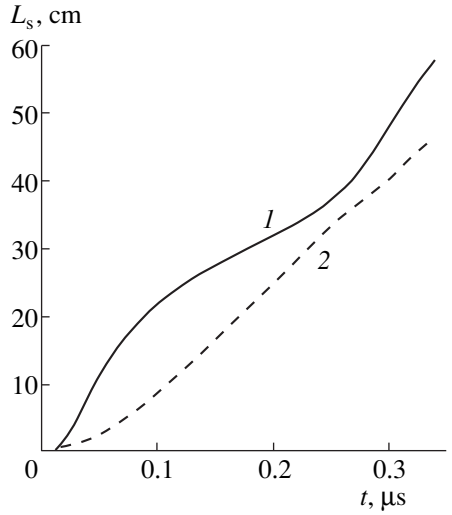

Fig. 9. Computed time variation of the streamer length (*1*) and position of the secondary wave of ionization (*2*) in air at linearly growing voltage pulse with slope 3400 kV/μs.

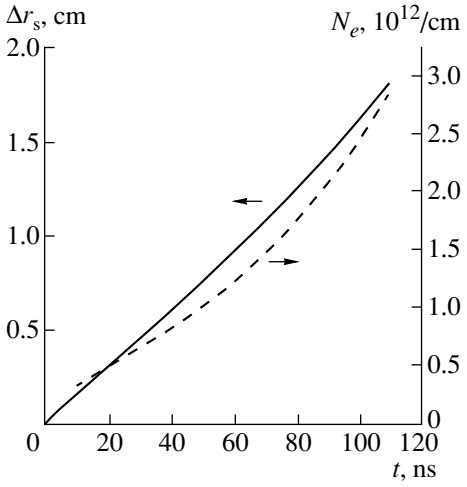

Fig. 10. Time variation of the radius of the streamer channel and electron density per unit length with ionization expansion of a channel in air for a linearly growing voltage pulse with slope 2000 kV/μs.

The dynamics of the variation of the per unit length conductivity of the expanding channel, taking account of the production and loss of electrons, was monitored for voltage pulses of different steepness. The estimates presented below were performed with channel length $L_s = 40$ cm.

It was found that a clear threshold effect is characteristic for ionization expansion. As long as the growth rate of the potential at a given location of the channel is less than 1700–1800 kV/μs the radial expansion time of the channel does not exceed 10–20 ns. Subsequently, the ionization wave cannot exist because the field on its cylindrical front decreases to the threshold value E_i. Such a short-time additional electron production in a radial field adds little to the previously produced electrons in the streamer head. The role of radial expansion in maintaining the per unit length conductivity of a streamer is negligible. The situation changes radically when the growth rate of the potential increases to $A_x \approx$ 2000 kV/μs. Now the weakening of the field on the front of the cylindrical wave with increasing radius,

$$E_r \approx \frac{U_x(t)}{r_s \ln(L_s/r_s)},$$

can be compensated by the linearly growing potential $U_x(t)$. This is what guarantees the long-time nature of the process. It is evident from Fig. 10 that in a time of about 100 ns, when the potential U_x increases to 250 kV, the radius of the conducting channel along which the streamer current is transported increases from 0.1 cm to 1.6 cm. About 2×10^{12} electrons per centimeter of channel length remains in the channel only as a result of ionization in a radial field 100 ns after the production of the channel section under study (Fig. 10). This is even somewhat greater than the number of electrons produced initially on the ionization front in the streamer head. Hence, the conductivity per unit length of the streamer does not decrease as long as the ionization expansion of the channel remains and the longitudinal electric field in the channel does not increase. This eliminates the reason for the creation of a secondary wave of ionization. A secondary wave is indeed not observed experimentally when the rate of growth of the voltage across the gap is high (Fig. 8).

A change in the mechanism maintaining the conductivity in the channel behind the streamer head has a beneficial effect on the development of long streamers in a sharply nonuniform field. To maintain their growth under the conditions considered it is possible to get away with a much lower axial field than the threshold field determined above (~20 kV/cm). Thus, in an experiment with $A_i = 3400$ kV/μs, by the time the length of the streamer channel has reached 75 cm the voltage across the gap has increased to 600 kV. Even if the head potential is neglected and it is assumed that the voltage in the channel is entirely lost, then the average field there will not exceed 8 kV/cm.

Although the ionization expansion in a radial field is a real means for maintaining the streamer channel in a conducting state with a high rate of rise of the voltage across the gap, there is no need to count on this effect for very long streamers. As the distance from the anode increases, the growth rate of the potential at a specific location of the channel, $A_x(x)$, becomes increasingly less than the slope A_i because of the voltage drop across it. Sooner or later the value of A_x will drop below the limiting value for which ionization expansion can still be maintained, and this factor will no longer operate. Then, once again, there will remain only one means for the streamer to remain in a conducting state for a long time—a secondary longitudinal wave of ionization, which requires a field of about 20 kV/cm. It has not been ruled out that now it will start not from the surface of the anode but rather from deep inside the gap at a location of the channel where ionization expansion stopped first (or does not start at all).

It should be noted that everything examined above is valid if the propagation of streamers does not substantially change the distribution of the electric field inside the discharge gap. This is undoubtedly correct for a single streamer, but it may not be true for a powerful, strongly branched, streamer burst, for example, such as a burst that develops from an anode with a radius of tens of cm with an initial voltage of several megavolts. Here the space charge of the many simultaneously forming streamers, which is comparable to the anode charge, displaces the field from the near-anode region, moving it in the direction of the cathode. Streamers in the region of a field intensified in this manner can advance even with complete loss of galvanic coupling with the location where they started; this is facilitated by the polarization of "young" channel sections which have still not lost their conductivity [3]. It is

known from model investigations and experiments that to maintain such a propagation regime the external field must not be less than the average axial field in the streamer channel, which in air under normal conditions is approximately 5 kV/cm [3, 13]. The analysis of the question of the collective development of streamers is beyond the scope of the present paper.

5. POSSIBILITY OF OBSERVING STEP MOTION OF A STREAMER IN VARIOUS GASES

The choice of air as a medium for simulation was dictated only by the fact that the results can be more easily checked experimentally. Essentially, the gas composition is not fundamentally important. An appreciable decrease in the conductivity of a streamer plasma is characteristic not only of electronegative but also electropositive gases. The difference between them is purely quantitative, because sooner or later any streamer channel will appreciably lose conductivity. The only question is the length and development time of the channel. But, the stronger the electronegative properties of the medium are, the shorter the streamer length and the shorter the times at which the effects being studied here will appear.

To confirm this a long streamer with increasing voltage in heated air at atmospheric pressure as well as in oxygen and SF_6 with $p = 1$ atm and $T = 300$ K was simulated. All geometric parameters were the same as before. As is well known, the effective rate of the loss of electrons in a streamer channel decreases substantially as the air is heated, because the three-particle attachment of electrons to oxygen molecules slows down, electron attachment is partially compensated by detachment of electrons from negative ions, and the electron-ion recombination slows down because of the destruction of the complex positive ions [14]. Conversely, the rate of the loss of electrons in oxygen and in the strongly electronegative gas SF_6 is much higher than in air under normal conditions because of the intense attachment of electrons to molecules. The results of the simulation showed (Fig. 11) that heating air up to 600 K at a voltage pulse with the same slope of 100 kV/μs increased the streamer length, formed in the first step, by a factor of 6, so that the length reached 23 cm, while the duration of the step approximately doubled. The simulation in pure oxygen, as expected, resulted in the opposite effect. In reality step development appeared only with a voltage rise rate of the order of 1000 kV/μs and higher. It is evident from the computational data in Fig. 12 that for such a slope of the leading edge of the pulse the step length is only 1.5–2 cm, and the development time of the step decreases to 40–50 ns. However, if the slope of the front is decreased to 500 kV/cm, then there is virtually no time for a step to form. The streamer advances continuously, and the primary and secondary waves of ionization move along the gap with virtually the same velocities, leaving behind them a channel with a field of

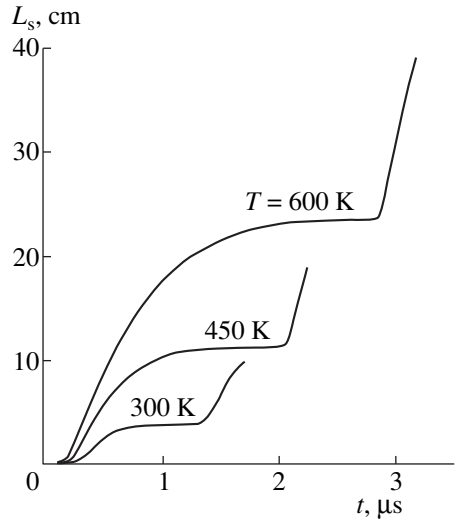

Fig. 11. Time variation of the streamer length in air at various temperatures for a linearly growing voltage pulse with slope 100 kV/μs.

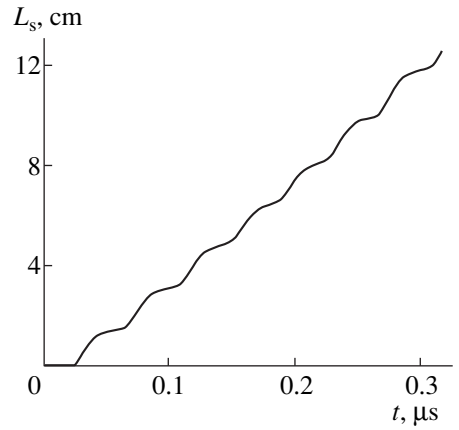

Fig. 12. Time variation of the streamer length in pure oxygen at a linearly growing voltage pulse with slope 1000 kV/μs.

about 25 kV/cm. With such a field the electron density in the channel is maintained almost at a constant level.

Finally, in SF_6 the exceedingly strong attachment completely eliminated the possibility of step development. The streamer plasma decayed so rapidly that no appreciable channel section with a longitudinal field below the threshold value ($E_i/N \approx 3.6 \times 10^{-15}$ V cm^2 in SF_6) could form behind the head. As a result, the steps almost completely degenerated. Only rapidly decaying oscillations of the velocity at the very start of the process remained (Fig. 13a). Subsequently, the streamer length increased uniformly, and the longitudinal field was about 90 kV/cm along the entire length of the channel (Fig. 13). The decrease in the field was appreciable only immediately behind the head, on a channel section

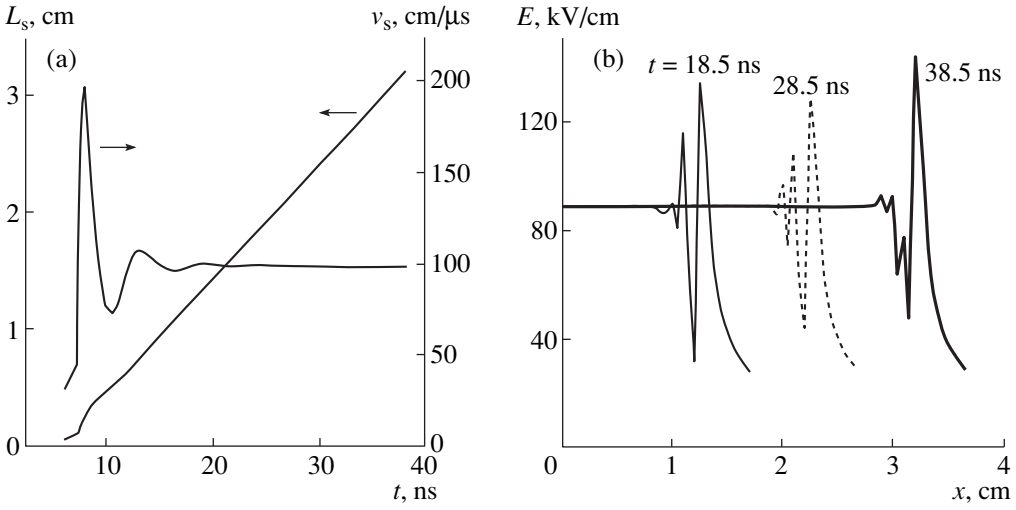

Fig. 13. Time variation of the streamer length and streamer velocity in SF_6 at a linearly growing voltage pulse with slope 10^4 kV/μs.

equal in length to several radii of the section (Fig. 13b). Increasing the rate of growth of the voltage up to 10^4 kV/μs did not change the course of the process. Such continuous advancement of the streamer in SF_6 was observed experimentally and confirmed by numerical calculations in [17] for a voltage pulse with a 15 ns leading edge.

6. CONCLUSIONS

We shall now formulate the basic results of this work. It was shown theoretically and experimentally that with increasing voltage the streamer can propagate in a gas gap by a step mechanism. The mechanism of this phenomenon is explained as follows. A streamer moving in a sharply nonuniform field slows down and stops as a result of a large loss of voltage on the decaying channel. The "revival" of the streamer is due to the appearance of a new wave of ionization, caused by redistribution of the electric field in the channel because of the nonuniform decay of the plasma, near the electrode. Moving along the streamer channel, the new wave increases the conductivity of the channel and the head potential, thereby creating conditions for further advancement of the streamer. The streamer rapidly moves away from the secondary wave of ionization and after a certain time, comparable to the lifetime of the plasma in the channel, once again stops because of the decay of the channel and, consequently, the breakdown of the galvanic coupling between the streamer head and the high-voltage electrode. The secondary wave of ionization, reaching the weakly conducting section of the channel, restores the conductivity of the channel and a new step in the motion of the streamer starts.

The longitudinal electric field in the channel behind the front of the secondary wave is close to the threshold field E_i, in which an equilibrium between electron production and loss is established for the conditions under study. A field much less than E_i is characteristic only for the section between the fronts of the secondary and primary waves. The average velocity of the step motion of a streamer on a section multifold greater than the step length is directly proportional to the slope of the leading edge of the voltage pulse. For a high voltage rise rate, the conductivity of the streamer channel can be maintained as a result of ionization expansion of the channel in the radial field. In air under normal conditions this effect is actually noticeable with voltage rise rates exceeding 2000 kV/μs, as a result of which the steps are disappeared. In gases with strongly pronounced electronegative properties, such as SF_6, the length of the steps becomes comparable with the streamer radius for any actually attainable voltage rise rate across the discharge gap, as a result of which streamer propagation is seen to be continuous.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project no. 00-02-17399) and INTAS (grant no. 96-2120).

REFERENCES

1. J. M. Meek and J. D. Craggs, *Electrical Breakdown of Gases* (Wiley, New York, 1978).
2. Yu. P. Raizer, *Gas Discharge Physics* (Springer, Berlin, 1991).
3. E. M. Bazelyan and Yu. P. Raizer, *Spark Discharge* (Chemical Rubber Company, New York, 1998).
4. B. N. Gorin and A. V. Shkilev, Élektrichestvo, No. 6, 31 (1976).

5. D. Dawson and W. Winn, in *Electron Avalanches and Breakdown in Gases* (Butterworths, London, 1964; Mir, Moscow, 1968].
6. E. Marode, J. Appl. Phys. **46**, 2005 (1975); **46**, 2016 (1975).
7. R. S. Sigmond, J. Appl. Phys. **56**, 1355 (1984).
8. M. Aints, A. Haljaste, and K. Kudu, in *Proceedings of the 6th International Conference on Gas Discharges* (London, 1978); IEE Conf. Publ., No. 189, 265 (1978).
9. N. L. Aleksandrov, A. É. Bazelyan, É. M. Bazelyan, and I. V. Kochetov, Fiz. Plazmy **21**, 60 (1995) [Plasma Phys. Rep. **21**, 57 (1995)].
10. M. I. D'yakonov and V. Yu. Kachorovskiĭ, Zh. Éksp. Teor. Fiz. **95**, 1850 (1989) [Sov. Phys. JETP **68**, 1070 (1989)].
11. M. I. D'yakonov and V. Yu. Kachorovskiĭ, Zh. Éksp. Teor. Fiz. **98**, 895 (1990) [Sov. Phys. JETP **71**, 498 (1990)].
12. É. M. Bazelyan and A. Yu. Goryunov, Élektrichestvo, No. 11, 27 (1986).
13. N. L. Aleksandrov and E. M. Bazelyan, J. Phys. D **29**, 740 (1996).
14. N. L. Aleksandrov and E. M. Bazelyan, J. Phys. D **29**, 2873 (1996).
15. R. Morrow, IEEE Trans. Plasma Sci. PS-14, 234 (1986).
16. N. L. Aleksandrov and É. E. Bazelyan, Fiz. Plazmy **22**, 458 (1996) [Plasma Phys. Rep. **22**, 417 (1996)].
17. R. Morrow, IEEE Trans. Plasma Sci. **19**, 86 (1991).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 4, 2000, pp. 736–747.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 4, 2000, pp. 849–862.

PLASMA, GASES

Stochastic Transport in Random Wave Fields

V. I. Klyatskin[a, b, *] and I. G. Yakushkin[a, **]

[a]*Oboukhov Institute of Atmospheric Physics, Russian Academy of Sciences, Moscow, 109017 Russia*
[b]*Pacific Institute of Oceanology, Far Eastern Division, Russian Academy of Sciences, Vladivostok, 690041 Russia*
**e-mail: klyatskin@hotmail.com*
***e-mail: yakushk@omega.ifaran.ru*
Received April 4, 2000

Abstract—An analysis is made of particle diffusion and the field of a passive impurity in random wave fields. A characteristic of this problem is that the statistical transport coefficients (diffusion coefficients) vanish in the approximations normally used (delta-correlated random field or diffusion) giving the Fokker–Planck equation. In this study perturbation theory is used in the first nonvanishing order of smallness which allows these transport coefficients to be calculated for waves of various types. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION. CHARACTERISTICS OF PARTICLE DIFFUSION IN RANDOM WAVE FIELDS OF VELOCITY AND EXTERNAL FORCES

Particle motion in rapidly varying random velocity fields or under the action of rapidly varying random forces is an important problem having numerous applications in mechanics, hydrodynamics, plasma physics, and so on. It is well known that stochastic transport in rapidly-varying vibrational and wave fields leads to various important physical phenomena such as Fermi acceleration, stochastic plasma heating, and so on [1, 2]. These phenomena are generally described using the Fokker–Planck equation whose coefficients are expressed in terms of the correlation functions of random fields and are calculated using methods of averages developed for nonlinear equations. Although the results thus obtained reflect the main features of these phenomena, they do not have a universal character or a clearly defined range of validity.

At the same time we know that a broad class of problems can be described fairly comprehensively using the approximation of a delta-correlated random process, the diffusion approximation, or various generalizations of these based on a functional technique with variational derivatives (see, e.g., [3–7]). Calculations of the transport coefficients for random wave fields indicate that in many cases these may have values of the second order of smallness relative to the values appearing in ordinary variants of the theory of short-correlated random fields. It is therefore of considerable interest to develop a general method of deriving equations for the statistical characteristics of stochastic particle transport and fields and to calculate the transport coefficients taking into account terms of the second order of smallness if the coefficients of the first order of smallness vanish.

A general method for such calculations was proposed in [8] (see, also [3, 9]). This method can be used to analyze a range of phenomena described by different authors using different approaches from a common viewpoint, it can be used to calculate the statistical characteristics of particle ensembles and fields, and also to indicate the ranges of validity of the equations obtained.

Particle diffusion in the random velocity field $\mathbf{u}(\mathbf{r}, t)$ is usually described using the first-order differential equation

$$\frac{d}{dt}\mathbf{r}(t) = \mathbf{u}(\mathbf{r}, t), \quad \mathbf{r}(0) = \mathbf{r}_0. \tag{1}$$

Particle diffusion in a field of random external forces $\mathbf{f}(\mathbf{r}, t)$ with linear friction is described by the system of equations

$$\frac{d}{dt}\mathbf{r}(t) = \mathbf{v}(t), \quad \frac{d}{dt}\mathbf{v}(t) = -\lambda\mathbf{v}(t) + \mathbf{f}(\mathbf{r}, t), \qquad \mathbf{r}(0) = \mathbf{r}_0, \quad \mathbf{v}(0) = \mathbf{v}_0. \tag{2}$$

We introduce some indicator functions for Eqs. (1) and (2)

$$\varphi(\mathbf{r}, t) = \delta(\mathbf{r}(t) - \mathbf{r}), \qquad \varphi(\mathbf{r}, \mathbf{v}, t) = \delta(\mathbf{r}(t) - \mathbf{r})\delta(\mathbf{v}(t) - \mathrm{v}), \tag{3}$$

which are described by the Liouville equations (see, e.g., [3])

$$\frac{\partial}{\partial t}\varphi(\mathbf{r}, t) = -\frac{\partial}{\partial \mathbf{r}}\{\mathbf{u}(\mathbf{r}, t)\varphi(\mathbf{r}, t)\}, \qquad \varphi(\mathbf{r}, 0) = \delta(\mathbf{r} - \mathbf{r}_0),$$
$$\left(\frac{\partial}{\partial t} + \mathbf{v}\frac{\partial}{\partial \mathbf{r}} - \lambda\frac{\partial}{\partial \mathbf{v}}\mathbf{v}\right)\varphi(\mathbf{r}, \mathbf{v}, t) = -\mathbf{f}(\mathbf{r}, t)\frac{\partial}{\partial \mathbf{v}}\varphi(\mathbf{r}, \mathbf{v}, t), \qquad \varphi(\mathbf{r}, \mathbf{v}, 0) = \delta(\mathbf{r} - \mathbf{r}_0)\delta(\mathbf{v} - \mathbf{v}_0). \tag{4}$$

1063-7761/00/9104-0736$20.00

The average of the indicator function $\varphi(\mathbf{r}, t)$ over the ensemble of realizations of the random field $\{\mathbf{u}(\mathbf{r}, t)\}$ will then describe the single-point probability density of the particle position,

$$P(\mathbf{r}, t) = \langle \varphi(\mathbf{r}, t)\rangle_{\mathbf{u}} = \langle \delta(\mathbf{r}(t) - \mathbf{r})\rangle_{\mathbf{u}},$$

and the average of the indicator function $\varphi(\mathbf{r}, \mathbf{v}, t)$ over the ensemble of realization of the random field $\{\mathbf{f}(\mathbf{r}, t)\}$ will describe the joint single-point probability density of the particle position and its velocity

$$P(\mathbf{r}, \mathbf{v}, t) = \langle \varphi(\mathbf{r}, \mathbf{v}, t)\rangle_{\mathbf{f}} = \langle \delta(\mathbf{r}(t) - \mathbf{r})\delta(\mathbf{v}(t) - \mathbf{v})\rangle_{\mathbf{f}},$$

We shall average equation (4) over the ensemble of realizations of the random fields $\{\mathbf{u}(\mathbf{r}, t)\}$ and $\{\mathbf{f}(\mathbf{r}, t)\}$. As a result we obtain the open equations

$$\begin{aligned}
\frac{\partial}{\partial t}P(\mathbf{r}, t) &= -\frac{\partial}{\partial \mathbf{r}}\langle \mathbf{u}(\mathbf{r}, t)\varphi(\mathbf{r}, t)\rangle,\\
P(\mathbf{r}, 0) &= \delta(\mathbf{r} - \mathbf{r}_0),\\
\left(\frac{\partial}{\partial t} + \mathbf{v}\frac{\partial}{\partial \mathbf{r}} - \lambda\frac{\partial}{\partial \mathbf{v}}\mathbf{v}\right)P(\mathbf{r}, \mathbf{v}, t) &\\
= -\frac{\partial}{\partial \mathbf{v}}\langle \mathbf{f}(\mathbf{r}, t)\varphi(\mathbf{r}, \mathbf{v}, t)\rangle,&\\
P(\mathbf{r}, \mathbf{v}, 0) &= \delta(\mathbf{r} - \mathbf{r}_0)\delta(\mathbf{v} - \mathbf{v}_0),
\end{aligned} \tag{5}$$

containing the correlations $\langle \mathbf{u}(\mathbf{r}, t)\varphi(\mathbf{r}, t)\rangle$ and $\langle \mathbf{f}(\mathbf{r}, t)\varphi(\mathbf{r}, \mathbf{v}, t)\rangle$. We shall assume that the fields $\mathbf{u}(\mathbf{r}, t)$ and $\mathbf{f}(\mathbf{r}, t)$ are Gaussian random fields, spatially uniform and steady-state in time having zero averages and correlation tensors

$$\begin{aligned}
B_{ij}^{(\mathbf{u})}(\mathbf{r} - \mathbf{r}', t - t') &= \langle u_i(\mathbf{r}, t)u_j(\mathbf{r}', t')\rangle,\\
B_{ij}^{(\mathbf{f})}(\mathbf{r} - \mathbf{r}', t - t') &= \langle f_i(\mathbf{r}, t)f_j(\mathbf{r}', t')\rangle.
\end{aligned} \tag{6}$$

The correlations can then be split using the Furutsu–Novikov formula (see, e.g., [3,6,10])

$$\begin{aligned}
&\langle f_k(\mathbf{x}, t)R[t; \mathbf{f}(\mathbf{y}, \tau)]\rangle\\
&= \int d\mathbf{x}'\int dt' B_{kl}(\mathbf{x}, t; \mathbf{x}', t')\left\langle \frac{\delta}{\delta f_l(\mathbf{x}', t')}R[t; \mathbf{f}(\mathbf{y}, \tau)]\right\rangle,
\end{aligned} \tag{7}$$

which holds for the Gaussian random field $\mathbf{f}(\mathbf{x}, t)$ with the arbitrary functional $R[t; \mathbf{f}(\mathbf{x}, \tau)]$ of it. Consequently, Eqs. (5) may be rewritten in the form

$$\begin{aligned}
\frac{\partial}{\partial t}P(\mathbf{r}, t) &= -\frac{\partial}{\partial r_i}\int d\mathbf{r}'\int_0^t dt' B_{ij}^{(\mathbf{u})}(\mathbf{r} - \mathbf{r}', t - t')\\
&\times\left\langle \frac{\delta}{\delta u_j(\mathbf{r}', t')}\varphi(\mathbf{r}, t)\right\rangle_{\mathbf{u}},\\
P(\mathbf{r}, 0) &= \delta(\mathbf{r} - \mathbf{r}_0),\\
\left(\frac{\partial}{\partial t} + \mathbf{v}\frac{\partial}{\partial \mathbf{r}} - \lambda\frac{\partial}{\partial \mathbf{v}}\mathbf{v}\right)P(\mathbf{r}, \mathbf{v}, t) &\\
= \frac{\partial}{\partial v_i}\int d\mathbf{r}'\int_0^t dt' B_{ij}^{(\mathbf{f})}(\mathbf{r} - \mathbf{r}', t - t')&\\
\times\left\langle \left[\frac{\partial}{\partial r_k}\frac{\delta r_k(t)}{\delta f_j(\mathbf{r}', t')} + \frac{\partial}{\partial v_k}\frac{\partial v_k(t)}{\delta f_j(\mathbf{r}', t')}\right]\varphi(\mathbf{r}, \mathbf{v}, t)\right\rangle_{\mathbf{f}},&\\
P(\mathbf{r}, \mathbf{v}, 0) &= \delta(\mathbf{r} - \mathbf{r}_0)\delta(\mathbf{v} - \mathbf{v}_0).
\end{aligned} \tag{8}$$

We subsequently need to use asymptotic methods to obtain closed equations. The simplest of these methods are the approximation that the random fields $\mathbf{u}(\mathbf{r}, t)$ and $\mathbf{f}(\mathbf{r}, t)$ are delta-correlated in time, and the diffusion approximation.

1.1. Delta-Correlated Approximation

In the approximation that the random fields $\mathbf{u}(\mathbf{r}, t)$ and $\mathbf{f}(\mathbf{r}, t)$ are delta correlated in time, the correlation tensors (6) are approximated by the expressions

$$B_{ij}(\mathbf{r}, t) = 2B_{ij}^{\text{eff}}(\mathbf{r})\delta(t - t'), \tag{9}$$

where

$$B_{ij}^{\text{eff}}(\mathbf{r}) = \frac{1}{2}\int_{-\infty}^{\infty} dt B_{ij}(\mathbf{r}, t) = \int_0^{\infty} dt B_{ij}(\mathbf{r}, t). \tag{10}$$

Now taking into account the equalities

$$\begin{aligned}
\left.\frac{\delta\varphi(\mathbf{r}, t)}{\delta u_j(\mathbf{r}', t')}\right|_{t = t'} &= -\frac{\partial}{\partial r_j}\{\delta(\mathbf{r} - \mathbf{r}')\varphi(\mathbf{r}, t')\},\\
\left.\frac{\delta r_k(t)}{\delta f_j(\mathbf{r}', t')}\right|_{t = t'} &= 0,\\
\left.\frac{\delta v_k(t)}{\delta f_j(\mathbf{r}', t')}\right|_{t = t'} &= \delta_{kj}\delta(\mathbf{r}(t') - \mathbf{r}'),
\end{aligned} \tag{11}$$

derived from (4) and (2), Eqs. (8) may be rewritten in a closed form corresponding to the Fokker–Planck equation

$$\begin{aligned}
\frac{\partial}{\partial t}P(\mathbf{r}, t) &= D_{ij}^{(\mathbf{u})}\frac{\partial^2}{\partial r_i \partial r_j}P(\mathbf{r}, t),\\
P(\mathbf{r}, 0) &= \delta(\mathbf{r} - \mathbf{r}_0),
\end{aligned} \tag{12}$$

$$\begin{aligned}
&\left(\frac{\partial}{\partial t} + \mathbf{v}\frac{\partial}{\partial \mathbf{r}} - \lambda\frac{\partial}{\partial \mathbf{v}}\mathbf{v}\right)P(\mathbf{r}, \mathbf{v}, t)\\
&= D_{ij}^{(\mathbf{f})}\frac{\partial^2}{\partial v_i \partial v_j}P(\mathbf{r}, \mathbf{v}, t),\\
&P(\mathbf{r}, \mathbf{v}, 0) = \delta(\mathbf{r} - \mathbf{r}_0)\delta(\mathbf{v} - \mathbf{v}_0),
\end{aligned} \tag{13}$$

with the diffusion tensors

$$D_{ij}^{(\mathbf{u})} = \int_0^\infty d\tau B_{ij}^{(\mathbf{u})}(0, \tau),$$
$$D_{ij}^{(\mathbf{f})} = \int_0^\infty d\tau B_{ij}^{(\mathbf{f})}(0, \tau). \quad (14)$$

In this case, the solutions of the stochastic equations (1) and (2) are vector Markov processes whose transition probability density is also described by Eqs. (12) and (13). The condition for validity of the approximation that the random fields $\mathbf{u}(\mathbf{r}, t)$ and $\mathbf{f}(\mathbf{r}, t)$ are delta-correlated in time is evidently that the time correlation radius τ_0 of the random fields $\mathbf{u}(\mathbf{r}, t)$ and $\mathbf{f}(\mathbf{r}, t)$ is small compared with the time scales appearing in the problem, which requires corresponding smallness of the fluctuating parameters.

In general, integrating Eq. (13) over $\mathbf{r}$, we obtain a closed equation for the probability density of the particle velocity:

$$\left(\frac{\partial}{\partial t} - \lambda \frac{\partial}{\partial \mathbf{v}} \mathbf{v}\right) P(\mathbf{v}, t) = \frac{\partial}{\partial v_i} D_{ij}^{(\mathbf{f})} \frac{\partial}{\partial v_j} P(\mathbf{v}, t),$$
$$P(\mathbf{r}, \mathbf{v}, 0) = \delta(\mathbf{v} - \mathbf{v}_0), \quad (15)$$

which has a steady-state probability distribution described by

$$\lambda \frac{\partial}{\partial \mathbf{v}} \mathbf{v} P(\mathbf{v}) = -\frac{\partial}{\partial v_i} D_{ij}^{(\mathbf{f})} \frac{\partial}{\partial v_j} P(\mathbf{v}). \quad (16)$$

The rate of establishment of this distribution depends on the parameter λ. For a fairly high value of λ the steady-state distribution (16) is established rapidly and the time evolution of the probability density for the particle position will be described by Eq. (12) with the diffusion tensor

$$D_{ij}^{(\mathbf{u})} = \frac{1}{\lambda^2} \int_0^\infty d\tau B_{ij}^{(\mathbf{f})}(0, \tau),$$

which implies that this problem is statistically equivalent to the problem of particle diffusion in a random velocity field having the form

$$\mathbf{u}(\mathbf{r}, t) = \frac{1}{\lambda} \mathbf{f}(\mathbf{r}, t). \quad (17)$$

The transition from Eq. (13) to Eq. (12) forms the so-called Kramers problem (see, e.g., [11]).

1.2. Diffusion Approximation

Allowance for the finite time-correlation radius of the random fields $\mathbf{u}(\mathbf{r}, t)$ and $\mathbf{f}(\mathbf{r}, t)$ may be made using the diffusion approximation. This approximation is clearer and more physical than the formal mathematical approximation of a delta-correlated random field. It also holds for fairly small fluctuations of the parameters of a stochastic dynamic system and can be used to obtain not only the conditions of validity of the delta-correlated approximation but can also describe new physical effects generated by the finite time correlation radius of the random parameters. In the diffusion approximation it is also assumed that the influence of random actions on time scales of the order of τ_0 is negligible, i.e., on these scales the system evolves as a free system.

In the diffusion approximation the equations for the corresponding probability density (8) are exact. The corresponding simplifications of the problem are made at the level of the functional dependence of the solution of the problem on the fluctuating parameters. For instance, for the problem (1) in the diffusion approximation the indicator function and its variational derivative on time scales of the order of the correlation time radius τ_0 of the random field $\mathbf{u}(\mathbf{r}, t)$ are described by the system of dynamic equations

$$\frac{\partial}{\partial t} \frac{\delta \varphi(\mathbf{r}, t)}{\delta u_i(\mathbf{r}', t')} = 0,$$
$$\left.\frac{\delta \varphi(\mathbf{r}, t)}{\delta u_i(\mathbf{r}', t')}\right|_{t = t'} = -\frac{\partial}{\partial r_i} \{\delta(\mathbf{r} - \mathbf{r}') \varphi(\mathbf{r}, t')\}, \quad (18)$$
$$\frac{\partial}{\partial t} \varphi(\mathbf{r}, t) = 0, \quad \varphi(\mathbf{r}, t)|_{t = t'} = \varphi(\mathbf{r}, t').$$

Substituting the system (18) into the first of Eqs. (8), we obtain the equation

$$\frac{\partial}{\partial t} P(\mathbf{r}, t) = D_{ij}^{(\mathbf{u})}(t) \frac{\partial^2}{\partial r_i \partial r_j} P(\mathbf{r}, t),$$
$$P(\mathbf{r}, 0) = \delta(\mathbf{r} - \mathbf{r}_0), \quad (19)$$

where the diffusion tensor now has the form

$$D_{ij}^{(\mathbf{u})}(t) = \int_0^t d\tau B_{ij}^{(\mathbf{u})}(0, \tau). \quad (20)$$

Equation (19) with the diffusion tensor (20) holds for all times t. In this case, however, the solution of the problem (1) $\mathbf{r}(t)$ is not a vector Markov random process since its multitime probability density does not allow factorization using the transition probability density. In the asymptotic case $t \gg \tau_0$ the solution of the initial dynamic system (1) in the diffusion approximation will be a Markov random process described by Eq. (12) with the diffusion coefficient (14), i.e., in this asymptotic case the diffusion approximation is the same as the approximation that the random field $\mathbf{u}(\mathbf{r}, t)$ is delta-correlated in time. This is evidently because of the absence of an average flux in Eq. (1).

Similarly, for the dynamic problem (2) the variational derivatives of the functions $\mathbf{r}(t)$ and $\mathbf{v}(t)$ contained in Eq. (8) are described in the diffusion approximation on scales of the order of the time correlation

radius τ_0 of the random field $\mathbf{f}(\mathbf{r}, t)$ by a deterministic system of equations derived from (2) for $t' < t$,

$$\frac{d}{dt}\frac{\delta r_k(t)}{\delta f_j(\mathbf{r}', t')} = \frac{\delta v_k(t)}{\delta f_j(\mathbf{r}', t')},$$
$$\frac{d}{dt}\frac{\delta v_k(t)}{\delta f_j(\mathbf{r}', t')} = -\lambda\frac{\delta v_k(t)}{\delta f_j(\mathbf{r}', t')}, \tag{21}$$

for which the initial conditions

$$\left.\frac{\delta r_k(t)}{\delta f_j(\mathbf{r}', t')}\right|_{t=t'} = 0,$$
$$\left.\frac{\delta v_k(t)}{\delta f_j(\mathbf{r}', t')}\right|_{t=t'} = \delta_{kj}\delta(\mathbf{r}(t') - \mathbf{r}') \tag{22}$$

however contain randomnesses in the form of the stochastic function $\mathbf{r}(t')$.

The solution of the system (21) with the initial conditions (22) has the form

$$\frac{\delta v_k(t)}{\delta f_j(\mathbf{r}', t')} = \delta_{kj}e^{-\lambda(t-t')}\delta(\mathbf{r}(t') - \mathbf{r}'),$$
$$\frac{\delta r_k(t)}{\delta f_j(\mathbf{r}', t')} = \frac{1}{\lambda}\delta_{kj}[1 - e^{-\lambda(t-t')}]\delta(\mathbf{r}(t') - \mathbf{r}'). \tag{23}$$

Subsequently assuming that on these time scales the action of random forces is negligible even for the particle dynamics, we can replace the value of $\mathbf{r}(t')$ appearing in (23) with $\mathbf{r}(t)$ using a simplified system of Eqs. (2),

$$\frac{d}{dt}\mathbf{r}(t) = \mathbf{v}(t), \quad \frac{d}{dt}\mathbf{v}(t) = -\lambda\mathbf{v}(t) \tag{24}$$

with the initial conditions

$$\mathbf{r}(t)|_{t=t'} = \mathbf{r}(t'), \quad \mathbf{v}(t)|_{t=t'} = \mathbf{v}(t'), \tag{25}$$

from which it follows that

$$\mathbf{r}(t') = \mathbf{r}(t) = \frac{1}{\lambda}\mathbf{v}[e^{\lambda(t-t')} - 1],$$
$$\mathbf{v}(t') = e^{\lambda(t-t')}\mathbf{v}(t). \tag{26}$$

Using Eqs. (23) and (26), we can now write Eq. (8) in the closed form [9]:

$$\left(\frac{\partial}{\partial t} + \mathbf{v}\frac{\partial}{\partial \mathbf{r}} - \lambda\frac{\partial}{\partial \mathbf{v}}\mathbf{v}\right)P(\mathbf{r}, \mathbf{v}, t)$$
$$= \frac{\partial}{\partial v_i}\left\{D_{ij}^{(1)}(\mathbf{v}, t)\frac{\partial}{\partial v_j} + D_{ij}^{(2)}(\mathbf{v}, t)\frac{\partial}{\partial r_j}\right\}P(\mathbf{r}, \mathbf{v}, t), \tag{27}$$

where we introduce the diffusion coefficients

$$D_{ij}^{(1)}(\mathbf{v}, t) = \int_0^t d\tau e^{-\lambda\tau}B_{ij}^{(\mathbf{f})}\left(\frac{1}{\lambda}[e^{\lambda\tau} - 1]\mathbf{v}, \tau\right),$$
$$D_{ij}^{(2)}(\mathbf{v}, t) = \frac{1}{\lambda}\int_0^t d\tau[1 - e^{-\lambda\tau}]B_{ij}^{(\mathbf{f})}\left(\frac{1}{\lambda}[e^{\lambda\tau} - 1]\mathbf{v}, \tau\right). \tag{28}$$

Equation (28) correctly describes the single-point probability density for times $t < \tau_0$. In this case, however, the solution of the problem (2) $\{\mathbf{r}(t), \mathbf{v}(t)\}$ is not a vector Markov random process since its multitime probability density cannot be factorized using the transition probability density. In the asymptotic case $t \gg \tau_0$ the solution of the initial dynamic system (2) will be a Markov random process. In this case, the upper limits in the integrals (28) may be replaced by infinity. We then obtain the Fokker–Planck equation for the single-time probability density

$$\left(\frac{\partial}{\partial t} + \mathbf{v}\frac{\partial}{\partial \mathbf{r}} - \lambda\frac{\partial}{\partial \mathbf{v}}\mathbf{v}\right)P(\mathbf{r}, \mathbf{v}, t)$$
$$= \frac{\partial}{\partial v_i}\left\{D_{ij}^{(1)}(\mathbf{v})\frac{\partial}{\partial v_j} + D_{ij}^{(2)}(\mathbf{v})\frac{\partial}{\partial r_j}\right\}P(\mathbf{r}, \mathbf{v}, t), \tag{29}$$
$$P(\mathbf{r}, \mathbf{v}, 0) = \delta(\mathbf{r} - \mathbf{r}_0)\delta(\mathbf{v} - \mathbf{v}_0),$$

with the diffusion coefficients

$$D_{ij}^{(1)}(\mathbf{v}) = \int_0^\infty d\tau e^{-\lambda\tau}B_{ij}^{(\mathbf{f})}\left(\frac{1}{\lambda}[e^{\lambda\tau} - 1]\mathbf{v}, \tau\right),$$
$$D_{ij}^{(2)}(\mathbf{v}) = \frac{1}{\lambda}\int_0^\infty d\tau[1 - e^{-\lambda\tau}]B_{ij}^{(\mathbf{f})}\left(\frac{1}{\lambda}[e^{\lambda\tau} - 1]\mathbf{v}, \tau\right). \tag{30}$$

We note that the approximation of a delta-correlated random field [i.e., Eq. (12)] corresponds to Eq. (27) with the diffusion coefficients

$$D_{ij}^{(1)}(\mathbf{v}) = \int_0^\infty d\tau B_{ij}^{(\mathbf{f})}(0, \tau), \quad D_{ij}^{(2)}(\mathbf{v}) = 0.$$

In the absence of friction, Eq. (29) is simplified and has the form [1]

$$\left(\frac{\partial}{\partial t} + \mathbf{v}\frac{\partial}{\partial \mathbf{r}}\right)P(\mathbf{r}, \mathbf{v}, t)$$
$$= \frac{\partial}{\partial v_i}\left\{D_{ij}^{(1)}(\mathbf{v})\frac{\partial}{\partial v_j} + D_{ij}^{(2)}(\mathbf{v})\frac{\partial}{\partial r_j}\right\}P(\mathbf{r}, \mathbf{v}, t), \tag{31}$$
$$P(\mathbf{r}, \mathbf{v}, 0) = \delta(\mathbf{r} - \mathbf{r}_0)\delta(\mathbf{v} - \mathbf{v}_0)$$

with the diffusion coefficients

$$D_{ij}^{(1)}(\mathbf{v}) = \int_0^\infty d\tau B_{ij}^{(\mathbf{f})}(\mathbf{v}\tau, \tau),$$
$$D_{ij}^{(2)}(\mathbf{v}) = \int_0^\infty \tau d\tau B_{ij}^{(\mathbf{f})}(\mathbf{v}\tau, \tau). \quad (32)$$

For a fairly high value of the parameter λ this problem is similar to the Kramers problem for the delta-correlated approximation of the random force field $\mathbf{f}(\mathbf{r}, t)$. In this case, the probability density of the particle position will be described by Eq. (19) which is statistically equivalent to a transition to the dynamic equation (1) with the velocity field (17).

We introduce the new field $\tilde{\mathbf{u}}(\mathbf{r}, t)$ with unit dispersion such that

$$\mathbf{u}(\mathbf{r}, t) = \sigma_{\mathbf{u}}\tilde{\mathbf{u}}(\mathbf{r}, t),$$

where the dispersion of the velocity field is

$$\sigma_{\mathbf{u}}^2 = B_{ii}^{(\mathbf{u})}(0, 0).$$

We shall assume that this random field has a wave origin and thus its correlation tensor has the structure

$$B_{ij}^{(\mathbf{u})}(\mathbf{r}, t) = \int d\mathbf{k} F_{ij}^{(\mathbf{u})}(\mathbf{k}) \cos\{\mathbf{k}\cdot\mathbf{r} - \omega(\mathbf{k})t\}, \quad (33)$$

where the spectral function $F_{ij}^{(\mathbf{u})}(\mathbf{k})$ is such that $\int d\mathbf{k} F_{ii}^{(\mathbf{u})}(\mathbf{k}) = 1$ and $\omega = \omega(\mathbf{k}) > 0$ is the dispersion curve for the wave motion.

For acoustic waves, e.g., we have $\omega(\mathbf{k}) = ck$ where c is the sound propagation velocity, for gravitational waves at the surface of a deep liquid $\omega(\mathbf{k}) = \sqrt{gk}$, for internal gravitational waves in a stratified medium we have $\omega(\mathbf{k}) = N\sqrt{k^2 - k_z^2}/k$, where N is the Brunt–Vaisala frequency, and for Rossby waves in the atmosphere and the ocean $\omega(\mathbf{k}) = -\beta k_x/k^2$, where β is the gradient of the Coriolis force in the y direction, and so on.

Let us assume that the spectral function satisfies the condition $\Phi_{ij}(0) = 0$ where $\Phi_{ij}(\omega) = \int d\mathbf{k} F_{ij}^{(\mathbf{u})}(\mathbf{k})\,\delta(\omega - \omega(\mathbf{k}))$ (we shall subsequently assume that the stronger condition $\Phi_{ij}(\omega)/\omega^2 \longrightarrow 0$ is satisfied where $\omega \longrightarrow 0$). For the correlation function of the random wave field $\tilde{\mathbf{u}}(\mathbf{r}, t)$ we then have

$$\int_0^\infty B_{ij}^{(\mathbf{u})}(0, t)dt = 0 \quad (34)$$

and consequently neither the approximation of a delta-correlated velocity field nor the diffusion approximation for the most interesting case $t \gg \tau_0$ give the final result since the diffusion coefficients vanish. In order to obtain the final result we need to allow for terms of a higher order of smallness. For the case $t \lesssim \tau_0$ the diffusion coefficients are nonzero although the particle behavior over this time period does not have a universal character and depends strongly on the model of the random velocity field.

However, we note that in the general case of the stochastic problem (2) of particle diffusion in a random force field having a wave nature whose correlation function is similar to (33), i.e.,

$$B_{ij}^{(\mathbf{f})}(\mathbf{r}, t) = \int d\mathbf{k} F_{ij}^{(\mathbf{f})}(\mathbf{k}) \cos\{\mathbf{k}\cdot\mathbf{r} - \omega(\mathbf{k})t\}, \quad (35)$$

Eqs. (19) and (31) with the diffusion coefficients (30) and (32) have the same meaning as before since in this case the diffusion coefficients do not vanish. Substituting Eq. (35) into (32) and integrating over time, we find that the diffusion coefficient is nonzero,

$$D_{ij}^{(1)}(\mathbf{v}) \sim \int d\mathbf{k} F_{ij}^{(\mathbf{f})}(\mathbf{k})\,\delta(\mathbf{k}\cdot\mathbf{v} - \omega(\mathbf{k})) \neq 0,$$

because of the existence of "wave–particle" resonance:

$$\mathbf{k}\cdot\mathbf{v} = \omega(\mathbf{k}.)$$

Let us assume that the maximum of the spectral function $F_{ij}(\mathbf{k})$ corresponds to a certain wave number k_m and the maximum of the spectral function $\Phi_{ij}(\omega)$ corresponds to the frequency ω_m. We define the spatial and temporal correlation scales as $l \sim 2\pi/k_m$, $\tau_0 \sim 2\pi/\omega_m$. In this case the value of $\varepsilon = \sigma_{\mathbf{u}}\tau_0/l$ for real wave fields is generally small and may be considered as the fundamental small parameter of the problem, i.e., $\varepsilon \ll 1$. We shall also assume that over the entire range where the spectrum of the velocity field is determined, the inequality $\sigma_{\mathbf{u}}k \ll \omega(k)$ is valid. This last condition is responsible for the absence of resonances between the different components of the velocity field.

The existence of maxima of the spectral functions $F_{ij}(\mathbf{k})$ and $\Phi_{ij}(\omega)$ by no means implies the presence of a quasi-regular component in the random velocity field. These exist because the velocity field itself is the result of differentiating (over space and time) other auxiliary wave fields (for example, the potential fields for the potential velocity field or the displacement fields of the interface, and so on). Naturally, if the spectral functions are very "narrow", i.e., they have a delta-shaped form relative to the central frequency (wave number), the problem can be preliminarily simplified by dynamic averaging over fast oscillations having the central frequency (wave number) of the initial stochastic equations. However, for most geophysical wave problems this situation does not occur.

Note that the hypothesis of statistical spatial uniformity generally has limited validity and is not valid, for example, for waves in atmospheric or oceanic waveguides, in analyses of transport by bounded wave packets, and so on. We shall subsequently confine our analysis to a Gaussian statistically uniform wave veloc-

ity field, focusing our attention on the fundamental aspects of the problem. In order to obtain specific quantitative results we need to consider statistical models of the wave field to within quadratic terms. In this case, average transport (Stokes drift) and particle diffusion generally occur and these were analyzed for various particular cases in [12–14], for example, using an approach proposed by Taylor [15]. Our aim is to apply a more general and systematic approach to this class of problems, which is valid for waves of different types and can be used to obtain various generalizations of transport theory based on the Fokker–Planck equation. This approach can be applied to calculate various statistical characteristics of particle ensembles transported by wave flows and to analyze effects associated with clustering and formation of coherent structures in impurity density fields based on methods of statistical topography [5–7].

We shall subsequently analyze in greater detail the diffusion of a passive impurity (particles and impurity concentration density fields) in a random velocity field having a wave nature for which, as we have already seen, the diffusion approximation is equivalent to the approximation of a delta-correlated velocity field and yields zero diffusion coefficients.

2. DIFFUSION OF A PASSIVE IMPURITY IN A RANDOM VELOCITY FIELD

The diffusion of a passive impurity in a random velocity field $\mathbf{u}(\mathbf{r}, t)$ is described by a linear first-order partial differential equation which is the equation of continuity for the density of a conservative impurity:

$$\left(\frac{\partial}{\partial t} + \sigma_{\mathbf{u}}\frac{\partial}{\partial \mathbf{r}}\tilde{\mathbf{u}}(\mathbf{r}, t)\right)\rho(\mathbf{r}, t) = 0, \qquad \rho(\mathbf{r}, 0) = \rho_0(\mathbf{r}), \tag{36}$$

where $\tilde{\mathbf{u}}(\mathbf{r}, t)$ is the random wave vector field, statistically uniform in space and steady-state in time, having the average value $\langle \tilde{\mathbf{u}}(\mathbf{r}, t)\rangle = 0$ and the correlation tensor

$$\langle \tilde{u}_i(\mathbf{r}, t)\tilde{u}_j(\mathbf{r}', t')\rangle = B_{ij}(\mathbf{r} - \mathbf{r}', t - t')$$

$$(B_{ii}(0, 0) = 1).$$

In this case, the total impurity mass is conserved, i.e.,

$$M = M(t) = \int d\mathbf{r}\rho(\mathbf{r}, t) = \int d\mathbf{r}\rho_0(\mathbf{r}) = M_0.$$

The linear first-order partial differential equation (36) can be solved using the method of statistics. Introducing characteristic curves (particle trajectories) whose dynamics is described by Eq. (1)

$$\frac{d}{dt}\mathbf{r}(t) = \sigma_{\mathbf{u}}\tilde{\mathbf{u}}(\mathbf{r}, t), \quad \mathbf{r}(0) = \mathbf{r}_0, \tag{37}$$

we can write (36) in the form

$$\frac{d}{dt}\rho(t) = -\sigma_{\mathbf{u}}\frac{\partial\tilde{\mathbf{u}}(\mathbf{r}, t)}{\partial \mathbf{r}}\rho(t), \qquad \rho(0) = \rho_0(\mathbf{r}_0) \tag{38}$$

This formulation of the problem corresponds to a Lagrangian description whereas the initial dynamic equation (36) corresponds to a Eulerian description.

The solution of the system of Eqs. (37) and (38) depends on a characteristic parameter, the initial value $\mathbf{r}_0$, which will be denoted by the vertical bar:

$$\mathbf{r}(t) = \mathbf{r}(t|\mathbf{r}_0), \quad \rho(t) = \rho(t|\mathbf{r}_0). \tag{39}$$

The first of the equalities (39) may be considered to be an algebraic equation for the parameter $\mathbf{r}_0$ for which the solution $\mathbf{r}_0 = \mathbf{r}_0(\mathbf{r}, t)$ exists since the divergence $j(t|\mathbf{r}_0) = \det\|j_{ik}(t|\mathbf{r}_0)\|$ is nonzero where

$$j_{ik}(t|\mathbf{r}_0) = \frac{\partial r_i(t|\mathbf{r}_0)}{\partial r_{0k}}.$$

Consequently, the solution of the initial equation (36) may be written as the equality

$$\rho(\mathbf{r}, t) = \int d\mathbf{r}_0\rho_0(\mathbf{r}_0)\delta(\mathbf{r}(t|\mathbf{r}_0) - \mathbf{r}), \tag{40}$$

which establishes a link between the Lagrangian and Eulerian characteristics.

The delta function on the right-hand side of equality (40) is an indicator function for the position of a Lagrangian particle and consequently, after averaging it over an ensemble of realizations of the random velocity field $\{\tilde{\mathbf{u}}(\mathbf{r}, t)\}$ we obtain the well-known relationship between the simultaneous probability density of the particle position in the Lagrangian description $P(t, \mathbf{r}|\mathbf{r}_0) = \langle\delta(\mathbf{r}(t|\mathbf{r}_0) - \mathbf{r})\rangle$ and the average density in the Euler description (see, for e.g., [10])

$$\langle\rho(\mathbf{r}, t)\rangle = \int d\mathbf{r}_0\rho_0(\mathbf{r}_0)P(t, \mathbf{r}|\mathbf{r}_0). \tag{41}$$

3. LAGRANGIAN DESCRIPTION

We introduce the notation for the indicator function of the coordinate of a Lagrangian particle,

$$\varphi(\mathbf{r}, t) = \delta(\mathbf{r}(t) - \mathbf{r}), \tag{42}$$

and also the first and second variational derivatives needed for the subsequent calculations of the statistical averages:

$$\frac{\delta\varphi(\mathbf{r}, t)}{\delta\tilde{u}_i(\mathbf{r}', t')} = \sigma_{\mathbf{u}}S_i(\mathbf{r}, t; \mathbf{r}', t'),$$
$$\frac{\delta^2\varphi(\mathbf{r}, t)}{\delta\tilde{u}_i(\mathbf{r}', t')\delta\tilde{u}_j(\mathbf{r}'', t'')} = \sigma_{\mathbf{u}}^2S_{ij}(\mathbf{r}, t; \mathbf{r}', t'; \mathbf{r}'', t''). \tag{43}$$

For the indicator function, we have the stochastic Liouville equation

$$\frac{\partial}{\partial t}\varphi(\mathbf{r}, t) = -\sigma_{\mathbf{u}}\frac{\partial}{\partial r_k}\{\tilde{u}_k(\mathbf{r}, t)\varphi(\mathbf{r}, t)\}, \qquad \varphi(\mathbf{r}, 0) = \delta(\mathbf{r} - \mathbf{r}_0), \tag{44}$$

which may be rewritten in the form of an integral equation

$$\varphi(\mathbf{r}, t) = \delta(\mathbf{r} - \mathbf{r}_0) - \sigma_{\mathbf{u}}\frac{\partial}{\partial r_k}\int_0^t d\tau\tilde{u}_k(\mathbf{r}, \tau)\varphi(\mathbf{r}, \tau). \tag{45}$$

Thus, for the first variational derivative (43), bearing in mind that this is only nonzero for $t \geq t'$, we obtain the stochastic integral equation

$$S_i(\mathbf{r}, t; \mathbf{r}', t') = \hat{L}_i(\mathbf{r}, \mathbf{r}')\varphi(\mathbf{r}, t')\theta(t - t') - \sigma_{\mathbf{u}}\frac{\partial}{\partial r_k}\int_0^t d\tau\tilde{u}_k(\mathbf{r}, \tau)S_i(\mathbf{r}, \tau; \mathbf{r}', t'). \tag{46}$$

where the action of the operator $\hat{L}_i(\mathbf{r}, \mathbf{r}')$ on the function $f(\mathbf{r})$ is described by the formula

$$\hat{L}_i(\mathbf{r}, \mathbf{r}')f(\mathbf{r}) = -\frac{\partial}{\partial r_i}\{\delta(\mathbf{r} - \mathbf{r}')f(\mathbf{r})\}. \tag{47}$$

Similarly, for the second variational derivative we obtain the stochastic integral equation

$$\begin{aligned} S_{ij}(\mathbf{r}, t; \mathbf{r}', t'; \mathbf{r}'', t'') &= \hat{L}_i(\mathbf{r}, \mathbf{r}')S_i(\mathbf{r}, t'; \mathbf{r}'', t'')\theta(t - t')\theta(t' - t'') \\ &= \hat{L}_j(\mathbf{r}, \mathbf{r}'')S_i(\mathbf{r}, t''; \mathbf{r}', t')\theta(t - t'')\theta(t'' - t') \\ &- \sigma_{\mathbf{u}}\frac{\partial}{\partial r_l}\int_{\max\{t', t''\}}^t d\tau\tilde{u}_k(\mathbf{r}, \tau)S_{ij}(\mathbf{r}, \tau; \mathbf{r}', t'; \mathbf{r}'', t''). \end{aligned} \tag{48}$$

We now average Eq. (44) over the ensemble of realizations of the field $\{\tilde{\mathbf{u}}(\mathbf{r}, t)\}$. For the Lagrangian probability density of the particle position $P(\mathbf{r}, t) = \langle\varphi(\mathbf{r}, t)\rangle$, using the approach from [3–6] and taking into account the Furutsu–Novikov formula (7) we then obtain

$$\frac{\partial}{\partial t}P(\mathbf{r}, t) = -\sigma_{\mathbf{u}}^2\frac{\partial}{\partial r_k}\int d\mathbf{r}' \int_0^t dt' B_{ki}(\mathbf{r} - \mathbf{r}', t - t')\langle S_i(\mathbf{r}, t; \mathbf{r}', t')\rangle, \qquad P(\mathbf{r}, 0) = \delta(\mathbf{r} - \mathbf{r}_0). \tag{49}$$

Integrating Eq. (49) with respect to time over the range (t_1, t) where $t_1 < t$, we obtain the equality

$$P(\mathbf{r}, t) - P(\mathbf{r}, t_1) = -\sigma_{\mathbf{u}}^2\frac{\partial}{\partial r_l}\int_{t_1}^t d\tau\int d\mathbf{r}'' \int_0^\tau dt'' B_{lj}(\mathbf{r} - \mathbf{r}'', \tau - t'')\langle S_j(\mathbf{r}, \tau; \mathbf{r}'', t'')\rangle. \tag{50}$$

We now average Eq. (46) over the ensemble of realizations of the field $\{\tilde{u}_k(\mathbf{r}, t)\}$. For $\langle S_i(\mathbf{r}, t; \mathbf{r}', t')\rangle$ we then obtain the equation

$$\begin{aligned} \langle S_i(\mathbf{r}, t; \mathbf{r}', t')\rangle &= \hat{L}_i(\mathbf{r}, \mathbf{r}')P(\mathbf{r}, t')\theta(t - t') \\ &- \sigma_{\mathbf{u}}^2\frac{\partial}{\partial r_l}\int_{t'}^t d\tau\int d\mathbf{r}''\int_0^t dt'' \\ &\times B_{lj}(\mathbf{r} - \mathbf{r}'', \tau - t'')\langle S_{ij}(\mathbf{r}, \tau; \mathbf{r}', t'; \mathbf{r}'', t'')\rangle. \end{aligned} \tag{51}$$

For the function $\langle S_{ij}(\mathbf{r}, t; \mathbf{r}', t''; \mathbf{r}'', t)\rangle$ we use the approximate expression

$$\begin{aligned} &\langle S_{ij}(\mathbf{r}, t; \mathbf{r}', t''; \mathbf{r}'', t)\rangle \\ &= \hat{L}_i(\mathbf{r}, \mathbf{r}')\langle S_j(\mathbf{r}, t'; \mathbf{r}'', t'')\rangle\theta(t - t')\theta(t' - t'') \\ &+ \hat{L}_j(\mathbf{r}, \mathbf{r}'')\langle S_i(\mathbf{r}, t''; \mathbf{r}', t')\rangle\theta(t - t'')\theta(t'' - t'), \end{aligned} \tag{52}$$

which corresponds to neglecting the third-order variational derivatives in (48). Using this approximation and equality (50), we can write Eq. (51) as a closed integral equation:

$$\begin{aligned} \langle S_i(\mathbf{r}, t; \mathbf{r}', t')\rangle &= \hat{L}_i(\mathbf{r}, \mathbf{r}')P(\mathbf{r}, t)\theta(t - t') \\ &+ \sigma_{\mathbf{u}}^2\hat{L}_i(\mathbf{r}, \mathbf{r}')\frac{\partial}{\partial r_l}\int_{t'}^t d\tau\int d\mathbf{r}''\int_0^\tau dt'' \\ &\times B_{lj}(\mathbf{r} - \mathbf{r}'', \tau - t'')\langle S_j(\mathbf{r}, \tau; \mathbf{r}'', t'')\rangle \\ &- \sigma_{\mathbf{u}}^2\frac{\partial}{\partial r_l}\int_{t'}^t d\tau\int d\mathbf{r}''\int_0^{t'} dt'' \\ &\times B_{lj}(\mathbf{r} - \mathbf{r}'', \tau - t'')\hat{L}_i(\mathbf{r}, \mathbf{r}')\langle S_j(\mathbf{r}, t'; \mathbf{r}'', t'')\rangle \\ &- \sigma_{\mathbf{u}}^2\frac{\partial}{\partial r_l}\int_{t'}^t d\tau\int d\mathbf{r}''\int_{t'}^\tau dt'' \\ &\times B_{lj}(\mathbf{r} - \mathbf{r}'', \tau - t'')\hat{L}_j(\mathbf{r}, \mathbf{r}'')\langle S_i(\mathbf{r}, t''; \mathbf{r}', t')\rangle. \end{aligned} \tag{53}$$

Solving Eq. (53) for $\langle S_i(\mathbf{r}, t; \mathbf{r}', t')\rangle$ by the method of successive approximations in terms of the parameter $\sigma_{\mathbf{u}}^2$ to within small terms (in this case the time argu-

ments t_i of the functions $P(\mathbf{r}, t_i)$ may be replaced by t), we have

$$\begin{aligned}\langle S_i(\mathbf{r}, t; \mathbf{r}', t')\rangle \;=\; &\Big\{\hat{L}_i(\mathbf{r}, \mathbf{r}')\\ &+\sigma_{\mathbf{u}}^2\hat{L}_i(\mathbf{r}, \mathbf{r}')\frac{\partial}{\partial r_l}\int_{t'}^{t}d\tau\int d\mathbf{r}''\int_0^{\tau}dt''\,B_{lj}(\mathbf{r}-\mathbf{r}'', t'')\hat{L}_j(\mathbf{r}, \mathbf{r}'')\\ &-\sigma_{\mathbf{u}}^2\frac{\partial}{\partial r_l}\int_{t'}^{t}d\tau\int d\mathbf{r}''\int_0^{t'}dt''\\ &\times B_{lj}(\mathbf{r}-\mathbf{r}'', \tau-t'')\hat{L}_i(\mathbf{r}, \mathbf{r}')\hat{L}_j(\mathbf{r}, \mathbf{r}'')\\ &-\sigma_{\mathbf{u}}^2\frac{\partial}{\partial r_l}\int_{t'}^{t}d\tau\int d\mathbf{r}''\int_{t'}^{\tau}dt''\\ &\times B_{lj}(\mathbf{r}-\mathbf{r}'', \tau-t'')\hat{L}_j(\mathbf{r}, \mathbf{r}'')\hat{L}_i(\mathbf{r}, \mathbf{r}')\Big\}P(\mathbf{r}, t).\end{aligned} \tag{54}$$

Substituting (54) into (49), we can integrate over all spatial variables and obtain a third-order equation in $\mathbf{r}$ (in which we can omit terms with first-order derivatives proportional to $\sigma_{\mathbf{u}}^4$):

$$\begin{aligned}\frac{\partial}{\partial t}P(\mathbf{r}, t) \;=\; &-\sigma_{\mathbf{u}}^2\int_0^t dt'\frac{\partial B_{ki}(0, t')}{\partial r_i}\frac{\partial}{\partial r_k}P(\mathbf{r}, t)\\ &+\sigma_{\mathbf{u}}^2\int_0^t dt' B_{ki}(0, t')\frac{\partial^2}{\partial r_k\partial r_i}P(\mathbf{r}, t)\\ &+\sigma_{\mathbf{u}}^4\frac{\partial^2}{\partial r_k\partial r_l}\int_0^t dt' B_{ki}(0, t-t')\int_{t'}^{t}d\tau\\ &\times\int_0^{t'}dt''\frac{\partial^2 B_{lj}(0, \tau-t'')}{\partial r_i\partial r_j}P(\mathbf{r}, t)\\ &+\sigma_{\mathbf{u}}^4\frac{\partial^2}{\partial r_k\partial r_l}\int_0^t dt'\frac{\partial^2 B_{ki}(0, t-t')}{\partial r_i\partial r_j}\int_{t'}^{t}d\tau\\ &\times\int_0^{\tau}dt''\,B_{lj}(0, \tau-t'')P(\mathbf{r}, t)\\ &+\sigma_{\mathbf{u}}^4\frac{\partial^2}{\partial r_k\partial r_j}\int_0^t dt'\frac{\partial B_{ki}(0, t-t')}{\partial r_l}\int_{t'}^{t}d\tau\\ &\times\int_0^{t'}dt''\frac{\partial B_{lj}(0, \tau-t'')}{\partial r_i}P(\mathbf{r}, t)\end{aligned} \tag{55}$$

$$\begin{aligned}&+\sigma_{\mathbf{u}}^4\frac{\partial^2}{\partial r_k\partial r_i}\int_0^t dt'\frac{\partial B_{ki}(0, t-t')}{\partial r_l}\int_{t'}^{t}d\tau\int_0^{\tau}dt''\frac{\partial B_{lj}(0, t'')}{\partial r_j}P(\mathbf{r}, t)\\ &+\sigma_{\mathbf{u}}^4\frac{\partial^2}{\partial r_k\partial r_j}\int_0^t dt'\frac{\partial B_{ki}(0, t-t')}{\partial r_l}\int_{t'}^{t}d\tau\int_0^{\tau}dt''\,B_{lj}(0, t'')P(\mathbf{r}, t)\\ &+\sigma_{\mathbf{u}}^4\frac{\partial^2}{\partial r_k\partial r_i}\int_0^t dt'\frac{\partial^2 B_{ki}(0, t-t')}{\partial r_l\partial r_j}\int_{t'}^{t}d\tau\\ &\times\int_{t'}^{\tau}dt''\,B_{lj}(0, \tau-t'')P(\mathbf{r}, t)\\ &-\sigma_{\mathbf{u}}^4\frac{\partial^3}{\partial r_k\partial r_l\partial r_j}\int_0^t dt' B_{ki}(0, t-t')\int_{t'}^{t}d\tau\\ &\times\int_0^{\tau}dt''\frac{\partial B_{lj}(0, \tau-t'')}{\partial r_i}P(\mathbf{r}, t)\\ &-\sigma_{\mathbf{u}}^4\frac{\partial^3}{\partial r_k\partial r_i\partial r_j}\int_0^t dt'\frac{\partial B_{ki}(0, t-t')}{\partial r_j}\int_{t'}^{t}d\tau\\ &\times\int_{t'}^{\tau}dt''\,B_{lj}(0, \tau-t'')P(\mathbf{r}, t)\\ &-\sigma_{\mathbf{u}}^4\frac{\partial^3}{\partial r_k\partial r_i\partial r_j}\int_0^t dt'\frac{\partial B_{ki}(0, t-t')}{\partial r_l}\int_{t'}^{t}d\tau\int_0^{\tau}dt''\,B_{lj}(0, t'')P(\mathbf{r}, t).\end{aligned}$$

Equation (55) is not generally an equation for the probability density since it may lead to negative quantities in the range of low values. At the same time, its solution accurately describes the statistical moments and in this sense is a generalization of the Fokker–Planck equation. Now using the spectral representation of the velocity field (33) and its properties, we can integrate with respect to time in the coefficients of the equation and for large time values ($t \gg \tau_0$) Eq. (55) may be written in the form

$$\begin{aligned}\frac{\partial}{\partial t}P(\mathbf{r}, t) \;=\; &-\sigma_{\mathbf{u}}^2\int\frac{d\mathbf{k}}{\omega(\mathbf{k})}k_i F_{ki}(\mathbf{k})\frac{\partial}{\partial r_k}P(\mathbf{r}, t)\\ &+\sigma_{\mathbf{u}}^4\frac{\pi}{2}\int d\mathbf{k}_1\int\frac{d\mathbf{k}_2}{\omega_2^2}k_{1l}k_{1j}F_{ki}(\mathbf{k}_1)F_{lj}(\mathbf{k}_2)\end{aligned}$$

$$\times \delta(\omega_1 - \omega_2)\frac{\partial^2}{\partial r_k \partial r_i}P(\mathbf{r}, t) \tag{56}$$

$$+ \sigma_{\mathbf{u}}^4 \frac{\pi}{2}\int d\mathbf{k}_1 \int \frac{d\mathbf{k}_2}{\omega_2^2} k_{1l}k_{2i}F_{ki}(\mathbf{k}_1)F_{lj}(\mathbf{k}_2)$$

$$\times \delta(\omega_1 - \omega_2)\frac{\partial^2}{\partial r_k \partial r_j}P(\mathbf{r}, t),$$

where $\omega_1 = \omega(\mathbf{k}_1)$, $\omega_2 = \omega(\mathbf{k}_2)$.

Equation (56) is the Fokker–Planck equation describing the probability density of the position of a particle transported by a statistically uniform Gaussian wave velocity field.

For isotropic fluctuations of the field $\tilde{\mathbf{u}}(\mathbf{r}, t)$ Eq. (56) is simplified and has the form

$$\frac{\partial}{\partial t}P(\mathbf{r}, t) = D_N \frac{\partial^2}{\partial \mathbf{r}^2}P(\mathbf{r}, t), \tag{57}$$

which corresponds to the Gaussian random vector process $\mathbf{r}(t)$ with the average $\langle \mathbf{r}(t)\rangle = \mathbf{r}_0$ and the dispersion

$$\sigma_{\mathbf{r}}^2(t) = \langle (\mathbf{r}(t) - \mathbf{r}_0)^2\rangle = 2ND_N t, \tag{58}$$

where N is the dimensions of space and D_N is the diffusion coefficient,

$$D_N = \sigma_{\mathbf{u}}^4 \frac{\pi}{2N}\int d\mathbf{k}_1 \times \int \frac{d\mathbf{k}_2}{\omega_2^2} k_{1l}k_{1j}F_{ii}(\mathbf{k}_1)F_{lj}(\mathbf{k}_2)\delta(\omega_1 - \omega_2). \tag{59}$$

In this case, the spectral tensor of the wave velocity field has the structure

$$F_{ki}(\mathbf{k}) = F^s(k)\left(\delta_{ik} - \frac{k_i k_k}{\mathbf{k}^2}\right) + F^p(k)\frac{k_i k_k}{\mathbf{k}^2}, \tag{60}$$

where $F^s(k)$ and $F^p(k)$ are the solenoidal and potential components of the spectral tensor, respectively, $\omega(\mathbf{k}) \equiv \omega(k)$, and consequently we obtain the following expression for the diffusion coefficient:

$$D_N = \sigma_{\mathbf{u}}^4 \frac{\pi}{2N}\int \frac{d\mathbf{k}_1}{\omega_1^2} k_1^2 F_{ii}(\mathbf{k}_1)\int d\mathbf{k}_2 F_{ll}(\mathbf{k}_2)\delta(\omega_1 - \omega_2)$$

$$= \sigma_{\mathbf{u}}^4 \frac{\pi}{2N}\int \frac{d\mathbf{k}_1}{\omega_1^2} k_1^2 [F^s(k_1)(N-1) + F^p(k_1)]^2 \tag{61}$$

$$\times \int d\mathbf{k}_2 \delta(\omega_1 - \omega_2).$$

For an anisotropic medium spatial asymmetry of the vector process $\mathbf{r}(t)$ appears. Its average and dispersion are described by the expressions

$$\langle r_m(t)\rangle = r_{0m} + t\sigma_{\mathbf{u}}^2 \int \frac{d\mathbf{k}}{\omega(\mathbf{k})} k_i F_{mi}(\mathbf{k}),$$

$$\sigma_{\mathbf{r}}^2(t) = \langle \mathbf{r}^2(t) - \langle \mathbf{r}(t)\rangle^2\rangle$$

$$= t\sigma_{\mathbf{u}}^4 \pi \int d\mathbf{k}_1 \int \frac{d\mathbf{k}_2}{\omega_2^2} k_{1l}k_{1j}F_{ii}(\mathbf{k}_1)F_{lj}(\mathbf{k}_2)\delta(\omega_1 - \omega_2) \tag{62}$$

$$+ t\sigma_{\mathbf{u}}^4 \pi \int d\mathbf{k}_1 \int \frac{d\mathbf{k}_2}{\omega_2^2} k_{1l}k_{2i}F_{ki}(\mathbf{k}_1)F_{lk}(\mathbf{k}_2)\delta(\omega_1 - \omega_2).$$

As we can see, the diffusion coefficient is proportional not to the dispersion of the velocity field but to its square. This is because no "wave–particle" resonances appear in this problem, which leads to a reduction in the order of dispersion of the random particle drift velocity. This problem is similar to the problems involving the oscillations of a Kapitza pendulum or the vortex drift of charged particles in a rapidly varying electric field [1], where the dominant effect is also quadratic.

4. EULER DESCRIPTION

We shall now give a statistical description of the Euler representation. For simplicity, we shall assume that the initial distribution of the density field is constant, $\rho_0(\mathbf{r}) = \rho_0 = \text{const}$ and consequently the random function $\rho(\mathbf{r}, t)$ will be statistically uniform in space, i.e., all its single-point statistical characteristics will not depend on the spatial point $\mathbf{r}$.

We shall introduce an indicator function similar to the function (42) in the Lagrange description,

$$\varphi(\mathbf{r}, t; \rho) = \delta(\rho(\mathbf{r}, t) - \rho), \tag{63}$$

and the first and second variational derivatives:

$$\frac{\delta\varphi(\mathbf{r}, t; \rho)}{\delta\tilde{u}_i(\mathbf{r}', t')} = \sigma_{\mathbf{u}} S_i(\mathbf{r}, t; \mathbf{r}', t'; \rho),$$

$$\frac{\delta^2\varphi(\mathbf{r}, t; \rho)}{\delta\tilde{u}_i(\mathbf{r}', t')\delta\tilde{u}_j(\mathbf{r}'', t'')} = \sigma_{\mathbf{u}}^2 S_{ij}(\mathbf{r}, t; \mathbf{r}', t'; \mathbf{r}'', t''''; \rho). \tag{64}$$

For the indicator function $\varphi(\mathbf{r}, t; \rho)$ using Eq. (36) we obtain the stochastic Liouville equation (see, e.g., [4–7]), which we write in the form

$$\frac{\partial}{\partial t}\varphi(\mathbf{r}, t; \rho) = \sigma_{\mathbf{u}}\hat{N}(\mathbf{r}, t; \rho)\varphi(\mathbf{r}, t; \rho),$$
$$\varphi(\mathbf{r}, 0; \rho) = \delta(\rho(\mathbf{r}, 0) - \rho_0), \tag{65}$$

where

$$\hat{N}(\mathbf{r}, t; \rho) = -\frac{\partial}{\partial \mathbf{r}}\tilde{\mathbf{u}}(\mathbf{r}, t) + \frac{\partial\tilde{\mathbf{u}}(\mathbf{r}, t)}{\partial \mathbf{r}}\left(1 + \frac{\partial}{\partial\rho}\rho\right). \tag{66}$$

Equation (65) may be rewritten as an integral equation

$$\varphi(\mathbf{r}, t; \rho) = \delta(\rho(\mathbf{r}, 0) - \rho_0) + \sigma_{\mathbf{u}}\int_0^t d\tau \hat{N}(\mathbf{r}, \tau; \rho)\varphi(\mathbf{r}, \tau; \rho). \tag{67}$$

Thus, for the first variational derivative (64) we obtain the stochastic integral equation

$$S_i(\mathbf{r}, t; \mathbf{r}', t'; \rho) = \hat{N}_i(\mathbf{r}, \mathbf{r}'; \rho)\varphi(\mathbf{r}, t'; \rho)\theta(t-t') + \sigma_{\mathbf{u}}\int_{t'}^{t} d\tau \hat{N}(\mathbf{r}, \tau; \rho)S_I(\mathbf{r}, \tau; \mathbf{r}', t'; \rho), \quad (68)$$

where

$$\hat{N}_i(\mathbf{r}, \mathbf{r}'; \rho) = \{\hat{L}_i(\mathbf{r}, \mathbf{r}') + \hat{M}_i(\mathbf{r}, \mathbf{r}'; \rho)\},$$

the operator $\hat{L}_i(\mathbf{r}, \mathbf{r}')$ is described by formula (47) and the action of the operator $\hat{M}_i(\mathbf{r}, \mathbf{r}'; \rho)$ on the function $f(\mathbf{r}; \rho)$ is described by the formula

$$\hat{M}_i(\mathbf{r}, \mathbf{r}'; \rho) f(\mathbf{r}; \rho) = \frac{\partial\delta(\mathbf{r}-\mathbf{r}')}{\partial r_i}\left(1 + \frac{\partial}{\partial\rho}\rho\right) f(\mathbf{r}; \rho). \quad (69)$$

Similarly, for the second variational derivative we obtain the stochastic integral equation

$$\begin{aligned} S_{ij}(\mathbf{r}, t; \mathbf{r}', t'; \mathbf{r}'', t''; \rho) &= \hat{N}_i(\mathbf{r}, \mathbf{r}'; \rho)S_j(\mathbf{r}, t'; \mathbf{r}'', t''; \rho)\theta(t-t')\theta(t'-t'') \\ &+ \hat{N}_j(\mathbf{r}, \mathbf{r}''; \rho)S_i(\mathbf{r}, t''; \mathbf{r}', t'; \rho)\theta(t-t'')\theta(t''-t') \quad (70) \\ &+ \sigma_{\mathbf{u}}\int_{\max\{t', t''\}}^{t} d\tau \hat{N}(\mathbf{r}, \tau''''; \rho)S_{ij}(\mathbf{r}, \tau; \mathbf{r}', t'; \mathbf{r}'', t''; \rho). \end{aligned}$$

We shall subsequently proceed as in the Lagrange description. We shall average Eq. (65) over the ensemble of realizations of the field $\{\tilde{u}_k(\mathbf{r}, t)\}$. Then for the Euler probability density $P(t; \rho) = \langle\varphi(\mathbf{r}, t; \rho)\rangle\}$, taking into account the Furutsu–Novikov formula (7), we obtain

$$\begin{aligned} \frac{\partial}{\partial t}P(t; \rho) &= \sigma_{\mathbf{u}}^2\int d\mathbf{r}'\int_0^t dt' \frac{\partial B_{ki}(\mathbf{r}-\mathbf{r}', t-t')}{\partial r_k} \\ &\times\left(1 + \frac{\partial}{\partial\rho}\rho\right)\langle S_i(\mathbf{r}, t; \mathbf{r}', t')\rangle, \quad (71) \\ P(0; \rho) &= \delta(\rho - \rho_0). \end{aligned}$$

We shall now average Eq. (68) over the ensemble of realizations of the field $\{\tilde{u}_k(\mathbf{r}, t)\}$. For $\langle S_i(\mathbf{r}, t; \mathbf{r}', t'; \rho)\rangle$ we then obtain

$$\begin{aligned} \langle S_i(\mathbf{r}, t; \mathbf{r}', t'; \rho)\rangle &= \hat{N}_i(\mathbf{r}, \mathbf{r}'; \rho)P(t'; \rho)\theta(t-t') \\ &- \sigma_{\mathbf{u}}^2\int_{t'}^{t} d\tau\int d\mathbf{r}''\int_0^t dt'' \frac{\partial}{\partial r_k}B_{kj}(\mathbf{r}-\mathbf{r}'', \tau-t'') \\ &\times\langle S_{ij}(\mathbf{r}, \tau; \mathbf{r}', t'; \mathbf{r}'', t''; \rho)\rangle \quad (72) \\ &+ \sigma_{\mathbf{u}}^2\int_{t'}^{t} d\tau\int d\mathbf{r}''\int_0^t dt'' \frac{\partial B_{kj}(\mathbf{r}-\mathbf{r}'', \tau-t'')}{\partial r_k} \\ &\times\left(1 + \frac{\partial}{\partial\rho}\rho\right)\langle S_{ij}(\mathbf{r}, \tau; \mathbf{r}', t'; \mathbf{r}'', t''; \rho)\rangle. \end{aligned}$$

For the function $\langle S_{ij}(\mathbf{r}, t; \mathbf{r}', t'; \mathbf{r}'', t'')\rangle$ we use the approximate expression

$$\begin{aligned} &\langle S_{ij}(\mathbf{r}, t; \mathbf{r}', t'; \mathbf{r}'', t''; \rho)\rangle \\ &= \hat{N}_i(\mathbf{r}, \mathbf{r}'; \rho)\langle S_j(\mathbf{r}, t'; \mathbf{r}'', t''; \rho)\rangle\theta(t-t')\theta(t'-t'') \quad (73) \\ &+ \hat{N}_j(\mathbf{r}, \mathbf{r}''; \rho)\langle S_i(\mathbf{r}, t''; \mathbf{r}', t'; \rho)\rangle\theta(t-t'')\theta(t''-t'), \end{aligned}$$

which corresponds to neglect of the third-order variational derivatives in (70). Using this approximation, Eq. (72) may be written in the form of a closed integral equation:

$$\begin{aligned} \langle S_i(\mathbf{r}, t; \mathbf{r}', t'; \rho)\rangle &= \hat{N}_i(\mathbf{r}, \mathbf{r}'; \rho)P(t; \rho)\theta(t-t') \\ &- \sigma_{\mathbf{u}}^2\hat{N}_i(\mathbf{r}, \mathbf{r}'; \rho)\int_{t'}^{t} d\tau\int d\mathbf{r}''\int_0^{\tau} dt'' \frac{\partial B_{lj}(\mathbf{r}-\mathbf{r}'', \tau-t'')}{\partial r_l} \\ &\times\left(1 + \frac{\partial}{\partial\rho}\rho\right)\langle S_j(\mathbf{r}, \tau; \mathbf{r}'', t''; \rho)\rangle \\ &- \sigma_{\mathbf{u}}^2\int_{t'}^{t} d\tau\int d\mathbf{r}''\int_0^{\tau} dt'' \frac{\partial}{\partial r_l}B_{lj}(\mathbf{r}-\mathbf{r}'', \tau-t'') \\ &\times\hat{N}_i(\mathbf{r}, \mathbf{r}'; \rho)\langle S_j(\mathbf{r}, t'; \mathbf{r}'', t''; \rho)\rangle \\ &- \sigma_{\mathbf{u}}^2\int_{t'}^{t} d\tau\int d\mathbf{r}''\int_{t'}^{\tau} dt'' \frac{\partial}{\partial r_l}B_{lj}(\mathbf{r}-\mathbf{r}'', \tau-t'') \quad (74) \\ &\times\hat{N}_i(\mathbf{r}, \mathbf{r}''; \rho)\langle S_i(\mathbf{r}, t''; \mathbf{r}', t'; \rho)\rangle \\ &+ \sigma_{\mathbf{u}}^2\int_{t'}^{t} d\tau\int d\mathbf{r}''\int_0^{\tau} dt'' \frac{\partial B_{lj}(\mathbf{r}-\mathbf{r}'', \tau-t'')}{\partial r_l} \\ &\times\left(1 + \frac{\partial}{\partial\rho}\rho\right)\hat{N}_i(\mathbf{r}, \mathbf{r}'; \rho)\langle S_j(\mathbf{r}, t'; \mathbf{r}'', t''; \rho)\rangle \\ &+ \sigma_{\mathbf{u}}^2\int_{t'}^{t} d\tau\int d\mathbf{r}''\int_{t'}^{\tau} dt'' \frac{\partial B_{lj}(\mathbf{r}-\mathbf{r}'', \tau-t'')}{\partial r_l} \\ &\times\left(1 + \frac{\partial}{\partial\rho}\rho\right)\hat{N}_j(\mathbf{r}, \mathbf{r}''; \rho)\langle S_i(\mathbf{r}, t''; \mathbf{r}', t'; \rho)\rangle. \end{aligned}$$

Solving Eq. (74) for $\langle S_i(\mathbf{r}, t; \mathbf{r}', t'; \rho)\rangle$ by the method of successive approximations for the parameter $\sigma_{\mathbf{u}}^2$ to within small terms [in this case, the time argu-

ments t_i of the functions $P(t_i; \rho)$ may be replaced by t] and integrating with respect to $\mathbf{r}''$ we obtain for $t > t'$

$$\begin{aligned}&\langle S_i(\mathbf{r}, t; \mathbf{r}', t'; \rho)\rangle \\ &= \hat{T}_i(\mathbf{r}, t; \mathbf{r}', t'; \rho)P(t; \rho)\theta(t-t'),\end{aligned} \tag{75}$$

where the operator $\hat{T}_i(\mathbf{r}, t; \mathbf{r}', t'; \rho)$ is given by

$$\begin{aligned}\hat{T}_i(\mathbf{r}, t; \mathbf{r}', t'; \rho) &= \frac{\partial\delta(\mathbf{r}-\mathbf{r}')}{\partial r_i}\frac{\partial}{\partial\rho}\rho \\ &+ \sigma_{\mathbf{u}}^2\frac{\partial\delta(\mathbf{r}-\mathbf{r}')}{\partial r_i}\frac{\partial}{\partial\rho}\rho\int_{t'}^{t}d\tau\int_0^{\tau}dt''\\ &\times\frac{\partial B_{lj}(0, \tau-t'')}{\partial r_l\partial r_j}\left(1+\frac{\partial}{\partial\rho}\rho\right)\frac{\partial}{\partial\rho}\rho \\ &+ \sigma_{\mathbf{u}}^2\frac{\partial^2\delta(\mathbf{r}-\mathbf{r}')}{\partial r_i\partial r_l}\int_{t'}^{t}d\tau\int_0^{\tau}dt''\frac{\partial B_{lj}(0, \tau-t'')}{\partial r_l}\frac{\partial}{\partial\rho}\rho\frac{\partial}{\partial\rho}\rho \\ &+ \sigma_{\mathbf{u}}^2\frac{\partial^3\delta(\mathbf{r}-\mathbf{r}')}{\partial r_i\partial r_j\partial r_l}\int_{t'}^{t}d\tau\int_0^{\tau}dt''B_{lj}(0, \tau-t'')\frac{\partial}{\partial\rho}\rho \\ &+ \sigma_{\mathbf{u}}^2\frac{\partial^2\delta(\mathbf{r}-\mathbf{r}')}{\partial r_i\partial r_l}\int_{t'}^{t}d\tau\int_0^{\tau}dt''\frac{\partial B_{lj}(0, \tau-t'')}{\partial r_j}\frac{\partial}{\partial\rho}\rho\frac{\partial}{\partial\rho}\rho \\ &- \sigma_{\mathbf{u}}^2\int_{t'}^{t}d\tau\int_0^{t'}dt''\frac{\partial B_{lj}(0, \tau-t'')}{\partial r_l\partial r_j}\left(1+\frac{\partial}{\partial\rho}\rho\right) \\ &\times\frac{\partial\delta(\mathbf{r}-\mathbf{r}')}{\partial r_i}\frac{\partial}{\partial\rho}\rho\frac{\partial}{\partial\rho}\rho \\ &- \sigma_{\mathbf{u}}^2\frac{\partial^2\delta(\mathbf{r}-\mathbf{r}')}{\partial r_i\partial r_j}\int_{t'}^{t}d\tau\int_{t'}^{\tau}dt''\frac{\partial B_{lj}(0, \tau-t'')}{\partial r_l}\left(1+\frac{\partial}{\partial\rho}\rho\right)\frac{\partial}{\partial\rho}\rho \\ &- \sigma_{\mathbf{u}}^2\frac{\partial\delta(\mathbf{r}-\mathbf{r}')}{\partial r_i}\int_{t'}^{t}d\tau\int_{t'}^{\tau}dt''\frac{\partial^2 B_{lj}(0, \tau-t'')}{\partial r_l\partial r_j} \\ &\times\left(1+\frac{\partial}{\partial\rho}\rho\right)\frac{\partial}{\partial\rho}\rho\frac{\partial}{\partial\rho}\rho.\end{aligned} \tag{76}$$

Substituting (76) into (71) and integrating with respect to $\mathbf{r}'$, we obtain

$$\begin{aligned}\frac{\partial}{\partial t}P(t; \rho) &= -\sigma_{\mathbf{u}}^2\int_0^t dt'\frac{\partial^2 B_{ki}(0, t')}{\partial r_k\partial r_i}\frac{\partial^2}{\partial\rho^2}\rho^2 P(t; \rho) \\ &- \sigma_{\mathbf{u}}^4\int_0^t dt'\frac{\partial^2 B_{ki}(0, t')}{\partial r_k\partial r_i}\int_0^{t'}d\tau\int_{\tau}^{t}dt'' \\ &\times\frac{\partial^2 B_{lj}(0, t''-\tau)}{\partial r_l\partial r_j}\frac{\partial^2}{\partial\rho^2}\rho^2\frac{\partial^2}{\partial\rho^2}\rho^2 P(t; \rho) \\ &+ \sigma_{\mathbf{u}}^4\int_0^t dt'\frac{\partial^2 B_{ki}(0, t')}{\partial r_k\partial r_i}\int_0^{t'}d\tau\int_0^{t}dt'' \\ &\times\frac{\partial^2 B_{lj}(0, t''-\tau)}{\partial r_j}\frac{\partial^2}{\partial\rho^2}\rho^2\frac{\partial}{\partial\rho}\rho P(t; \rho) \\ &- \sigma_{\mathbf{u}}^4\int_0^t dt'\frac{\partial^4 B_{ki}(0, t')}{\partial r_k\partial r_i\partial r_l}\int_0^{t'}d\tau\int_{\tau}^{t}dt''B_{lj}(0, t''-\tau)\frac{\partial^2}{\partial\rho^2}\rho^2 P(t; \rho) \\ &- \sigma_{\mathbf{u}}^4\int_0^t dt'\frac{\partial^3 B_{ki}(0, t')}{\partial r_k\partial r_i\partial r_l}\int_0^{t'}d\tau\int_{\tau}^{t}dt'' \\ &\times\frac{\partial B_{lj}(0, t''-\tau)}{\partial r_j}\frac{\partial^2}{\partial\rho^2}\rho^2 P(t; \rho) \\ &+ \sigma_{\mathbf{u}}^4\int_0^t dt'\frac{\partial^2 B_{ki}(0, t')}{\partial r_k\partial r_i}\int_0^{t'}d\tau\int_{t'}^{t}dt'' \\ &\times\frac{\partial^2 B_{lj}(0, t''-\tau)}{\partial r_l\partial r_j}\left(1+\frac{\partial}{\partial\rho}\rho\right)\frac{\partial^2}{\partial\rho^2}\rho^2\frac{\partial}{\partial\rho}\rho P(t; \rho).\end{aligned} \tag{77}$$

Now using the spectral representation (33) and integrating with respect to time in the coefficients of the equation for long times, we obtain the final equation

$$\begin{aligned}\frac{\partial}{\partial t}P(t; \rho) &= \tilde{D}_N^{(2)}\frac{\partial^2}{\partial\rho^2}\rho^2 P(t; \rho) \\ &+ \tilde{D}_N^{(3)}\frac{\partial^2}{\partial\rho^2}\rho^2\frac{\partial}{\partial\rho}\rho P(t; \rho),\end{aligned} \tag{78}$$

where N is the spatial dimension and

$$\begin{aligned}\tilde{D}_N^{(2)} &= \sigma_{\mathbf{u}}^4\frac{\pi}{2}\int d\mathbf{k}_1 k_{1k}k_{1i}k_{1l}(k_{1j}-k_{2j})F_{ki}(\mathbf{k}_1) \\ &\times\int\frac{d\mathbf{k}_2}{\omega_2^2}F_{lj}(\mathbf{k}_2)\delta(\omega_1-\omega_2), \\ \tilde{D}_N^{(3)} &= -\sigma_{\mathbf{u}}^4\frac{\pi}{2}\int d\mathbf{k}_1 k_{1k}k_{1i}k_{1l}k_{2j}F_{ki}(\mathbf{k}_1) \\ &\times\int\frac{d\mathbf{k}_2}{\omega_2^2}F_{lj}(\mathbf{k}_2)\delta(\omega_1-\omega_2).\end{aligned} \tag{79}$$

Equation (78) is valid for isotropic and nonisotropic fluctuations of the velocity field. Consequently, in random isotropic compressible wave fields the probability distribution $P(t; \rho)$ is lognormal in this approximation and the impurity field should become clustered (see

[4–7]). In this case, for the coefficient $\tilde{D}_N^{(2)}$ allowing for formula (60) we obtain

$$\tilde{D}_N^{(2)} = \sigma_{\mathbf{u}}^4 \frac{\pi}{2N} \int \frac{d\mathbf{k}_1}{\omega^2(k_1)} k_1^4 F^p(k_1)$$

$$\times [F^s(k_1)(N-1) + F^p(k_1)] \int d\mathbf{k}_2 \delta(\omega_1 - \omega_2).$$

For the case of an anisotropic velocity field the solution of Eq. (78) is expressed in terms of the Airy function of the density logarithm. In this case, for low values of ρ the solution of the equation has negative values. However, the range of high densities and consequently the moments of the field $\rho(\mathbf{r}, t)$ are described correctly. Some change in the distribution function at high densities does not impede clustering of the impurity field.

5. CONCLUSIONS

Thus, allowance for the first nonvanishing corrections to the equation for the probability density of diffusing particles and the field of a passive conservative impurity in random wave fields yields nonzero transport coefficients. For compressible anisotropic wave velocity fields average particle transport (Stokes drift) occurs and the probability distribution of the position of Lagrangian particles becomes anisotropic. In this case, the field of a passive conservative impurity also undergoes clustering. However, it should be noted that these processes take place on different spatial scales which is expressed by different powers of the wave vectors $\mathbf{k}_i$ in the diffusion coefficients in Eqs. (56) and (78). For instance, small-scale fluctuations of the velocity field have a substantially stronger influence on impurity clustering in the Euler description compared with the diffusion of Lagrangian particles. If the wave field has a fairly broad spectrum, for example, exponentially damped at fairly high wave numbers, as is characteristic of turbulence, it is possible for divergence to appear in the expressions for the diffusion coefficients (79). In this case, the contribution of resonance effects to the diffusion coefficient (59) can be calculated.

ACKNOWLEDGMENTS

The authors are grateful to the referee for comments which resulted in a considerably larger introduction. However, the authors hope that the position of this study in the enormous field of analysis of stochastic dynamic systems has become clearer.

This work was partly supported by the Russian Foundation for Basic Research (project nos. 98-05-64479, 99-05-64350, and 00-15-98608).

REFERENCES

1. G. M. Zaslavskiĭ and R. Z. Sagdeev, *Introduction to the Nonlinear Physics* (Nauka, Moscow, 1988).
2. A. J. Lichtenberg and M. A. Lieberman, *Regular and Stochastic Motion* (Springer-Verlag, New York, 1982; Mir, Moscow, 1984).
3. V. I. Klyatskin, *Stochastic Equations and Waves in Randomly-Inhomogeneous Media* (Nauka, Moscow, 1980).
4. V. I. Klyatskin, Usp. Fiz. Nauk **164**, 531 (1994) [Phys. Usp. **37**, 501 (1994)].
5. V. I. Klyatskin and A. I. Saichev, Zh. Éksp. Teor. Fiz. **111**, 1297 (1997) [JETP **84**, 716 (1997)].
6. V. I. Klyatskin and D. Gurarie, Usp. Fiz. Nauk **169**, 171 (1999).
7. V. I. Klyatskin, Izv. Akad. Nauk, Fiz. Atmos. Okeana **36**, 177 (2000).
8. V. I. Klyatskin and V. I. Tatarskiĭ, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **14**, 1400 (1971).
9. V. I. Klyatskin, in *Lectures in Applied Mathematics*, Vol. 27: *Mathematics of Random Media*, Ed. by W. Kohler and B. S. White (American Mathematical Society, Providence, 1991), p. 447.
10. A. S. Monin and A. M. Yaglom, in *Statistical Fluid Mechanics* (Nauka, Moscow, 1965, 1967; MIT Press, Cambridge, Mass., 1971, 1975), Vols. 1, 2.
11. G. E. Ulenbek, Usp. Fiz. Nauk **103**, 275 (1971).
12. K. Herterich and R. J. Hasselmann, J. Phys. Oceanogr. **12**, 704 (1982).
13. B. J. Sanderson and A. Okubo, J. Geophys. Res. **93**, 3570 (1988).
14. R. L. Walterscheid and W. K. Hocking, J. Atmos. Sci. **48**, 2213 (1991).
15. G. I. Taylor, Proc. London Math. Soc. **20**, 196 (1921).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 4, 2000, pp. 748–760.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 4, 2000, pp. 863–876.

PLASMA, GASES

Neutrino Interaction with Strongly Magnetized Electron–Positron Plasma

A. V. Kuznetsov* and N. V. Mikheev**

Yaroslavl State University, Yaroslavl, 150000 Russia
**e-mail: avkuzn@uniyar.ac.ru*
***e-mail: mikheev@yars.free.net*

Received April 27, 2000

Abstract—A study is made of a complete set of neutrino–electron processes in a magnetized plasma. It is shown that processes involving neutrinos in the initial and final states $\nu e^{\mp} \longrightarrow \nu e^{\mp}$ and $\nu \longleftrightarrow \nu e^{-} e^{+}$ have kinematic amplification in the ultrarelativistic limit. Relatively simple expressions are obtained for the probability and average neutrino energy–momentum loss which are convenient for quantitative analysis. It is observed that the total contribution of νe processes did not depend on the chemical potential of the magnetized electron–positron plasma. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Over the last few decades one of the most rapidly developing physical sciences has been cosmoparticle physics which lies at the junction between the physics of elementary particles, astrophysics, and cosmology [1, 2]. The most important stimulus to its development was understanding the important role of quantum processes in the dynamics of astrophysical objects and also in the early universe. However, the extreme physical conditions existing inside these objects and, specifically, the presence of a hot dense plasma and strong electromagnetic fields should have a strong influence on quantum processes. As a result, studies of the interactions of elementary particles in an external active medium have attracted ongoing interest.

So far, essentially one-dimensional problems have been solved in astrophysical calculations of processes such as supernova explosions, and analyses of the influence of the active medium on quantum processes have only contained the plasma contribution. However, serious arguments have been put forward to suggest that the physics of supernovas is considerably more complex. In particular, we need to allow for rotation of the shell and also for the possible existence of a strong magnetic field, with these two phenomena being interrelated. In fact, the magnetic field generated during the collapse of a supernova nucleus may reach the critical Schwinger value $B_e = m_e^2/e \simeq 4.41 \times 10^3$ G.[1] The presence of rotation may increase the magnetic field by an additional factor of 10^3–10^4 [3].

In astrophysical phenomena such as stellar collapse, the absence of strong magnetic fields is an exotic rather than a typical case. It is appropriate to discuss the following set of questions.

(1) Which can be considered to be the more exotic object: a star possessing a magnetic field or a star without such a field? As far as we know from astrodynamics, a star without a magnetic field should be taken as an exotic rather than a typical case. In exactly the same way the presence of a primary magnetic field may be considered natural for a presupernova. As we know, a primary magnetic field of 100 G leads to the generation of a field on the scale of 10^{12}–10^{13} G during the collapse process as a result of the conservation of magnetic flux.

(2) Which can be considered to be the more typical case: a star possessing rotation or a star without rotation? Evidently a star without rotation appears to be the more exotic object.

(3) Which type of collapse looks more exotic: compression without or with an angular velocity gradient? Since the velocities at the edge of a compressible astrophysical object may reach relativistic scales, compression with differential rotation, i.e., with an angular velocity gradient, seems more probable.

All these factors are required to achieve the Bisnovatyi-Kogan scenario for the rotational explosion of a supernova [3]. The main component of this scenario is that the initially poloidal magnetic field lines of a field of 10^{12}–10^{13} G are twisted and compacted as a result of the angular velocity gradient to form an almost toroidal field of 10^{15}–10^{17} G.

We stress that this field is in fact a very dense medium having the mass density

$$\rho = \frac{B^2}{8\pi} \simeq 0.4 \times 10^{10} \frac{\mathrm{g}}{\mathrm{cm}^3} \left(\frac{B}{10^{16}\ \mathrm{G}} \right)^2, \qquad (1.1)$$

[1] We use the natural system of units $c = \hbar = 1$. Everywhere in the article $e > 0$ is the elementary charge.

which is comparable with the characteristic mass density of the shell of an exploding supernova, 10^{10}–10^{12} g/cm^3. Thus, in detailed studies of astrophysical processes such as supernova collapse it is absolutely essential to take into account the influence of the complex active medium including the plasma and the magnetic field.

We know that neutrino physics plays a decisive role in astrophysical cataclysms such as supernova explosions and coalescence of neutron stars, and also in the early universe. Consequently, studies of neutrino interactions and in particular neutrino–electron processes in an external active medium are of considerable interest. At the same time, an investigation of neutrino processes under such extreme physical conditions is interesting from the conceptual viewpoint since it affects fundamental problems of quantum field theory.

The first studies of neutrino–electron processes in an external electromagnetic field were devoted to the "synchrotron" radiation of neutrino pairs, $e \longrightarrow e\nu\bar{\nu}$ [4] and neutrino creation of electron–positron pairs, $\nu \longrightarrow \nu e^- e^+$ [5]. The analysis was made under conditions of a so-called relatively weak magnetic field when the initial particle energy is the dominant parameter, $E^2 \gg eB$; in fact, this limit corresponds to the crossed field approximation. In our studies [6, 7] the $\nu \longrightarrow \nu e^- e^+$ process was investigated for arbitrary values of the magnetic field and in particular in the strong field limit $E^2 \gg eB$ when an electron and a positron can only be created in states corresponding to the Landau ground level. A canonical neutrino–electron interaction channel, $\nu e^- \longrightarrow \nu e^-$ scattering, was investigated in [8] under conditions of a degenerate electron plasma taking into account the influence of a relatively weak magnetic field.

In the present study we show that a correct analysis of the neutrino propagation process in a hot dense plasma in the presence of a strong magnetic field requires us to consider the complete set of neutrino–electron processes. Specifically, in addition to the $\nu e^{\mp} \longrightarrow \nu e^{\mp}$ scattering reactions which also take place in the absence of a field, and the $\nu \longrightarrow \nu e^- e^+$ pair creation process which is only possible in a magnetic field, we also need to take into account the "exotic" process when a neutrino captures an electron–positron pair from the plasma: $\nu e^- e^+ \longrightarrow \nu$. This process is only allowed when both a magnetic field and a plasma are present. Then only the probability of the process summed over all initial states of the plasma electrons and positrons is physically meaningful. The probability of the $\nu e^{\mp} \longrightarrow \nu e^{\mp}$ scattering channels is defined similarly as the sum over all e^- or e^+ initial states. The total probability of neutrino interaction with an electron–positron plasma in a magnetic field is made up of the probabilities of these processes.

The article is constructed as follows. In Section 2 we describe the concept of a strongly magnetized electron–positron plasma and justify the feasibility of these physical processes occurring under the physical conditions during stellar collapse. In Section 3 we obtain the total amplitude of the neutrino–electron processes under conditions of a strongly magnetized electron–positron plasma in the local $\nu\nu ee$ interaction limit derived from the standard model of electroweak interactions. A kinematic analysis is made and it is shown that those neutrino–electron processes in which a neutrino is present in the initial and final states: $\nu e^{\mp} \longrightarrow \nu e^{\mp}$ and $\nu \longleftrightarrow \nu e^- e^+$ exhibit kinematic amplification. Section 4 is devoted to a detailed description of the procedure for calculating the probability of the $\nu \longleftrightarrow \nu e^- e^+$ process in a strongly magnetized electron–positron plasma. In Section 5 we give the probabilities of other νe processes and obtain the total probability of neutrino interaction with a magnetized $e^- e^+$ plasma. In Section 6 we calculate the average energy and momentum losses of a neutrino propagating through a magnetized plasma. In Section 7 we calculate the characteristics of the integral action of a neutrino on a magnetized plasma: the volume density of the energy transferred from the neutrino to the plasma per unit time, $\dot{\mathscr{E}}$, and the volume density of the force acting on the plasma from the neutrino. An analysis is made of possible astrophysical manifestations of neutrino–electron processes under these extreme physical conditions. It is shown that these processes may be important for a detailed description of the evolution of astrophysical objects.

2. WHAT WE UNDERSTAND BY STRONGLY MAGNETIZED $e^- e^+$ PLASMA

Here we are talking of conditions where, among all the physical parameters characterizing an electron–positron plasma, the field parameter is the dominant one. These conditions can be characterized simply by the relationship: $eB \gg \mu^2, T^2$, where μ is the chemical potential of the electrons and T is the plasma temperature. In order to find a better substantiated relationship we compare the energy densities of the magnetic field $B^2/8\pi$ and the electron–positron plasma.

As we know, a magnetic field changes the statistical properties of an electron–positron gas [9]. Taking into account degeneracy of the transverse momentum, the dependences of the concentration and energy density of an electron–positron gas on the chemical potential and temperature are described by the following sums over Landau levels:

$$n = n_{e^-} - n_{e^+} = \frac{eB}{2\pi^2}\int_0^\infty dp\,[\Phi(p, \mu, T) - \Phi(p, -\mu, T)], \tag{2.1}$$

$$\mathcal{E} = \mathcal{E}_{e^-} + \mathcal{E}_{e^+} = \frac{eB}{2\pi^2}\int_0^\infty p\,dp[\Phi(p, \mu, T) + \Phi(p, -\mu, T)], \tag{2.2}$$

$$\Phi(p, \mu, T) = \left[\exp\left(\frac{p-\mu}{T}\right) + 1\right]^{-1} + 2\sum_{k=1}^{\infty}\left[\exp\left(\frac{\sqrt{p^2 + 2keB} - \mu}{T}\right) + 1\right]^{-1}. \tag{2.3}$$

Here we used the approximation of an ultrarelativistic electron–positron gas since astrophysical processes are characterized by fairly high neutrino and plasma electron energies $E \gg m_e$. Thus, we shall neglect the electron mass wherever this causes no misunderstandings.

In a strong field and specifically, when the condition $\sqrt{eB} - \mu \gg T$ is satisfied, in practice only the Landau ground level is filled. From (2.1) and (2.2) we then obtain

$$n = \frac{eB\mu}{2\pi^2}, \tag{2.4}$$

$$\mathcal{E} = \frac{eB\mu^2}{4\pi^2} + \frac{eBT^2}{12}. \tag{2.5}$$

Thus, a more exact condition that the electron–positron plasma is strongly magnetized may be written in the form

$$\frac{B^2}{8\pi} \gg \frac{\pi^2 n^2}{eB} + \frac{eBT^2}{12}. \tag{2.6}$$

Selecting values of the physical parameters typical of a supernova shell as scales in the relationship (2.6), we rewrite this in the form

$$0.8 \times 10^{32} B_3^2 \gg 1.7 \times 10^{30}\frac{\rho_{12}^2 Y_{0.1}^2}{B_3} + 1.1 \times 10^{27} B_3 T_5^2 \left(\frac{\text{erg}}{\text{cm}^3}\right), \tag{2.7}$$

where

$$B_3 = \frac{B}{10^3 B_e}, \quad \rho_{12} = \frac{\rho}{10^{12}\ \text{g/cm}^3}, \quad Y_{0.1} = \frac{Y_e}{0.1}, \quad T_5 = \frac{T}{5\ \text{MeV}}, \tag{2.8}$$

ρ is the total plasma density in the shell, and Y_e is the ratio of the number of electrons to the number of baryons. It can be seen that the plasma magnetization condition is definitely satisfied.

3. NEUTRINO–ELECTRON PROCESSES IN A STRONGLY MAGNETIZED PLASMA

3.1. Total Amplitude

Calculations will be made for the relatively small momentum transfers $|q^2| \ll m_W^2$, where m_W is the W boson mass. An analysis shows that when studying the processes in a magnetized plasma we need to add the conditions $m_W^2 \gg eB$ and $m_W^3 \gg eBT, eB\mu$. Weak neutrino–electron interaction can then be described in the local limit by the effective Lagrangian

$$\mathcal{L} = \frac{G_F}{\sqrt{2}}[\bar{e}\gamma_\alpha(g_V + g_A\gamma_5)e][\bar{\nu}\gamma^\alpha(1 + \gamma_5)\nu], \tag{3.1}$$

where $g_V = \pm 1/2 + 2\sin^2\theta_W$, $g_A = \pm 1/2$. Here the upper signs refer to an electron neutrino ($\nu = \nu_e$) when exchange with both Z and W bosons contributes to the process and the lower signs corresponds to muon and tau-neutrinos ($\nu = \nu_\mu, \nu_\tau$) when only exchange with a Z boson contributes to the Lagrangian (3.1).

The Lagrangian (3.1) is written assuming massless neutrinos and consequently the absence of mixing in the lepton sector. A generalization to the case of mass-possessing neutrinos with mixing can be found, e.g., in [10]. However, it should be noted that the kinematics of charged particles is a magnetic field is such that all interaction processes of high-energy neutrinos with electrons become insensitive to lepton mixing and are possible even in the massless neutrino limit. This implies that the flavor of an ultrarelativistic neutrino is conserved in these processes in a magnetic field to within terms of the order m_ν^2/E_ν^2 regardless of the lepton mixing angles which makes the question of neutrino mixing irrelevant.

The total amplitude for neutrino–electron processes is obtained directly from the Lagrangian (3.1) where known solutions of the Dirac equation in a magnetic field must be used for an electron and positron. As we know (see, e.g., [11]), these solutions for the Landau ground level may be expressed in the following form [the magnetic field is directed along the z-axis and the vector potential is taken in the form $\mathbf{A} = (0, xB, 0)$]:

(i) for an electron of energy ω and "momentum" k_y, k_z[2]

$$\psi_k = \frac{(eB)^{1/4}}{(\sqrt{\pi}2\omega L_y L_z)^{1/2}} \times \exp[-i(\omega t - k_y y - k_z z)]\exp\left(-\frac{\xi^2}{2}\right)u(k_\parallel), \tag{3.2}$$

[2] Here k_z is the kinetic momentum along the z-axis, and k_y is the generalized momentum which determines the position of the center of a Gaussian packet along the x axis, and $x_0 = -k_y/eB$, see (3.4).

(ii) for a positron of energy ω' and "momentum" k'_y, k'_z

$$\Psi_{k'} = \frac{(eB)^{1/4}}{(\sqrt{\pi}2\omega' L_y L_z)^{1/2}} \times \exp[-i(\omega' t - k'_y y - k'_z z)]\exp\left(-\frac{\xi'^2}{2}\right)u(k'_\parallel), \quad (3.3)$$

where $k_\parallel(k'_\parallel)$ is the electron (positron) energy–momentum vector in the Minkowski (0,3) plane. Here and subsequently L_x, L_y, and L_z are auxiliary parameters which determine the normalization volume $V = L_x L_y L_z$, $\omega = \sqrt{m^2 + k_z^2}$, $\omega' = \sqrt{m^2 + k_z'^2}$, m is the electron mass, and ξ and ξ' are dimensionless coordinates which describe the motion along the x-axis:

$$\xi = \sqrt{eB}\left(x + \frac{k_y}{eB}\right), \qquad \xi' = \sqrt{eB}\left(x - \frac{k'_y}{eB}\right). \quad (3.4)$$

The bispinor amplitudes have the form

$$u(k_\parallel) = \frac{1}{\sqrt{\omega + m}}\begin{pmatrix}(\omega + m)\Psi\\ -k_z\Psi\end{pmatrix}, \qquad u(k'_\parallel) = \frac{1}{\sqrt{\omega' - m}}\begin{pmatrix}(\omega' - m)\Psi\\ -k'_z\Psi\end{pmatrix}, \quad (3.5)$$

where $\Psi = \begin{pmatrix}0\\1\end{pmatrix}$ corresponds to a state whose spin is directed opposite to the field. It is interesting to note that the bispinor amplitudes (3.5) are exactly the same as the solutions of the free Dirac equation for an electron and positron having momenta directed along the z-axis. This separation of bispinor amplitudes which do not depend on the spatial coordinate x is only typical of the Landau ground level.

Using the Lagrangian (3.1) and the wave functions (3.2), (3.3), and (3.5), we write the S-matrix element of the process $\nu \longrightarrow \nu e^- e^+$ in the following form (the amplitudes of the other neutrino–electron processes are then obtained by crossing transformations):

$$S = i\frac{G_F}{\sqrt{2}}\frac{(2\pi)^3}{\sqrt{2EV2E'V2\omega L_y L_z 2\omega' L_y L_z}} \times \delta(\omega + \omega' - q_0)\delta(k_y + k'_y - q_y)\delta(k_z + k'_z - q_z) \times \exp\left(-\frac{q_\perp^2}{4eB} - \frac{iq_x(k_y - k'_y)}{2eB}\right)[\bar{u}(k_\parallel)\hat{j}(g_V + g_A\gamma_5)u(-k'_\parallel)], \quad (3.6)$$

where $q = p - p'$ is the change in the four-vector of the neutrino momentum, $q_\perp$ is the projection of the vector $\mathbf{q}$ on the plane perpendicular to the vector $\mathbf{B}$, $q_\perp^2 = q_x^2 + q_y^2$, $j_a = \bar{\nu}(p')\gamma_\alpha(1 + \gamma_5)\nu(p)$ is the Fourier transform of the current of left neutrinos. Note that in this approximation where the field is the largest energy parameter of the problem, the exponential factor $\exp(-q_\perp^2/4eB)$ in the amplitude (3.6) differs little from unity and may be omitted. Direct calculations using the bispinors (3.5) and taking into account the conservation laws in (3.6) give

$$[\bar{u}(k_\parallel)\hat{j}(g_V + g_A\gamma_5)u(-k'_\parallel)] = \frac{2m}{\sqrt{q_\parallel^2}}\frac{q_z}{|q_z|}[g_V(j\tilde{\varphi}q) + g_A(j\tilde{\varphi}\tilde{\varphi}q)], \quad (3.7)$$

where $\varphi_{\alpha\beta} = F_{\alpha\beta}/B$, $\tilde{\varphi}_{\alpha\beta} = \frac{1}{2}\varepsilon_{\alpha\beta\mu\nu}\varphi^{\mu\nu}$ are the dimensionless tensor and the dual tensor of the magnetic field, and $q_\parallel^2 = q_0^2 - q_z^2 = (q\tilde{\varphi}\tilde{\varphi}q)$. Inside the parentheses the tensor indices are contracted systematically, for example $(j\tilde{\varphi}\tilde{\varphi}q) = j_\alpha\tilde{\varphi}^{\alpha\beta}\tilde{\varphi}_{\beta\lambda}q^\lambda$.

Expression (3.7) and thus the total amplitude (3.6) for an arbitrary neutrino–electron process contain suppression associated with the relative smallness of the electron mass. This suppression is not random and reflects the angular momentum conservation law. For example, in the $\nu\bar{\nu} \longrightarrow e^-e^+$ process the total spin of a neutrino–antineutrino pair in the center-of-inertia system is one whereas the total spin of an electron–positron pair in the Landau ground level is zero. Consequently, the amplitude of the process would be zero for massless particle and contain suppression in the relativistic limit under study. However, an analysis shows that when integration is performed over the phase volume kinematic regions exist where $\sqrt{q_\parallel^2} \sim m$ and this suppression disappears for some neutrino–electron processes.

3.2. Kinematic Analysis

All neutrino–electron processes determined by the Lagrangian (3.1) can be divided into two groups.

1. Processes in which a neutrino is present in the initial and final states: $\nu e^{\mp} \longrightarrow \nu e^{\mp}$, $\nu \longrightarrow \nu e^- e^+$, $\nu e^- e^+ \longrightarrow \nu$ and similar antineutrino processes.

2. Processes involving creation or absorption of a neutrino–antineutrino pair $e^-e^+ \longrightarrow \nu\bar{\nu}$, $\nu\bar{\nu} \longrightarrow e^-e^+$, $e \longrightarrow e\nu\bar{\nu}$, and $e\nu\bar{\nu} \longrightarrow e$.

It can be seen from (3.7) that the square of the amplitude of each neutrino–electron process contains the factor $m^2/q_\parallel^2$. However, the value of $q_\parallel^2 = q_0^2 - q_z^2$ differs fundamentally for processes of the first and second types. For processes with a neutrino–antineutrino

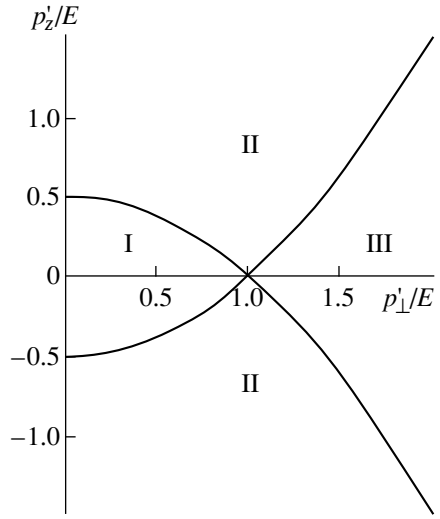

Fig. 1. Kinematic regions in momentum space of a final neutrino: I—for the $\nu \longrightarrow \nu e^- e^+$ pair creation process; II—for the $\nu e^- \longrightarrow \nu e^-$, $\nu e^+ \longrightarrow \nu e^+$ scattering channels; III—for the $\nu e^- e^+ \longrightarrow \nu$ pair capture process. The lines correspond to $q_\parallel^2 = 0$,

pair we have $q = p + p'$ (p and p' are the four-momenta of a neutrino and an antineutrino, respectively), and consequently $q^2 > 0$. Since $q_\parallel^2 = q^2 + q_\perp^2$, where both terms are positive, the value of $q_\parallel^2$ can only be small when both q^2 and $q_\perp^2$ are small which is only possible in a small region of phase space. This implies that almost everywhere in phase space $\sqrt{q_\parallel^2} \sim E \sim T \gg m$ which leads to reduction of the probability by a factor $m^2/T^2 \ll 1$.

At the same time, we have $q = p - p'$ for processes involving neutrinos in the initial and final states and consequently $q^2 < 0$ and the value of $q_\parallel^2$ may be small over a fairly wide region of phase space. Calculations confirm that kinematic amplification is achieved for these processes, leading to the disappearance of the factor m^2/T^2 in the probabilities.

Hence, neutrino interaction with a strongly magnetized electron–positron plasma is determined by the processes $\nu e^- \longrightarrow \nu e^-$, $\nu e^+ \longrightarrow \nu e^+$, $\nu \longrightarrow \nu e^- e^+$, and $\nu e^- e^+ \longrightarrow \nu$. Figure 1 shows kinematic regions in momentum space of a finite neutrino for the processes described above in a convenient frame of reference where the momentum of the initial neutrino is perpendicular to the magnetic field. The main contribution to the probability comes from regions near the parabola $q_\parallel^2 = 0$ where this kinematic amplification takes place.

4. PROBABILITY OF THE $\nu \longrightarrow \nu e^- e^+$ PROCESS

We express the probability of neutrino creation of an $e^- e^+$ pair per unit time in the form

$$W(\nu \longrightarrow \nu e^- e^+) = \frac{1}{\mathcal{T}} \int |S|^2 d\Gamma_{e^-}(1 - f_{e^-}) d\Gamma_{e^+}(1 - f_{e^+}) d\Gamma'_\nu (1 - f'_\nu), \tag{4.1}$$

where $\mathcal{T}$ is the total interaction time, $d\Gamma$ is an element of the particle phase volume,

$$d\Gamma_{e^-} = \frac{d^2 k L_y L_z}{(2\pi)^2}, \quad d\Gamma_{e^+} = \frac{d^2 k' L_y L_z}{(2\pi)^2}, \quad d\Gamma'_\nu = \frac{d^3 p' V}{(2\pi)^3}. \tag{4.2}$$

The electron and positron distribution functions

$$f_{e^-} = \frac{1}{\exp[(\omega - \mu)/T] + 1}, \quad f_{e^+} = \frac{1}{\exp[(\omega' + \mu)/T] + 1}, \tag{4.3}$$

allow for the presence of a plasma; here μ and T are the chemical potential and temperature of the electron–positron gas. To be general, we also allowed for the possible presence of a quasiequilibrium neutrino gas described by the distribution function f'_ν. In general, the question of the accuracy of the description of the state of a neutrino gas under conditions of stellar collapse or another astrophysical process using an equilibrium distribution function and also the determination of this function is a complex astrophysical problem (see, e.g., [12]). Quite clearly, the approximation of an equilibrium neutrino Fermi gas using the distribution function

$$f_\nu^{(eq)} = \frac{1}{\exp[(E - \mu_\nu)/T_\nu] + 1}, \tag{4.4}$$

where μ_ν and T_ν are the chemical potential and the temperature of the neutrino gas, should give satisfactory results inside the neutrinosphere. Outside the neutrinosphere, where an outgoing neutrino flux forms and the neutrino momenta become asymmetric, a factorization of the local distribution is usually assumed

$$f_\nu = \frac{\Phi(\vartheta, R)}{\exp[(E - \mu_\nu)/T_\nu] + 1}, \tag{4.5}$$

where the energy distribution is assumed to be approximately equilibrium, the function $\Phi(\vartheta, R)$ determines the neutrino angular distribution, $\vartheta = \cos\alpha$, α is the angle between the neutrino momentum and the radial direction in the star, and R is the distance from the center of the star. An analysis shows [12] that in the vicinity of the neutrinosphere the function $\Phi(\vartheta, R)$ differs negligibly from unity. In order to calculate the proba-

bility we shall use the neutrino distribution function in the form (4.4), neglecting the asymmetry. Later in Section 7 when analyzing possible astrophysical manifestations of these neutrino–electron processes, we shall also allow for asymmetry in the distribution function (4.5) for the initial and final neutrinos.

Substituting (3.6) into (4.1) and integrating using δ-functions with respect to d^2k' [where, as is usually the case $\delta^3(0) = \mathcal{T}L_yL_z/(2\pi)^3$], we obtain

$$W = \frac{G_F^2}{32(2\pi)^4 E}\frac{1}{L_x}\int\frac{d^3p'}{E'}(1-f'_\nu)\frac{dk_y dk_z}{\omega^2}\delta(\omega+\omega'-q_0)$$
$$\times(1-f_{e^-})(1-f_{e^+})\left|\bar{u}(k_\parallel)\hat{j}(g_V+g_A\gamma_5)u(-k'_\parallel)\right|^2, \tag{4.6}$$

where we need to substitute $\omega' = \sqrt{m^2+(q_z-k_z)^2}$, $k'_z = q_z - k_z$. It is easy to see that the expression in the integrand in (4.6) does not depend on k_y and consequently integration over k_y essentially determines the degree of degeneracy of an electron having a given energy [see footnote to Eq. (3.2)]

$$N_E = \frac{L_y}{2\pi}\int dk_y = \frac{eBL_y}{2\pi}\int\limits_{-L_x/2}^{L_x/2} dx_0 = \frac{eBL_xL_y}{2\pi}. \tag{4.7}$$

Integrating over the electron momentum in (4.6) taking into account (4.7) we obtain the probability of the $\nu \longrightarrow \nu e^- e^+$ process in the form of the following integral over the final neutrino momentum:

$$W = \frac{G_F^2 eBm^2}{64\pi^4 E}\int\frac{d^3p'}{E'}\theta(q_0-\sqrt{q_z^2+4m^2})$$
$$\times\frac{1}{(q_\parallel^2)^{3/2}(q_\parallel^2-4m^2)^{1/2}}\left|g_V(j\tilde{\varphi}q)+g_A(j\tilde{\varphi}\tilde{\varphi}q)\right|^2 \tag{4.8}$$
$$\times(1-f'_\nu)[(1-f_{e^-})(1-f_{e^+})+(q_z\longrightarrow -q_z)].$$

In this expression the electron and positron energies ω and ω' appearing in the distribution function $f_{e^\mp}$ are determined by the conservation law $\omega+\omega'-q_0=0$ and are given by

$$\omega = \frac{1}{2}\left(q_0+q_z\sqrt{1-\frac{4m^2}{q_\parallel^2}}\right),$$
$$\omega' = \frac{1}{2}\left(q_0-q_z\sqrt{1-\frac{4m^2}{q_\parallel^2}}\right). \tag{4.9}$$

Expression (4.8) is a generalization of Eq. (3.2) from our study [7], where we investigated the neutrino–electron process $\nu \longrightarrow \nu e^- e^+$ in a high-intensity purely magnetic field, to the case where electron–positron and neutrino gases are present.

Further integration over the final neutrino momentum can be conveniently performed in a reference frame where the initial neutrino momentum is perpendicular to the magnetic field $p_z = 0$. For the case of a purely magnetic field we could convert to this system without any loss of generality by performing a Lorentz transformation parallel to the field. In fact, we can see that, besides the statistical Fermi factors, the value of E_W determined from Eq. (4.8) only contains invariants with respect to this transformation (including the sign of the argument of the θ function). However, we now have an isolated reference frame, the plasma rest system, in which the distribution functions (4.3) and (4.5) are formulated. In order to convert to a system where $p_z = 0$ we express these functions in a relativistically invariant form:

$$f_{e^-} = \left[\exp\left(\frac{(kv)-\mu}{T}\right)+1\right]^{-1},$$
$$f_{e^+} = \left[\exp\left(\frac{(k'v)+\mu}{T}\right)+1\right]^{-1}, \tag{4.10}$$
$$f'_\nu = \left[\exp\left(\frac{(p'v)-\mu_\nu}{T_\nu}\right)+1\right]^{-1}.$$

Here we introduce the four-vector of the plasma velocity v^α, $v^2 = 1$ which in its rest system is $v^\alpha = (1, \mathbf{0})$ and the distribution functions (4.10) are exactly the same as the functions (4.3) and (4.5). In the system $p_z = 0$ we have

$$v^\alpha = (v_0, 0, 0, v_z), \quad v_0 = 1/\sin\theta,$$
$$v_z = -\cos\theta/\sin\theta,$$

where θ is the angle between the vectors of the initial neutrino momentum and the magnetic field induction in the plasma rest system.

In Eq. (4.8) it is convenient to convert the dimensionless cylindrical coordinates in the space of the final neutrino momentum vector $\mathbf{p}'$:

$$\rho = \frac{\sqrt{p_x'^2+p_y'^2}}{E_\perp}, \quad \tan\phi = \frac{p'_y}{p'_x}, \quad \zeta = \frac{p'_z}{E_\perp}.$$

Here $E_\perp$ is the energy of the initial neutrino in the system $p_z = 0$ which is related to its energy E in the plasma rest system by $E_\perp = E\sin\theta$. In terms of the variables ρ, ζ, expression (4.8) is rewritten in the form

$$EW = \frac{G_F^2 m^2 eBE_\perp^2}{4\pi^3}$$
$$\times\int\limits_0^{1-\lambda} d\rho\,\rho\int\limits_{-\zeta_m}^{\zeta_m}\frac{d\zeta}{\beta\sqrt{\rho^2+\zeta^2}(1-2\sqrt{\rho^2+\zeta^2}+\rho^2)^2}$$

$$\times\{(g_V^2+g_A^2)[(1+\rho^2)\sqrt{\rho^2+\zeta^2}-2\rho^2] \qquad (4.11)$$

$$-2g_Vg_A(1-\rho^2)\zeta\}(1-f'_\nu)$$

$$\times[(1-f_{e^-})(1-f_{e^+})\big|_{\sigma=+1}+(1-f_{e^-})(1-f_{e^+})\big|_{\sigma=-1}],$$

where we need to substitute in the distribution functions (4.10)

$$(kv)=\frac{E_\perp}{2\sin\theta}$$

$$\times[(1-\sqrt{\rho^2+\zeta^2})(1+\sigma\beta\cos\theta)-\zeta(\cos\theta+\sigma\beta)],$$

$$(k'v)=\frac{E_\perp}{2\sin\theta}$$

$$\times[(1-\sqrt{\rho^2+\zeta^2})(1-\sigma\beta\cos\theta)-\zeta(\cos\theta-\sigma\beta)],$$

$$(p'v)=\frac{E_\perp}{\sin\theta}(\sqrt{\rho^2+\zeta^2}+\zeta\cos\theta),$$

and also introduce the notation

$$\beta=\sqrt{1-\frac{4m^2}{q_\parallel^2}}=\sqrt{1-\frac{\lambda^2}{(1-2\sqrt{\rho^2+\zeta^2}+\rho^2)^2}},$$

$$\lambda=\frac{2m}{E_\perp},\quad \zeta_m=\frac{1}{2}\sqrt{(1+\rho^2-\lambda^2)^2-4\rho^2}.$$

Note that the expression in the integrand in (4.11) exhibits a gain which completely compensates for the reduction by the smallness of the electron mass. The main contribution then comes from the region near the upper limits of the integrals over ρ, ζ corresponding to the values $\sqrt{q_\parallel^2}\sim m$. Converting to the new integration variables β and $x=E_\perp(1-\rho^2)/4T\sin\theta$ in Eq. (4.11) and isolating the leading contribution $\sim E_\perp^2/m^2$, we transform the expression for the probability to the form

$$EW=\frac{G_F^2eBE_\perp^2T^2\sin^2\theta}{2\pi^3}$$

$$\times\int_0^{\varepsilon\tau/4}xdx\int_0^1d\beta\left\{\frac{(g_V+g_A)^2}{1+\exp[-\varepsilon+2x(1+u)/\tau+\eta_\nu]}\right.$$

$$\times[f(\beta,u,\eta)f(-\beta,u,-\eta)+f(\beta,u,-\eta)f(-\beta,u,\eta)]$$

$$+\frac{(g_V-g_A)^2}{1+\exp[-\varepsilon+2x(1-u)/\tau+\eta_\nu]} \qquad (4.12)$$

$$\left.\times[f(\beta,-u,\eta)f(-\beta,-u,-\eta)+f(\beta,-u,-\eta)f(-\beta,-u,\eta)]\right\},$$

where

$$\eta=\frac{\mu}{T},\quad \eta_\nu=\frac{\mu_\nu}{T_\nu},\quad \varepsilon=\frac{E_\perp}{T_\nu\sin\theta},$$

$$u=\cos\theta,\quad \tau=T_\nu/T,$$

$$f(\beta,u,\eta)=\frac{1}{1+\exp[-x(1+\beta)(1+u)+\eta]}.$$

Integrating (4.12) with respect to the variable β using the relationship

$$\int_0^1d\beta f(\beta,u,\eta)f(-\beta,u,-\eta)$$

$$=\frac{1}{a(1-e^{-2a})}\ln\left(\frac{1+e^{-2a+\eta}}{1+e^{\eta}}\frac{1+e^{a+\eta}}{1+e^{-a+\eta}}\right), \qquad (4.13)$$

where $a=x(1+u)$ and converting to the plasma rest system, we finally obtain

$$W(\nu\longrightarrow\nu e^-e^+)=\frac{G_F^2eBT^2E}{4\pi^3}\left\{(g_V+g_A)^2(1-u)^2\right.$$

$$\times\int_0^{\varepsilon\tau(1+u)/2}\frac{d\xi}{(1-e^{-\xi})[1+\exp(-\varepsilon+\xi/\tau+\eta_\nu)]}$$

$$\times\ln\frac{\cosh\xi+\cosh\eta}{1+\cosh\eta}+(g_V-g_A)^2(1+u)^2 \qquad (4.14)$$

$$\times\int_0^{\varepsilon\tau(1-u)/2}\frac{d\xi}{(1-e^{-\xi})[1+\exp(-\varepsilon+\xi/\tau+\eta_\nu)]}$$

$$\left.\times\ln\frac{\cosh\xi+\cosh\eta}{1+\cosh\eta}\right\},$$

where $\varepsilon=E/T_\nu$. The dependence of the probability (4.14) on the electron–positron gas concentration $n=n_{e^-}-n_{e^+}$ is defined in terms of its chemical potential [see (2.4)]. Note that the formula for the probability (4.14) holds for hot ($\mu\ll T$) and cold ($\mu\gg T$) plasmas. For low-density electron–positron and neutrino gases (T, μ, T_ν, $\mu_\nu\longrightarrow 0$), Eq. (4.14) reproduces our result [6, 7] for the probability of the process $\nu\longrightarrow\nu e^-e^+$ in the strong magnetic field limit $eB\gg E^2\sin^2\theta$ without a plasma:

$$W_B=\frac{G_F^2(g_V^2+g_A^2)}{16\pi^3}eBE^3\sin^4\theta. \qquad (4.15)$$

In the absence of a neutrino gas (T_ν, $\mu_\nu \longrightarrow 0$) the expression for the probability (4.14) for a hot electron–positron plasma ($T \longrightarrow \infty$) has the form $W_B/4$ as we indicated in [7] since the statistical factors for an electron and positron in this limit are 1/2.

5. TOTAL PROBABILITY OF NEUTRINO INTERACTION WITH A MAGNETIZED ELECTRON–POSITRON PLASMA

As we noted in the Introduction, the influence of the $\nu e^\mp \longrightarrow \nu e^\mp$ scattering and $\nu e^- e^+ \longrightarrow \nu$ pair capture channels on the neutrino propagation process in a plasma should be taken into account in terms of the probabilities summed over initial electron and/or positron states. Thus, the probabilities of scattering processes should be defined as

$$W(\nu e^\mp \longrightarrow \nu e^\mp) = \frac{1}{\mathcal{T}}\int |S|^2 d\Gamma_{e^\mp} f_{e^\mp} d\Gamma'_{e^\mp}(1 - f'_{e^\mp}) d\Gamma'_\nu (1 - f'_\nu). \tag{5.1}$$

Similarly for the pair capture process

$$W(\nu e^- e^+ \longrightarrow \nu) = \frac{1}{\mathcal{T}}\int |S|^2 d\Gamma_{e^-} f_{e^-} d\Gamma_{e^+} f_{e^+} d\Gamma'_\nu (1 - f'_\nu). \tag{5.2}$$

It can be seen from Fig. 1 that the scattering and pair capture processes correspond to infinite kinematic regions since the initial electrons and positrons can formally have any energy. Convergence of the integrals is ensured by the distribution functions.

The expressions (5.1) and (5.2) are integrated by the same scheme as that described above for the $\nu \longrightarrow \nu e^- e^+$ pair creation process. An important factor for the integration will be that the energy imparted from the neutrino to the active medium $q_0 = E - E'$ is not positive-definite. For the probability (per unit time) of neutrino scattering on magnetized plasma electrons we have

$$W(\nu e^- \longrightarrow \nu e^-) = \frac{G_F^2 eBT^2 E}{4\pi^3}\left\{(g_V + g_A)^2(1-u)^2 \times \int_0^{\varepsilon\tau(1+u)/2} \frac{d\xi}{(1 - e^{-\xi})[1 + \exp(-\varepsilon + \xi/\tau + \eta_\nu)]} \times \ln\frac{1 + e^\eta}{1 + e^{-\xi+\eta}} + (g_V - g_A)^2(1+u)^2 \times \int_0^{\varepsilon\tau(1-u)/2} \frac{d\xi}{(1 - e^{-\xi})[1 + \exp(-\varepsilon + \xi/\tau + \eta_\nu)]} \right. \tag{5.3}$$
$$\left. \times \ln\frac{1 + e^\eta}{1 + e^{-\xi+\eta}} + [(g_V + g_A)^2(1-u)^2 + (g_V - g_A)^2(1+u)^2] \times \int_0^\infty \frac{d\xi}{(e^\xi - 1)[1 + \exp(-\varepsilon - \xi/\tau + \eta_\nu)]} \ln\frac{1 + e^\eta}{1 + e^{-\xi+\eta}} \right\}.$$

Taking into account the distribution functions (4.10), the probability of scattering at positrons is obtained from (5.3) by substituting $\eta \longrightarrow -\eta$. For the pair capture channel we have

$$W(\nu e^- e^+ \longrightarrow \nu) = \frac{G_F^2 eBT^2 E}{4\pi^3} \times [(g_V + g_A)^2(1-u)^2 + (g_V - g_A)^2(1+u)^2] \times \int_0^\infty \frac{d\xi}{(e^\xi - 1)[1 + \exp(-\varepsilon - \xi/\tau + \eta_\nu)]} \times \ln\frac{\cosh\xi + \cosh\eta}{1 + \cosh\eta}. \tag{5.4}$$

As we have already noted, only the total probability of neutrino interaction with an electron–positron plasma is physically meaningful:

$$W(\nu \longrightarrow \nu) = W(\nu \longrightarrow \nu e^- e^+) + W(\nu e^- e^+ \longrightarrow \nu) + W(\nu e^- \longrightarrow \nu e^-) + W(\nu e^+ \longrightarrow \nu e^+). \tag{5.5}$$

It was found that this quantity has a substantially simpler form:

$$W(\nu \longrightarrow \nu) = \frac{G_F^2 eBT^2 E}{4\pi^3}\left\{(g_V + g_A)^2(1-u)^2 \times \left[F_1\left(\frac{\varepsilon\tau(1+u)}{2}\right) - F_1(-\infty)\right] + (g_A \longrightarrow -g_A;\ u \longrightarrow -u)\right\}, \tag{5.6}$$

where

$$F_k(z) = \int_0^z \frac{\xi^k d\xi}{(1 - e^{-\xi})[1 + \exp(-\varepsilon + \eta_\nu + \xi/\tau)]}, \tag{5.7}$$

and significantly, the dependence on the chemical potential μ of the electron–positron gas, which was present in the probabilities of the various processes, was cancelled in the total probability. At present, a physical cause of this reduction is unknown. Possibly some property involving the completeness of this set of processes in relation to electrons is manifest.

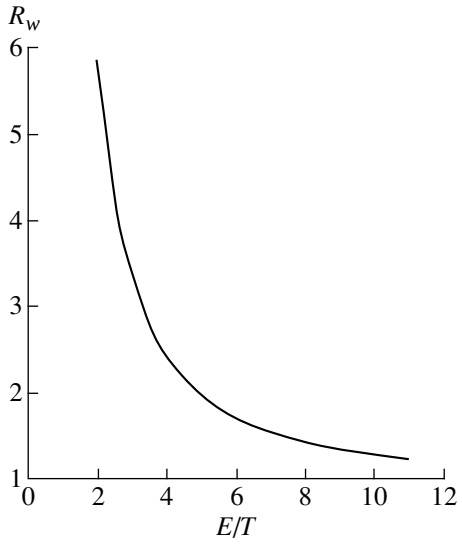

Fig. 2. Ratio of the probabilities of neutrino interaction with a magnetized plasma and a pure magnetic field, $R_w = W_{B+pl}/W_B$ for $\theta = \pi/2$ as a function of the ratio of the neutrino energy to the plasma temperature.

For a sparse neutrino gas the probability (5.6) is expressed in terms of the dilogarithm $\mathrm{Li}_2(x)$:

$$W(\nu \longrightarrow \nu) = \frac{G_F^2 eBT^2 E}{4\pi^3}\left\{(g_V^2 + g_A^2)\frac{E^2\sin^4\theta}{4T^2}\right.$$
$$+ (g_V + g_A)^2(1-\cos\theta)^2$$
$$\times \mathrm{Li}_2\left[1 - \exp\left(-\frac{E(1+\cos\theta)}{2T}\right)\right] \qquad (5.8)$$
$$+ (g_V - g_A)^2(1+\cos\theta)^2\mathrm{Li}_2\left[1 - \exp\left(-\frac{E(1-\cos\theta)}{2T}\right)\right]$$
$$\left. + \frac{\pi^2}{3}[(g_V^2 + g_A^2)(1+\cos^2\theta) - 4g_V g_A\cos\theta]\right\},$$

where $\mathrm{Li}_n(x)$ is an nth-order polylogarithm:

$$\mathrm{Li}_n(x) = \sum_{k=1}^{\infty}\frac{x^k}{k^n}. \qquad (5.9)$$

The relative contributions of the plasma and the magnetic field to the process of neutrino interaction with the active medium are illustrated in Fig. 2 which gives the ratio of the probabilities of neutrino interaction with a magnetized plasma and a pure magnetic field $R_w = W_{B+pl}/W_B$ for the angle $\theta = \pi/2$ as a function of the ratio of the neutrino energy to the plasma temperature. It can be seen that as the temperature increases, the interaction probability increases.

The probability (5.6) determines the partial contribution of these processes to the opacity for neutrino propagation in a medium. An estimate of the mean free path associated with neutrino–electron processes gives

$$\lambda_e = \frac{1}{W} \simeq 170\ \mathrm{km}\left(\frac{10^3 B_e}{B}\right)\left(\frac{5\ \mathrm{MeV}}{T}\right)^3. \qquad (5.10)$$

This should be compared with the neutrino mean free path as a result of interaction with nucleons, which is of the order of a kilometer at density $\rho \sim 10^{12}$ g/cm^3. At first glance the influence of neutrino–electron reactions on the neutrino propagation process is negligible. However, the mean free path does not exhaust the neutrino physics in the medium. Other important quantities in astrophysical applications are the neutrino energy and momentum losses. Of particular importance is the asymmetry of the neutrino momentum loss caused by the influence of an external magnetic field. Many attempts have been made to calculate these asymmetries caused by neutrino–nucleon processes associated with the problem of the high proper velocities of pulsars (see [13] and the references therein). As we shall show, despite the relatively low probability of neutrino–electron processes, their contribution to the asymmetry may be comparable to the contributions of neutrino–nucleon processes.

6. AVERAGE LOSSES OF NEUTRINO ENERGY AND MOMENTUM

In studies of these neutrino–electron interactions in a magnetic field and/or plasma [5, 8], the analysis has usually been confined to calculation of the probabilities and cross sections of processes. As we have noted, not only the probabilities of the processes are of practical interest for astrophysics but also the average loss of neutrino energy and momentum in the medium[3] which is determined by the four-vector

$$Q^\alpha = -E\left(\frac{dE}{dt}, \frac{d\mathbf{p}}{dt}\right) = E\int q^\alpha dW, \qquad (6.1)$$

where E and $\mathbf{p}$ are the energy and momentum of the initial neutrino, q is the difference between the momenta of the initial and final neutrinos, $q = p - p'$, and dW is the total differential probability of all the processes specified in (5.5). The zeroth component Q^0 is associated with the average energy lost by a single neutrino per unit time and the spatial components $\mathbf{Q}$ are associated with the loss of neutrino momentum per unit time.

For a purely magnetic field the four-vector of the losses Q^α was calculated in our studies [6,7]. In this case, the losses are caused by the only possible process in the absence of plasma, pair creation during motion of a neutrino in a strong magnetic field $\nu \longrightarrow \nu e^- e^+$. In the

[3] In general a neutrino can lose and acquire energy and momentum so that we shall subsequently understand "loss" of energy and momentum in the algebraic sense.

strong magnetic field limit for the zeroth and z-components of the vector Q^α we obtained (the field is directed along z)

$$Q_{0,z}^{(B)} = \frac{G_F^2 eBE^5 \sin^4\theta}{48\pi^3} \times \{g_V^2 + g_A^2 + 2g_V g_A \cos\theta, (g_V^2 + g_A^2)\cos\theta + 2g_V g_A\}. \quad (6.2)$$

It can be seen from Eq. (6.2) in particular that even for an isotropic neutrino momentum distribution the average momentum loss will be nonzero (proportional to $g_V g_A$) because of parity nonconservation in weak interaction. As we showed in [6,7] in fields of $\sim 10^3 B_e$ the integral asymmetry of the neutrino emission caused by the component Q_z and determined by the expression $A = \left|\sum \mathbf{p}\right| / \sum |\mathbf{p}|$ could reach the scale of ~1% required to explain the observed intrinsic pulsar velocities [14] as a result of the $\nu \longrightarrow \nu e^- e^+$ process only.

In the presence of a magnetized plasma our calculations yield the following result for the same components of the loss four-vector:

$$Q_{0,z} = \frac{G_F^2 eBT^3 E^2}{4\pi^3}\left\{(g_V + g_A)^2(1-u)^2 \times\left[F_2\left(\frac{\varepsilon\tau(1+u)}{2}\right) - F_2(-\infty)\right] \pm (g_A \longrightarrow -g_A;\ u \longrightarrow -u)\right\}, \quad (6.3)$$

where the function $F_2(z)$ was determined in expression (5.7), and the plus or minus signs correspond to the zeroth and z components. Our result for the loss four-vector obtained for the case of a purely magnetic field (6.2) is reproduced from Eq. (6.3) in the low-density plasma limit (T, T_μ, $\mu_\nu \longrightarrow 0$).

In order to illustrate the relationship between the contributions of the plasma and the magnetic field to the four-vector of the neutrino energy and momentum losses in an active medium we shall consider the simpler situation of a low-density neutrino gas and rewrite Eq. (6.3) for the angle $\theta = \pi/2$ in the following form:

$$Q_{0,z}\left(\theta = \frac{\pi}{2}\right) = \frac{G_F^2 eBE^5}{48\pi^3}(g_V^2 + g_A^2, 2g_V g_A)\mathcal{F}\left(\frac{E}{T}\right),$$
$$\mathcal{F}(x) = 1 + \frac{6}{x}\ln(1 - e^{-x/2}) - \frac{24}{x^2}\mathrm{Li}_2(e^{-x/2}) - \frac{48}{x^3}\mathrm{Li}_3(e^{-x/2}). \quad (6.4)$$

It can be seen from a comparison of (6.4) with Eq. (6.2) for $\theta = \pi/2$ that the function $\mathcal{F}(E/T)$ is the ratio of the components of the loss vector in a magnetized plasma and in a purely magnetic field. Figure 3 gives a graph of the function $\mathcal{F}(E/T)$. It can be seen that at $E = E_0 \simeq 3.4T$ there is a unique "window of transparency" when a neutrino does not exchange energy and momentum with a magnetized plasma. The negative values of the function $\mathcal{F}(E/T)$ at lower energies imply that the neutrino captures energy from the plasma and acquires momentum in the opposite direction to the magnetic field. At energies higher than E_0 the neutrino imparts energy to the plasma and also momentum in the direction of the field. This may have extremely interesting astrophysical consequences.

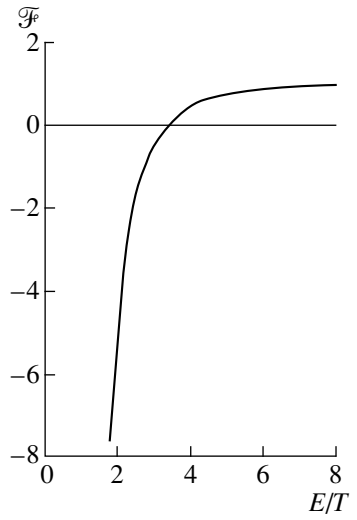

Fig. 3. The function $\mathcal{F}(E/T)$ introduced in (6.4) and determining the dependence of the components of the four-vector of the neutrino energy and momentum losses in a magnetized plasma on the ratio of the neutrino energy to the plasma temperature.

7. INTEGRAL ACTION OF NEUTRINOS ON A MAGNETIZED PLASMA

As an illustration of the application of our results to astrophysical conditions we estimate the volume energy density lost by neutrinos per unit time $\dot{\mathcal{E}}$ and the component $\mathcal{F}_z$ (parallel to the field) of the volume density of the force acting on the plasma from neutrinos

$$(\dot{\mathcal{E}}, \mathcal{F}_z) = \int dn_\nu \frac{1}{E} Q_{0,z}, \quad (7.1)$$

where dn_ν is the initial neutrino density:

$$dn_\nu = \frac{d^3p}{(2\pi)^3} \frac{\Phi(\vartheta, R)}{\exp[(E - \mu_\nu)/T_\nu] + 1}. \quad (7.1)$$

Here the angular distribution of the initial neutrinos is taken into account in the function $\Phi(\vartheta, R)$, $\vartheta = \cos\alpha$,

α is the angle between the neutrino momentum and the radial direction in the star, and R is the distance from the center of the star. At the same time, the similar function $\Phi(\vartheta', R)$ should be introduced in the statistical factor $1 - f'_\nu$ when integrating over the momenta of the final neutrino. In a supernova shell the neutrino angular distribution is close to isotropic [12] so that in the expansion of the function Φ in terms of ϑ, we can confine ourselves to the lowest Legendre polynomials and this function can be uniquely expressed in terms of the average values $\langle\vartheta\rangle$ and $\langle\vartheta^2\rangle$ (which depend on R) as follows:

$$\Phi(\vartheta, R) \simeq 1 + 3\langle P_1(\vartheta)\rangle P_1(\vartheta) + 5\langle P_2(\vartheta)\rangle P_2(\vartheta). \quad (7.2)$$

Neutrinos leaving the central region of a star at high temperature enter the peripheral region where a strong magnetic field is generated and the temperature of the electron–positron gas is lower. In this case the spectral temperatures for different types of neutrino differ [12, 15]:

$$T_{\nu_e} \simeq 4 \text{ MeV}, \quad T_{\bar\nu_e} \simeq 5 \text{ MeV},$$
$$T_{\nu_{\mu,\tau}} \simeq T_{\bar\nu_{\mu,\tau}} \simeq 8 \text{ MeV}. \quad (7.3)$$

The action of a neutrino on a plasma leads to the establishment of thermal equilibrium $\dot{\mathscr{E}}_{\text{tot}} = 0$. When analyzing this equilibrium we need to take into account the contributions to $\dot{\mathscr{E}}_{\text{tot}}$ made by all processes of neutrino interaction with the medium. As we have noted, the probability of the β processes $\nu_e + n \longleftrightarrow e^- + p$ is substantially higher than that for neutrino–electron processes so that these dominate in the energy balance. The energy transferred per unit time per unit plasma volume as a result of these processes involving only electron neutrinos may be expressed in the form

$$\dot{\mathscr{E}}(\beta) \simeq \mathscr{B}(T_{\nu_e} - T)/T.$$

From this it follows that as a result of neutrino heating the plasma temperature should be very close to the spectral temperature of the electron neutrinos ($T \simeq T_{\nu_e}$). However, the contribution to $\dot{\mathscr{E}}$ made by other types of neutrino whose spectral temperatures exceed T_{ν_e}, has the result that the plasma temperature is slightly higher ($T \gtrsim T_{\nu_e}$). It is therefore meaningful to make separate estimates of the contributions to $(\dot{\mathscr{E}}, \mathscr{F}_z)$ made by neutrino–electron processes involving ν_e and all other neutrinos and antineutrinos.

We stress that the appearance of the force density $\mathscr{F}_z$ in expression (7.1) is caused by interference between the vector and axial-vector coupling in the effective Lagrangian (3.1) and is a macroscopic manifestation of parity nonconservation in weak interactions. At first glance, the main contribution to $\mathscr{F}_z$ should be made by electron neutrinos since $g_V(\nu_e) \gg g_V(\nu_{\mu,\tau})$. However, as we shall show below, the main contributions are made by μ and τ neutrinos and antineutrinos (as a result of the conservation of CP parity neutrinos and antineutrinos push the plasma in the same direction). This is because in the vicinity of the ν_e neutrinosphere the spectral temperatures of the other types of neutrinos differ substantially from the plasma temperature $T \simeq T_{\nu_e}$.

7.1. Processes Involving Electron Neutrinos

We obtained the following expression for the volume density of the neutrino energy losses and the force density (7.1):

$$(\dot{\mathscr{E}}, \mathscr{F}_z)_{\nu_e} = \frac{G_F^2 eBT^7}{3\pi^5}(g_V^2 + g_A^2, 2g_V g_A)$$
$$\times\left\{(\tau_e - 1)\int_0^\infty \frac{x^3 dx}{e^x - 1}\int_0^\infty \frac{y^3 dy}{(1 + e^{-x-y+\eta_\nu})(1 + e^{y-\eta_\nu})}\right. \quad (7.4)$$
$$\left. + \frac{27}{8}\left(\langle\vartheta^2\rangle - \frac{1}{3}\right)\int_0^\infty \frac{x^3 dx}{e^x - 1}\int_0^\infty \frac{y^3(3y - x)dy}{(x + y)^2(1 + e^{y-\eta_\nu})}\right\},$$

where $\tau_e = T_{\nu_e}/T$. This formula is written assuming a small deviation from thermal equilibrium between the neutrino gas and the electron–positron plasma ($(\tau_e - 1) \ll 1$) and thus comparatively weak asymmetry of the neutrino distribution ($\langle\vartheta^2\rangle - 1/3) \ll 1$).

A numerical estimate gives

$$(\dot{\mathscr{E}}, \mathscr{F}_z)_{\nu_e} \simeq \left(2.0 \times 10^{30} \frac{\text{erg}}{\text{cm}^2\,\text{s}}, 0.57 \times 10^{20} \frac{\text{dyne}}{\text{cm}^3}\right)$$
$$\times\left(\frac{B}{10^{16}\text{ G}}\right)\left(\frac{T}{4\text{ MeV}}\right)^7 \quad (7.5)$$
$$\times \exp(\eta_\nu)\left[(\tau_e - 1) + 0.53\left(\langle\vartheta^2\rangle - \frac{1}{3}\right)\right].$$

7.2. Processes Involving $\bar\nu_e$, $\nu_{\mu,\tau}$, $\bar\nu_{\mu,\tau}$

In this case $T_\nu/T - 1$ cannot be considered to be a small parameter. However, the relative contribution of the asymmetry of the neutrino distribution is small [12] and can be neglected.

For numerical estimates we can conveniently express the values $\dot{\mathscr{E}}$ and $\mathscr{F}_z$ (7.1) in the following form:

$$(\dot{\mathscr{E}}, \mathscr{F})_{\nu_i} \simeq \mathscr{A}(g_V^2 + g_A^2, 2g_V g_A)\varphi(\eta_i)\psi(\tau_i), \quad (7.6)$$

where

$$\mathcal{A} = \frac{12G_F^2 eBT^7}{\pi^5}$$

$$= \left(\frac{B}{10^{16}\ \mathrm{Gs}}\right)\left(\frac{T}{4\ \mathrm{MeV}}\right)^7 \begin{cases} 1.6\times 10^{30}\ \dfrac{\mathrm{erg}}{\mathrm{cm}^3\ \mathrm{s}} \\ 0.55\times 10^{20}\ \dfrac{\mathrm{dyne}}{\mathrm{cm}^3}, \end{cases} \quad (7.7)$$

$$\varphi(\eta_i) = \frac{\eta_i^4}{24} + \frac{\pi^2\eta_i^2}{12} + \frac{7\pi^4}{360} + \mathrm{Li}_4[-\exp(-\eta_i)], \quad (7.8)$$

$$\varphi(0) = \frac{7\pi^4}{720} \simeq 0.947,$$

$$\psi(\tau_i) = \frac{\tau_i^7}{6}\int_0^\infty \frac{y^2 dy}{e^{\tau_i y}-1}\{\exp[(\tau_i-1)y]-1\}, \quad (7.9)$$

$$\psi(\tau_i)\big|_{\tau_i\to 1} \simeq \tau_i - 1.$$

Formulas (7.5)–(7.10) demonstrate in particular that the action of each individual neutrino fraction on an electron–positron plasma goes to zero when thermodynamic equilibrium is established between this fraction and the plasma $\tau_i = 1$, $\langle\vartheta\rangle = 0$, $\langle\vartheta^2\rangle = 1/3$.

We show that the main contribution to the neutrino action on the plasma is made by μ and τ neutrinos and antineutrinos. In fact the function $\psi(\tau_i)$ (7.10) increases rapidly as the difference between the spectral temperature of the neutrinos and the plasma temperature increases. For example, at temperatures (7.4) we have $\psi(1.25) \simeq 0.824$ for electron antineutrinos and $\psi(2) \simeq 38.47$ for μ and τ neutrinos and antineutrinos. This factor leads to compensation for the smallness of the constant $g_V(\nu_{\mu,\tau})$ and makes the contribution $\nu_{\mu,\tau}$, $\bar{\nu}_{\mu,\tau}$ not only comparable with the contribution of the electron neutrinos and antineutrinos but also dominant.

As we have noted, the contribution of neutrino–electron processes to the energy action of a neutrino on the plasma is small compared with the contribution of β processes and leads to a small departure from equilibrium between electron neutrinos and the plasma so that the total contribution of β processes and all νe processes to the value of $\dot{\mathcal{E}}$ is zero.

For the force action of a neutrino on the plasma parallel to the magnetic field described by $\mathcal{F}_z$ in Eqs. (7.5)–(7.10) the total contribution of all types of neutrino is given by

$$\mathcal{F}_z \simeq 3.6\times 10^{20}\ \frac{\mathrm{dyne}}{\mathrm{cm}^2}\left(\frac{B}{10^{16}\ \mathrm{G}}\right)\left(\frac{T}{4\ \mathrm{MeV}}\right)^7. \quad (7.10)$$

Here we assumed for estimates that the chemical potentials of the neutrinos are zero [15]. Note that the value (7.11) was independent of the chemical potential of an electron–positron plasma.

The force density (7.11) should be compared with the recent result for a similar force caused by β-processes [16]. Under the same physical conditions our value of the force as a result of neutrino–electron processes is of the same order of magnitude and, which is particularly important, of the same sign as the result of [16]. Thus, the role of neutrino–electron processes in a high-intensity magnetic field may be significant in addition to the contribution of β processes.

Note that the force density (7.11) is five orders of magnitude lower than the density of the gravitational force and thus negligibly influences the radial dynamics of the supernova shell. However, when a toroidal magnetic field [3] is generated in the shell, the force (7.11) directed along the field can fairly rapidly (within times of the order of a second[4] lead to substantial redistribution of the tangential plasma velocities. Then in two toroids in which the magnetic field has opposite directions, the tangential neutrino acceleration of the plasma will have different signs relative to the rotational motion of the plasma. This effect can then lead to substantial redistribution of the magnetic field lines, concentrating them predominantly in one of the toroids. This leads to considerable asymmetry of the magnetic field energy in the two hemispheres and may be responsible for the asymmetric explosion of the supernova [17] which could explain the phenomenon of high proper pulsar velocities [14]. In our view it is interesting to model the mechanism for toroidal magnetic field generation taking into account the neutrino force action on the plasma both via neutrino–nucleon and neutrino–electron processes.

8. CONCLUSIONS

As we know, in existing systems for numerical modeling of astrophysical cataclysms such as supernova explosions and coalescing of neutron stars, where the physical conditions being studied can be achieved in principle, the neutrino–electron interaction effects studied by us were neglected. However, in detailed analyses of these astrophysical processes it may be important to allow for the influence of an active medium such as a magnetized e^-e^+ plasma on quantum processes involving neutrinos.

In the present study we have investigated the entire range of neutrino–electron processes in a magnetized plasma. In addition to canonical $\nu e^{\mp} \longrightarrow \nu e^{\mp}$ scattering and $\nu\bar{\nu} \longrightarrow e^-e^+$ annihilation reactions we have also considered exotic processes of synchrotron radiation and pair absorption, $e \longleftrightarrow e\nu\bar{\nu}$, and also neutrino radiation and absorption of an electron–positron pair by a neutrino $\nu \longleftrightarrow \nu e^-e^+$. We have shown that among this

[4] We know that the cooling stage of a supernova shell, known as the Kelvin–Helmholz stage, lasts for around 10 s.

entire range processes involving neutrino pair creation and absorption are kinematically suppressed for the case of relatively high neutrino energies $E_\nu \gg m_e$ and hot dense plasmas $T, \mu \gg m_e$. The total probability of all processes including neutrinos in the initial and final states does not have this suppression. In addition, we observed that the total probability of these processes and also the average neutrino energy and momentum losses do not depend on the chemical potential of the e^-e^+ plasma whereas the contributions of the various processes do contain this dependence. This is a new and unexpected result.

We assume that these results will be useful for a detailed analysis of the dynamics of supernova explosion.

ACKNOWLEDGMENTS

We are grateful to G.G. Raffelt, V.A. Rubakov, and V.B. Semikov and also all participants at the international symposium on "Strong Magnetic Fields in Neutrino Astrophysics" (Yaroslavl, 1999) for stimulating discussions.

This work was partly financed by the Russian Foundation for Basic Research (project no. 98-02-16694).

REFERENCES

1. G. G. Raffelt, *Stars as Laboratories for Fundamental Physics* (Univ. of Chicago Press, Chicago, 1996).
2. M. Yu. Khlopov, *Cosmoparticle Physics* (World Scientific, Singapore, 1999).
3. G. S. Bisnovatyĭ-Kogan, Astron. Zh. **47**, 813 (1970) [Sov. Astron. **14**, 652 (1970)]; *Physical Problems of Theory of Stellar Evolution* (Nauka, Moscow, 1989).
4. V. N. Baĭer and V. M. Katkov, Dokl. Akad. Nauk SSSR **171**, 313 (1966) [Sov. Phys. Dokl. **11**, 947 (1967)]; V. I. Ritus, Tr. Fiz. Inst. Akad. Nauk SSSR **111**, 5 (1979).
5. É. A. Choban and A. N. Ivanov, Zh. Éksp. Teor. Fiz. **56**, 194 (1969) [Sov. Phys. JETP **29**, 109 (1969)]; A. V. Borisov, V. Ch. Zhukovskiĭ, and B. A. Lysov, Izv. Vyssh. Uchebn. Zaved., Fiz. **8**, 30 (1983); A. V. Borisov, A. I. Ternov, and V. Ch. Zhukovsky, Phys. Lett. B **318**, 489 (1993).
6. A. V. Kuznetsov and N. V. Mikheev, Phys. Lett. B **394**, 123 (1997).
7. A. V. Kuznetsov and N. V. Mikheev, Yad. Fiz. **60**, 2038 (1997) [Phys. At. Nucl. **60**, 1865 (1997)].
8. V. G. Bezchastnov and P. Haensel, Phys. Rev. D **54**, 3706 (1996).
9. L. D. Landau and E. M. Lifshitz, in *Statistical Physics* (Nauka, Moscow, 1976; Pergamon, Oxford, 1980), Part 1.
10. L. A. Vassilevskaya, A. V. Kuznetsov, and N. V. Mikheev, Yad. Fiz. **62**, 715 (1999) [Phys. At. Nucl. **62**, 666 (1999)].
11. A. I. Akhiezer and V. B. Berestetskii, *Quantum Electrodynamics* (Nauka, Moscow, 1981; Wiley, New York, 1965).
12. S. Yamada, H.-T. Janka, and H. Suzuki, astro-ph/9809009.
13. P. Arras and D. Lai, Phys. Rev. D **60**, 043001 (1999).
14. I. S. Shklovskiĭ, Astron. Zh. **46**, 715 (1969) [Sov. Astron. **13**, 562 (1969)]; A. G. Lyne and D. R. Lorimer, Nature **369**, 127 (1994); C. J. Horowitz and J. Piekarewicz, Nucl. Phys. A **640**, 281 (1998).
15. V. S. Imshennik and D. K. Nadezhin, Usp. Fiz. Nauk **156**, 561 (1988); D. K. Nadyozhin, in *Proceedings of the Baksan International School on Particles and Cosmology* (World Scientific, Singapore, 1992), p. 153.
16. A. A. Gvozdev and I. S. Ognev, Pis'ma Zh. Éksp. Teor. Fiz. **69**, 337 (1999) [JETP Lett. **69**, 365 (1999)].
17. G. S. Bisnovatyĭ-Kogan and S. G. Moiseenko, Astron. Zh. **69**, 563 (1992) [Sov. Astron. **36**, 285 (1992)]; G. S. Bisnovatyi-Kogan, Astron. Astrophys. Trans. **3**, 287 (1993).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 4, 2000, pp. 761–767.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 4, 2000, pp. 877–884.

PLASMA, GASES

Dynamics of the Magnetic Moment in a Polarized Boltzmann Gas Allowing for Dissipation

T. L. Andreeva and P. L. Rubin*

Lebedev Physical Institute, Russian Academy of Sciences, Moscow, 117924 Russia
**e-mail: rubin@sci.lebedev.ru*

Received March 28, 2000

Abstract—An equation is derived for the dynamics of the spin magnetic moment in a polarized Boltzmann gas allowing for spin loss processes. The general form of the T matrix for collisions between two spin 1/2 particles allowing for inelastic processes is used. It is shown that the rate of spin loss depends on the degree of polarization of the gas. As a result, the damping of deviations of the magnetic moment from the average becomes anisotropic where the degree of anisotropy depends on the amplitude of the zero-angle scattering of atoms. *© 2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

At present, among spin-polarized gases, of particular importance are vapors of alkali metals (rubidium, cesium) at temperatures close to room temperature. These gases are good donors for transferring the electron polarization of an alkali metal atom to the nuclear spin of He^3, which can give high concentrations of polarized helium. For its part, polarized He^3 is a unique object for investigating fundamental phenomena in nuclear physics and also for medical applications [1, 2]. From our point of view alkali metal vapor is also of independent interest. This is because in these gases, which are typical Boltzmann gases, it is possible for weakly damped spin waves to propagate in the presence of polarization [3]. It has been assumed so far that the only propagating collective mode in a Boltzmann gas is a sound wave. The parameters characterizing the spin wave spectrum in a Boltzmann gas may be expressed in terms of the exact scattering matrix of the atoms. In alkali metal vapor, unlike He^3, electron spin loss processes accompanying collisions of atoms ($\uparrow\uparrow \longrightarrow \uparrow\downarrow$) become important.

In the present study we derive equations for the dynamics of the spin magnetic moment in a polarized Boltzmann gas allowing for spin loss processes. It was found that the rate of spin loss depends on the degree of polarization of the gas. As a result, the damping of small deviations of the magnetic moment from the average becomes anisotropic where the degree of anisotropy depends on the amplitude of the inelastic zero-angle scattering of the atoms.

2. OPERATOR NOTATION OF THE COLLISION INTEGRAL

The collision integral for the Wigner matrix is conveniently written in the following invariant form [4]:

$$\begin{aligned}\mathrm{St}\, f(p) &= (2\pi)^3\hbar^2\mathrm{Tr}_1\langle p| - iTff_1 + iff_1T^+|p\rangle \\ &+ (2\pi)^4\hbar^2\mathrm{Tr}_1\langle p|Tff_1\Delta(P,E)T^+|p\rangle.\end{aligned} \quad (1)$$

Here $f(p)$ is the Wigner function (the matrix in the particle internal-state space), p is the particle momentum (for conciseness the dependence of f on the x coordinate is not given explicitly), $f_1 = f(p_1)$ (p_1 is the momentum of an impinging particle), Tr implies that the trace is taken over the momentum and quantum numbers of the particle internal state (Tr_1 refers to an impinging particle), T is the scattering t matrix, "+" denotes Hermitian conjugation; Dirac notation is used for the matrix elements, and finally

$$\Delta(P,E) = \delta(P-P')\delta(E-E'),$$

where P and P' are the relative momenta of the colliding particles:

$$P = p - p_1, \quad P' = p' - p_1',$$

and $E = P^2/4m$, $E' = P'^2/4m$ are the corresponding energies ($m/2$ is the reduced mass of the colliding particles). The product ff_1 is the direct product of the operators forming the operator in the variable space of two particles so that f and f_1 can be considered to commute. Note that the first term on the right-hand side of Eq. (1) only contains integration over the momentum of the impinging particle p_1 whereas the second term contains additional

1063-7761/00/9104-0761$20.00

integration over the scattered particle momentum p' since the T matrix is a two-particle operator:

$$T_{\alpha\alpha_1\alpha'\alpha'_1}(p - p_1, p' - p'_1) = T(\alpha p, \alpha_1 p_1 | \alpha' p', \alpha'_1 p'_1).$$

Note that Eq. (1) belongs to a system which is degenerate in terms of internal quantum numbers, i.e., all the energy levels E_α are the same.

Let us assume that $\hat{A}$ is an operator which is multiplicative in relation to the momentum variable like the Wigner matrix, i.e.,

$$\hat{A}\psi = A_{\alpha\beta}(p)\psi_\beta(p).$$

(ψ is the atomic wave function). We are dealing with the average value of $\hat{A}$, i.e., the quantity

$$A = \langle\langle \hat{A} \rangle\rangle = \mathrm{Tr}(\hat{A} f).$$

Using the kinetic equation for the Wigner matrix

$$\frac{df}{dt} = \mathrm{St}\, f(p), \tag{2}$$

we obtain an equation for the average value of A:

$$\frac{d\langle\langle A\rangle\rangle}{dt} = \mathrm{Tr}[A\,\mathrm{St}(f)] = -i(2\pi)^3\hbar^2 \times [\hat{\mathrm{T}}\mathrm{r}(ATff_1) - \hat{\mathrm{T}}\mathrm{r}(Aff_1T^+)] + (2\pi)^4\hbar^2\hat{\mathrm{T}}\mathrm{r}[ATff_1\Delta(P, E)T^+]. \tag{3}$$

Here $\hat{\mathrm{T}}\mathrm{r}$ is the complete trace in the combination of both spaces, including integration over momenta. When the particle numbering is changed, the T matrix remains the same:

$$T(\xi, \xi_1|\xi', \xi'_1) = T(\xi_1, \xi|\xi'_1, \xi')$$

($\xi = \{\alpha, p\}$). Thus, in (1) we can replace A with A_1 and therefore with

$$\tilde{A} = (A + A_1)/2.$$

Then, instead of (1) we can write the more symmetric formula:

$$\frac{d\langle\langle A\rangle\rangle}{dt} = -i(2\pi)^3\hbar^2[\hat{\mathrm{T}}\mathrm{r}(\tilde{A}Tff_1) - \hat{\mathrm{T}}\mathrm{r}(\tilde{A}ff_1T^+)] + (2\pi)^4\hbar^2\hat{\mathrm{T}}\mathrm{r}[\tilde{A}Tff_1\Delta(P, E)T^+]. \tag{4}$$

The formula is further transformed by replacing

$$\tilde{A}T = [\tilde{A}, T] + T\tilde{A}$$

(the symbol […, …] denotes a commutator) and, using a generalized optical theorem at the energy surface [5]:

$$T_{gg'} - T^*_{g'g} = -2\pi i\int \delta(E_g - E_h)T_{gh}T^*_{g'h}.$$

Here g, g', h is the complete set of quantum numbers of the particle state, including the momentum, the asterisk is the complex conjugation sign, E_g is the total particle energy in state g and it is assumed that $E_g = E'_g$. Finally we obtain

$$\frac{d\langle\langle A\rangle\rangle}{dt} = -(2\pi)^3\hbar^2\hat{\mathrm{T}}\mathrm{r}(i[\tilde{A}, T]ff_1) + (2\pi)^4\hbar^2\hat{\mathrm{T}}\mathrm{r}\{[\tilde{A}, T]ff_1\Delta(P, E)T^+\}. \tag{5}$$

If $[\tilde{A}, T] = 0$, both sides of the equality (5) are zero which implies that the average value of the quantity conserved during collisions is constant.

At first glance, the right-hand side of (3) is not necessarily real. However, by means of a completely similar transformation this can be reduced to the form

$$\frac{d\langle\langle A\rangle\rangle}{dt} = -(2\pi)^3\hbar^2\hat{\mathrm{T}}\mathrm{r}(i[\tilde{A}, T^+]ff_1) + (2\pi)^4\hbar^2\hat{\mathrm{T}}\mathrm{r}\{Tff_1\Delta(P, E)[T^+, \tilde{A}]\}. \tag{6}$$

The half-sum of the right-hand sides of the last two equalities is the trace of the Hermitian operator (the operator $\tilde{A}$ is Hermitian like the operator of the observable quantity) and is therefore real. A different procedure can also be adopted, simply taking the real part of the right-hand side of (3) and writing

$$\frac{d\langle\langle A\rangle\rangle}{dt} = (2\pi)^3\hbar^2\mathrm{Re}\{\hat{\mathrm{T}}\mathrm{r}(-i[\tilde{A}, T]ff_1)\} - (2\pi)^4\hbar^2\mathrm{Im}\left[\hat{\mathrm{T}}\mathrm{r}\left\{\frac{[\tilde{A}, T]}{i}ff_1\Delta(P, E)T^+\right\}\right]. \tag{7}$$

Since the operators f and A are diagonal in terms of momentum, the first term on the right-hand side contains the T matrix of zero-angle scattering.

3. RELAXATION OF THE MAGNETIC MOMENT IN A POLARIZED GAS

The relaxation of the magnetic moment in a polarized gas is described by Eq. (7) in which the spin magnetic moment of the atom (in units of $\hbar/2$) $\hat{m} = \sigma$ appears as the operator A, where σ is the vector of the Pauli matrices. The magnetic moment per unit volume of gas is then

$$M = \langle\langle f\hat{m}\rangle\rangle \tag{8}$$

(having in mind integration over momenta and summation over the quantum numbers of the internal atomic state). In the present study, only the spin moment, i.e., atoms in the S state is being considered.

The relaxation of the magnetic moment as a result of collisions of spin 1/2 atoms is determined by relativistic interactions which do not conserve spin. The scattering matrix T then has the following form [6]:

$$T = T_0 + Q_{ij}\sigma^1_i\sigma^2_j. \tag{9}$$

Here T_0 is the elastic scattering matrix which conserves spin ($[T_0, \hat{m} + \hat{m}_1] = 0$) and

$$Q_{ii} = 0, \quad Q_{ij} = Q_{ji}.$$

The main contribution to T_0 is made by ordinary elastic collisions allowing for exchange (see [3]). Strictly speaking, spin-conserving relativistic corrections also make some contribution. The only spin nonconserving term is the last term on the right-hand side of (9).

In this particular case the structure of the Wigner matrix has the following form:

$$f(p) = \frac{n_0(p)}{2}[R(p) + S_i(p)\sigma_i],$$

where $n_0(p)$ is the Maxwell momentum distribution function with the normalization

$$\int n_0(p)dp = N$$

(n is the local particle density). In the equilibrium state in a polarized gas neglecting spin damping we have

$$f = f_0 = \frac{n_0(p)}{2}[1 + M_i\sigma_i]. \tag{10}$$

As usual, averaging using $f_0(p)$ should give the same results as the exact function $f(p)$ so that

$$R(p) = 1, \quad S(p) = M.$$

In order to obtain an equation for the dynamics of the magnetic moment, we need to find the commutator

$$[(\sigma_i^1 + \sigma_i^2)/2, T] = 2iK_{ipq}\sigma_p^1\sigma_q^2, \tag{11}$$

where

$$K_{ijk} = \varepsilon_{is}({}_jQ_k)_s. \tag{12}$$

Symmetrization is implied with respect to the indices enclosed in parentheses. In Eq. (11) the particle numbering is changed for clarity: instead of the unnumbered (dominant) and first (perturbing) particle, we now have the first and second. Formulas (11) and (12) yield an expression for the first term in Eq. (7):

$$-i\hat{\mathrm{T}}\mathrm{r}([\tilde{\sigma}_i, T]ff_1) = 2\left\langle \varepsilon_{is}({}_jQ_k^0)_s S_j S_k^1 \right\rangle. \tag{13}$$

Here the particle numbering is that normally used in kinetics, the angular brackets denote integration over momenta or, more accurately, the trace over the momentum variables, and Q^0 denotes the amplitude of zero-angle inelastic scattering.

The second term is calculated similarly and has the form

$$\begin{aligned}&\mathrm{Im}\left[\hat{\mathrm{T}}\mathrm{r}\left\{\frac{[\tilde{A}, T]}{i}ff_1\Delta(P, E)T^+\right\}\right]\\ &= -4\,\mathrm{Re}[\langle Q_{qr}(RS_i^1 + R^1S_i)Q_{qr}^*\rangle\\ &\quad - \langle Q_{qr}(RS_r^1 + R^1S_r)Q_{qi}^*\rangle].\end{aligned} \tag{14}$$

Thus, the complete equation for relaxation of the magnetic moment M has the form

$$\begin{aligned}\frac{dM_i}{dt} &= 16\pi^3\hbar^2\mathrm{Re}\{\left\langle \varepsilon_{is}({}_jQ_k^0)_s S_j S_k^1\right\rangle\\ &\quad + 4\pi[\langle Q_{qr}(RS_i^1 + R^1S_i)Q_{qr}^*\rangle\\ &\quad - \langle Q_{qr}(RS_r^1 + R^1S_r)Q_{qi}^*\rangle]\}.\end{aligned} \tag{15}$$

In order to calculate the functions $R(p, t)$ and $S(P, t)$ which determine the dynamics of the magnetic moment, we need to use the kinetic equation (2). The function $f(p, t)$ can then be conveniently expressed in the form

$$f = f_0 + \phi, \tag{16}$$

where ϕ is a small correction to the function f_0 having the same structure:

$$\phi = \frac{n_0(p)}{2}[\rho(p) + s_i(p)\sigma_i].$$

The equation then has the form

$$\frac{d}{dt}(f_0 + \phi) = \mathrm{Coll}_0(f_0 + \phi) + \mathrm{st}(f_0).$$

Here Coll_0 is the collision integral describing purely elastic scattering where the spin is conserved (scattering matrix T_0) and coll is the relaxation component of the collision integral which only contains a small inelastic component of the T-matrix [see (9)]. As a result of this factor only the main component of the Wigner function f_0 remains within the coll.

The operator $\mathrm{Coll}_0(f_0 + \phi)$ can be expressed as follows:

$$\mathrm{Coll}_0(f_0 + \phi) = J(\phi),$$

where $J(\phi)$ is the linearized collision operator since $\mathrm{Coll}_0(f_0) \equiv 0$. Taking this factor into account, the kinetic equation has the form

$$\frac{d}{dt}(f_0 + \phi) = J(\phi) + \mathrm{st}(f_0). \tag{17}$$

We note that the elastic collision operator J is Hermitian with a suitably determined scalar product (see, e.g., [3]). It has an eightfold degenerate zero eigenvalue which corresponds to the laws of conservation of particle number, energy, momentum, and spin (three components each). These eight functions form a linear

space to which the function f_0 is assigned. The division (16) is assumed to be such that the function ϕ is orthogonal to the kernel (zero proper subspace) of the operator J. Finally the equation for ϕ has the form:

$$\frac{d\phi}{dt} = J(\phi) + \hat{P}\mathrm{coll}(f_0).$$

Here $\hat{P}$ is the projection operator on the subspace orthogonal to the kernel J.

Since the relaxation of the magnetic moment is a relatively slow process, the time derivative can be neglected in the last equation:

$$\phi = -J^{-1}\hat{P}\mathrm{coll}(f_0).$$

The operator J^{-1} is the inversion of J on the subspace orthogonal to the kernel. For the vector component ϕ we then obtain

$$s(p) = \mathrm{tr}[-J^{-1}\hat{P}\mathrm{coll}(f_0)\sigma],$$

where tr is the trace over the spin variables (but not over the momenta).

The operator of elastic collisions in a polarized gas J analyzed in detail in [3] has the form

$$\hat{J}\begin{pmatrix} \rho \\ s \end{pmatrix} = \begin{pmatrix} J_1 & MJ_3 \\ MJ_4 & J_2 + M\times J_5 \end{pmatrix}\begin{pmatrix} \rho \\ s \end{pmatrix}.$$

A similar expression for the integral operators J_n (n = 1, …, 5) in terms of the elastic scattering T-matrix is given in the Appendix.

We recall that in an unpolarized gas the nonzero eigenvalues of the operator $\hat{J}$ are of the order of the gas kinetic collision frequencies ν_s although they differ for the scalar J_1 and vector J_2 components of the Wigner function. For a polarized gas the spectrum of the operator $\hat{J}$ changes substantially. The most important change is the appearance of eigenvalues of the order ν_{ex} (the operator J_5), known as the "exchange" frequency [3] which is far higher than the gaskinetic frequency ν_s at least at room temperature.

The vector component of the operator coll(f_0) has the form

$$\mathrm{tr}[\mathrm{coll}(f_0)\sigma_i] = A_{ik}M_k + B_{ijk}M_jM_k.$$

Here we have

$$A_{ik}(p) = 128\pi^4\hbar^2\mathrm{Re}[Q_{qr}Q^*_{qr}\delta_{ik} - Q_{qk}Q^*_{qi}], \quad (18)$$

$$B_{ijk} = 16\pi^3\hbar^2\mathrm{Re}[\varepsilon_{is}({}_jQ^0_k)_s]. \quad (19)$$

The tensors $A_{ik} = A_{ik}(p)$ and $B_{ijk} = B_{ijk}(p)$ are expressed in terms of the T matrix of the dissipative collision process which does not conserve spin (but conserves the particle number) $Q_{pq}\sigma_p\sigma_q$. The tensor A_{ik} is proportional to the cross section of the dissipative process while the tensor B_{ijk} is proportional to the real part of the amplitude of zero-angle scattering. For the small correction s we can now give the estimate:

$$s_\alpha \simeq \sum_\beta \frac{\bar{A}_{\alpha\beta}}{\nu_\alpha}M_\beta + \sum_{\beta\gamma}\frac{\bar{B}_{\alpha\beta\gamma}}{\nu_\alpha}M_\beta M_\gamma. \quad (20)$$

The bar over A and B indicates that the corresponding values are exposed to the action of the operator $\hat{P}$. Using Greek subscripts in this last formula indicates a transition to the cyclic components of vectors and tensors since, in a polarized gas, the elastic collision operator J is diagonalized in this basis [3].

The value of ν_α for the longitudinal component has the order of the gaskinetic collision frequency ν_s whereas for the transverse component it is considerably higher than the "exchange" frequency $\nu_{ex} = 16\pi^3\hbar^2 n\mathrm{Re}T_0$[1] ($T_0$ is the amplitude of zero-angle elastic scattering). As a result of this factor we have $\nu_{ex} \gg \nu_s$ and thus only corrections which include ν_s are taken into account [3]. Substituting Eq. (20) into Eq. (15), we obtain the damping of the magnetic moment:

$$\begin{aligned} -\dot{M}_\alpha &= a_{\tau\beta}A_{\alpha\tau}M_\beta + A_{\alpha\tau}b_{\tau\beta\gamma}M_\beta M_\gamma \\ &+ a_{\sigma\kappa}a_{\tau\beta}B_{\alpha\tau\sigma}M_\beta M_\kappa + a_{\sigma\kappa}b_{\tau\beta\gamma}B_{\alpha\tau\sigma}M_\beta M_\gamma M_\kappa \\ &+ a_{\tau\beta}B_{\alpha\tau\sigma}M_\beta M_\sigma + b_{\tau\beta\gamma}B_{\alpha\tau\sigma}M_\beta M_\gamma M_\sigma \\ &+ A_{\alpha\tau}M_\tau + a_{\sigma\kappa}B_{\alpha\tau\sigma}M_\kappa M_\tau + B_{\alpha\tau\sigma}M_\sigma M_\tau \\ &+ a_{\tau\beta}b_{\sigma\varphi\chi}B_{\alpha\tau\sigma}M_\beta M_\varphi M_\chi + b_{\sigma\varphi\chi}B_{\alpha\tau\sigma}M_\tau M_\varphi M_\chi. \end{aligned} \quad (21)$$

Here we have

$$a_{\alpha\beta} = \frac{\bar{A}_{\alpha\beta}}{\nu_\alpha}, \quad b_{\alpha\beta\gamma} = \frac{\bar{B}_{\alpha\beta\gamma}}{\nu_\alpha}.$$

In this case we find

$$M = \int n_0(p)S(p)dp = \int n_0(p)[S_0 + s(p)]dp.$$

It follows from symmetry concepts that Eq. (21) can only contain odd powers of the vector components of the magnetic moment M_α. Consequently, the rate of damping of the magnetic moment may be expressed by the formula

$$\dot{M}_\alpha = -\gamma(M)M_\alpha. \quad (22)$$

Here $\gamma(M)$ may be expressed in the form

$$\begin{aligned} \gamma(M) &= A(1+a) + (bB + abB)M^2 \\ &= \gamma_0 + \gamma_1(M), \end{aligned} \quad (23)$$

where

$$\gamma^0\ (\mathrm{s}^{-1}) = A(1+a), \quad \gamma_1(M) = (bB + abB)M^2.$$

[1] Here unlike previous studies [3] the factor $|M|$ is not introduced in the definition of ν_{ex}.

The value of $\gamma_0 \sim A\,(s^{-1})$ determines the damping of the magnetic moment in an unpolarized gas and is proportional to the cross section of the corresponding process [see (14)]. The small parameter a allows for the influence of elastic collisions on the spin relaxation process and is of the order of magnitude

$$a \sim \frac{A}{\nu_s} \ll 1.$$

This parameter is the ratio of the cross section for inelastic scattering to that for elastic scattering. Then, we have

$$\gamma_1 \sim bBM^2 \sim \gamma_n^0 \frac{\gamma_n^0}{\nu_s},$$

where $\gamma_n^0\ (s^{-1}) \sim C|M|\mathrm{Re}\,Q^0$ (Q^0 is the inelastic component of the T matrix for zero-angle scattering). Here C is a constant having dimensions of frequency determined by the first term on the right-hand side of Eq. (15). Since $a \ll 1$, the second term in Eq. (23) can simply be replaced by bBM^2 and then

$$\gamma_1 \sim bBM^2 \sim \gamma_n^0 \frac{\gamma_n^0}{\nu_s}. \tag{24}$$

For a small deviation of the magnetic moment m from the steady-state average M^0 we have the following equation:

$$\frac{dm_i}{dt} = -(\gamma_0 + \gamma_1)m_i - 2\gamma_1 \frac{M_i^0 M_k^0}{M_0^2} m_k. \tag{25}$$

It can be seen from Eq. (25) that the rate of relaxation of the longitudinal components of the magnetic moment differs from the transverse components $(\gamma_0 + 3\gamma_1)/(\gamma_0 + \gamma_1)$ times as a result of the second term on the right-hand side of the formula.

4. EQUATION FOR THE DYNAMICS OF THE MAGNETIC MOMENT

The dynamics of the magnetic moment in a polarized Boltzmann gas allowing only for elastic collisions was analyzed in detail in our previous study [3]. Using the microscopic equations obtained there taking into account the results of the present study which allow for inelastic processes, we can write the complete macroscopic equation for the dynamics of the magnetic moment:

$$\frac{\partial M_i}{\partial t} + \frac{\partial}{\partial x_j}\left[D_0 \frac{\partial M_i}{\partial x_j}\right] + \frac{1}{2}\frac{\partial}{\partial x_j}\left[D_1 M_i \frac{\partial M^2}{\partial x_j}\right] + \frac{\partial}{\partial x_j}\left[D_2 \varepsilon_{ikl} M_l \frac{\partial M_k}{\partial x_j}\right] = -\gamma(M) M_i. \tag{26}$$

Microscopic expressions for the coefficients D_0, D_1, and D_2 may be obtained by comparing solutions of the linearized Eq. (26) with similar solutions obtained using the kinetic equation [3]. We then obtain

$$D_0 = \mathrm{Re}\langle f_+ | \hat{v} J_+^{-1} \hat{v} | f_+ \rangle, \tag{27}$$

$$D_2 = \frac{\mathrm{Im}\langle f_+ | \hat{v} J_+^{-1} \hat{v} | f_+ \rangle}{M}, \tag{28}$$

$$D_1 = \frac{\langle f_\parallel | \hat{v} J_\parallel^{-1} \hat{v} | f_\parallel \rangle - D_0}{M^2}. \tag{29}$$

Here we have

$$\hat{J}_\parallel \begin{pmatrix} \rho \\ s \end{pmatrix} = \begin{pmatrix} J_1 & MJ_3 \\ MJ_4 & J_2 \end{pmatrix} \begin{pmatrix} \rho \\ s \end{pmatrix},$$

$$J_\pm = U_\pm (J_2 + M \times J_5) U_\pm^{-1} = J_2 \mp iMJ_5,$$

where U is the operator of the transformation from Cartesian components of the vector s to cyclic components. The operators $J_\pm$ act in the vector subspace of the Wigner functions where they are diagonal with respect to the cyclic components s, and it is implied that J_2 and J_5 act on the cyclic components of the vectors as on the initial Cartesian components (see Appendix); $\hat{v}$ indicates multiplication by the particle velocity (considered as an operator); $\langle f_+|$ and $\langle f_\parallel|$ are the eigenfunctions of the transverse and longitudinal components of the magnetic moment (for further details see [3]). Formulas (27)–(29) in principle solve the problem of the microscopic calculation of diffusion coefficients in a polarized Boltzmann gas.

We shall now estimate the coefficients D_0 and D_2 using their microscopic representations (27), (28) and earlier estimates [3]:

$$D_0 \simeq \frac{D}{1 + M^2(\nu_{ex}/\nu_s)^2}, \tag{30}$$

$$D_2 \simeq \frac{D(\nu_{ex}/\nu_s)}{1 + M^2(\nu_{ex}/\nu_s)^2}. \tag{31}$$

Here $D = (kv)^2/\nu_s$ is the coefficient of spin diffusion in an unpolarized gas.

From the general Eq. (26) we can then easily obtain a linearized equation for the dynamics of small perturbations of the magnetic moment $m_\alpha = M_\alpha - \langle M_\alpha \rangle$ (these components are more conveniently analyzed using cyclic coordinates):

$$\frac{\partial m_\alpha}{\partial t} + D_{\alpha\beta}\Delta m_\beta = -\gamma_{\alpha\beta} m_\beta. \tag{32}$$

In this basis the tensors D and γ are diagonal. We then have

$$D_{00} = D_0 + D_1 M^2 \simeq D(1 + cM^2),$$

$$D_\pm = D_0 \pm iMD_2 \simeq D\frac{1 \mp iM(\nu_{ex}/\nu_s)}{1 + M^2(\nu_{ex}/\nu_s)^2}.$$

Here D is the diffusion coefficient in an unpolarized gas, which is of the order of magnitude of $\langle v^2\rangle/3\nu_s$, and c is a dimensionless constant. As we showed earlier,

$$\nu_{ex}/\nu_s \sim |U|a_0/\hbar\langle v\rangle \gg 1,$$

since the factor $|U|a_0/\hbar\langle v\rangle$ is the so-called "Born parameter" [3] which is usually large in classical gases. An estimate of the values of $D_\pm$ and D_{00} for moderately small $|M|$ gives [3]

$$\frac{\mathrm{Re}\, D_\pm}{D_{00}} \simeq \left(\frac{\nu_s}{\nu_{ex}}\right)^2 \frac{1}{M^2} \ll 1,$$
$$\mathrm{Im}\, D_\pm \simeq \pm\frac{(kv)^2}{3\nu_{ex}|M|}. \qquad (33)$$

These results yield the well-known dispersion relationship for spin waves. The frequency and damping of the spin waves are determined by the imaginary and real parts of D, respectively.

For the tensor γ we have

$$\gamma_\pm = \gamma_0 + \gamma_1, \quad \gamma_{00} = \gamma_0 + 3\gamma_1. \qquad (34)$$

Note that the frequencies γ_0 and γ_1 are much lower than the gaskinetic collision frequency ν_s but their interrelationship may be arbitrary.

5. CONCLUSIONS

The operator notation of the collision integral using the generalized optical theorem [Eq. (7)] has been used to obtain a universal equation for the dissipation of the magnetic moment of degenerate systems: a general form of the T matrix is used for the collision of two spin 1/2 particles allowing for elastic processes [6]. The explicit form of the equation only contains the inelastic component of the T matrix since its elastic component T_0 commutes with the spin operator. Nevertheless, the role of elastic processes in spin relaxation is not lost and is observed as a result of their influence on the nonequilibrium particle velocity distribution function.

The main result of the study is the equation for the dynamics of the macroscopic magnetic moment in a polarized gas allowing for spin relaxation (7). The left-hand side of the equation has the same form as the Leggett equation [7] while the right-hand side describes the damping of the magnetic moment. For the coefficients D_0, D_1, and D_2 on the left-hand side of the equation we obtained microscopic expressions containing elements of the T matrix for elastic scattering (27)–(29) whereas in the Leggett study these coefficients are phenomenological. The microscopic expressions for the diffusion coefficients in gases obtained in [8] differ from (27)–(29) because of differences in the form of the collision integral used (s-wave scattering is used in [8]). Nevertheless, if a suitable substitution of the parameters is made: $\mu \longrightarrow \nu_{ex}/\nu_s$ (Leggett) and $\Omega_{int}\tau_\perp \longrightarrow \nu_{ex}/\nu_s$ [8], the spin wave spectrum is the same in all three cases.

A microscopic approach to calculating the diffusion coefficients and spin dynamics in quantum gases for an arbitrary T-matrix was also used in [9]. However, the estimates of the diffusion coefficients in the Boltzmann limit made in [9] do not agree with the results of the present study [see (27)–(29)]. In particular, our condition $\nu_{ex}/\nu_s \gg 1$ which determines the possibility of spin wave propagation in a Boltzmann gas is reduced in [9] to the gas "quantization" condition (the de Broglie wavelength is much greater than the scattering wavelength) which makes it impossible for spin waves to propagate in a Boltzmann gas. A more detailed comparison between the results of the present study and [9] requires separate analysis.

Note that the rate of damping of the magnetic moment γ in a polarized gas depends on its magnitude. This is a consequence of the deviation of the Wigner matrix from the equilibrium (diagonal) matrix as a result of inelastic collisions which impair the conservation of spin. The perturbed Wigner matrix is generally off-diagonal with respect to the spin variables so that the rate of relaxation $\gamma(M)$ depends not only on the cross section of the dissipative process but also on the real part of the amplitude of zero-angle scattering, a factor which is impossible in the pure balance form of the Boltzmann equation. The damping of small perturbations of the magnetic moment in a polarized gas is anisotropic as was first noted by Snider [10]. It is therefore possible to explain the observed discrepancy (approximately an order of magnitude) between the observed rate of relaxation of the magnetic moment in polarized Rb vapor and the calculations of the corresponding relaxation cross section [2]: $\sigma \sim 10^{-17}$ cm^2 according to [1, 11] (experimental) and $\sigma \leq 10^{-18}$ cm^2 according to [12] (calculated).

It has been shown that it is generally inadequate to describe the relaxation of the magnetic moment merely using the cross section, since the rate of the longitudinal and transverse relaxation of the moment depends not only on the cross section but also on the real part of the amplitude of inelastic zero-angle scattering (34) where the relationship between them (γ_0 and γ_1) is unknown a priori and may be determined experimentally using the ratio of the rates of damping of the polarization in the longitudinal and transverse directions relative to the polarization vector of the gas. Note that the anisotropy of the diffusion coefficients in a polarized Boltzmann gas may be very high [see (33)] while the anisotropy of the polarization damping factor varies in the range 1–3.

ACKNOWLEDGMENTS

This work was supported in part by the Russian Foundation for Basic Research (project no. 99-02-16304).

APPENDIX

The operators J_n ($n = 1, \ldots, 5$) have the form [3]

$$J_1(\rho) = 32\pi^4\hbar^2\int dp'dp_1dp_1'W(p, p_1|p', p_1') \times A_0[\rho(p') + \rho(p_1') - \rho(p) - \rho(p_1)],$$

$$J_2(s) = 32\pi^4\hbar^2\int dp'dp_1dp_1'W(p, p_1|p', p_1') \times [A_1s(p_1') + A_2s(p') - A_3s(p_1) - A_0s(p)],$$

$$J_3(s) = 32\pi^4\hbar^2\int dp'dp_1dp_1'W(p, p_1|p', p_1') \times A_3M[s(p') + s(p_1') - s(p) - s(p_1)],$$

$$J_4(\rho) = 32\pi^4\hbar^2\int dp'dp_1dp_1'W(p, p_1|p', p_1') \times M[A_1\rho(p') + A_2\rho(p_1') - A_0\rho(p_1) - A_3\rho(p)],$$

$$J_5(s) = 32\pi^4\hbar^2\int dp'dp_1dp_1'W(p, p_1|p', p_1') \times A_4[s(p_1') + s(p')] + 16\pi^3\hbar^2\int dp_1 f^{(0)}(p) f^{(0)}(p_1) \times \mathrm{Re}\left[t_1\left(\frac{p-p_1}{2}, \frac{p-p_1}{2}\right) - t_2\left(\frac{p-p_1}{2}, \frac{p-p_1}{2}\right)\right] \times [s(p) - s(p_1)].$$

Here we have

$$W(p, p_1|p', p_1') = f^{(0)}(p_1)\delta(p + p_1 - p' - p_1') \times \delta[(p-p_1)^2/4m - (p'-p_1')^2/4m],$$

and the coefficients A_0–A_4, t_1, and t_2 are expressed directly in terms of the elements of the elastic-scattering T matrix [3].

REFERENCES

1. M. E. Wagshul and T. E. Chupp, Phys. Rev. A **49**, 3854 (1994).
2. N. N. Kolachevsky, A. A. Papchenko, Yu. V. Procofichev, *et al.*, Preprint No. 39, FIAN (Lebedev Institute of Physics, Russian Academy of Sciences, Moscow, 1999); Yu. V. Bogdanov, K. V. Volodchenko, S. I. Kanorskii, I. I. Sobelman, *et al.*, Preprint No. 129, FIAN (Lebedev Institute of Physics, Academy of Sciences of USSR, Moscow, 1991).
3. T. L. Andreeva and P. L. Rubin, Pis'ma Zh. Éksp. Teor. Fiz. **67**, 777 (1998) [JETP Lett. **67**, 820 (1998)]; T. L. Andreeva and P. L. Rubin, Zh. Éksp. Teor. Fiz. **115**, 865 (1999) [JETP **88**, 476 (1999)].
4. R. F. Snider, J. Chem. Phys. **32**, 1051 (1960).
5. M. L. Goldberger and K. M. Watson, *Collision Theory* (Wiley, New York, 1964; Mir, Moscow, 1967).
6. L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-Relativistic Theory* (Fizmatgiz, Moscow, 1963; Pergamon, Oxford, 1977).
7. A. J. Leggett, J. Phys. C **3**, 448 (1970).
8. A. E. Meyerovich, J. H. Naish, J. R. Owers-Bradly, and A. Stepaniants, Fiz. Nizk. Temp. **23**, 553 (1997) [Low Temp. Phys. **23**, 411 (1997)].
9. J. W. Jeon and W. J. Mullin, J. Low Temp. Phys. **88**, 483 (1992).
10. F. M. Chen and R. F. Snider, J. Chem. Phys. **46**, 3937 (1967).
11. R. J. Knize, Phys. Rev. A **40**, 6219 (1989).
12. E. I. Dashevskaya, Opt. Spektrosk. **51**, 71 (1981) [Opt. Spectrosc. **51**, 37 (1981)].

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 4, 2000, pp. 768–774.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 4, 2000, pp. 885–892.

FLUIDS

Manifestation of Dissipation and Compressibility in the Experiments on Quantum Cavitation¶

S. N. Burmistrov and L. B. Dubovskiĭ
Russian Research Centre Kurchatov Institute, Moscow, 123182 Russia
e-mail: dubovski@kurm.polyn.kiae.su
Received March 23, 2000

Abstract—Cavitation in the liquid helium isotopes of ^{3}He and ^{4}He is considered. It is shown that the dissipative processes play an important role in the growth of the stable phase nucleus in the normal liquid ^{3}He. This leads to the lack of the quantum behavior of cavitation in this system down to 2 mK, which is in contrast to the thermal–quantum crossover in the cavitation of the superfluid ^{4}He at 600 mK. Below 180 mK, the dissipative ^{3}He kinetics is of the Knudsen type. The high value (600 mK) for the transition into the quantum kinetic behavior in ^{4}He is related to the compressibility of a liquid, which leads to a noticeable emission of sound with cavitation. The recent experiments on quantum cavitation in the liquid helium isotopes ^{3}He and ^{4}He are discussed. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

At low temperatures, the first-order phase transition associated with the formation of droplets of a stable phase occurs via quantum tunneling through the potential barrier originating from the positive interfacial tension between the phases. Such a macroscopic underbarrier motion is associated with the flow of the metastable fluid towards the droplet due to the mass difference between the stable and metastable liquids [1]. The growth of the droplet in the metastable phase is also accompanied by dissipative effects due to the lack of equilibrium in the medium during droplet evolution [2]. Direct experimental observations of the dissipation effects accompanying the kinetics of the first-order phase transition are absent so far.

Recently, there was an experiment on the kinetics of the first-order phase transition at the cavitation in superfluid ^{4}He [3] and normal liquid ^{3}He [4]. In the experiment [5], the large pressure oscillations in helium are produced by focusing ultrasonic waves at the center of the cell that has four windows providing an optical access in two (perpendicular) directions. The method is used to obtain a negative pressure region in the bulk of the liquid in order to avoid the surface nucleation. The cavitation is found to be a stochastic process. A significant cavitation rate is observed near the spinodal pressure.

The investigation of cavitation has a long history. First of all, this involves the investigation of the crossover from thermal to quantum behavior. According to the first estimates [6] of the cavitation rates at which bubbles nucleate in a liquid ^{4}He, it has been expected that quantum nucleation should dominate over the thermally activated one at temperatures (below ≈0.3 K) and that for this temperature range the pressure providing a noticeable nucleation rate or the tensile strength should be about $P \approx -15$ atm. Later, Maris and Xiong [7] attracted attention to the possibility that before this pressure can be achieved, the liquid ^{4}He becomes unstable against the long-wavelength density fluctuations once the square of the sound velocity becomes negative. The extrapolations of the sound velocity into the negative pressure range and some numerical calculations suggest that the sound velocity at the pressure P vanishes as

$$c(P) \propto (P - P_c)^{\nu}$$

with the exponent ν close to 1/3–1/4. The critical pressure P_c, i.e., the pressure at the spinodal point, was estimated as $P_c = -8$ to -9 atm at absolute zero for ^{4}He. For liquid ^{3}He, it was expected that $P_c = -2$ to -3 atm [8]. According to [9], the crossover temperature T^* from thermal to quantum behavior must be about 125 mK for ^{3}He and 220 mK for ^{4}He. The analysis is based only on the thermodynamic properties of ^{3}He and ^{4}He, i.e., on the chemical potentials and surface energies. The kinetic properties of the system (relaxation processes) were not involved.

The above result for T^* was also supported by the description of homogeneous and inhomogeneous states of liquid helium within the density–functional method [10]. In addition, the liquid–vapor phase diagram in ^{3}He and ^{4}He was analyzed with the help of this method [10] in the vicinity of the spinodal line. The spinodal pressures $P_c = -9$ atm for liquid ^{4}He and $P_c = -2$ to -3 atm for liquid ^{3}He were found. Monte

¶This article was submitted by the authors in English.

Carlo simulation of the critical behavior of liquid ^{4}He at a negative pressure in the vicinity of the spinodal curve [11] was performed.

Dissipation [2] and sound emission [12] strongly affect the underbarrier nucleation kinetics [13]. Recently, the dissipation effects in liquid ^{3}He have also been considered within the density–functional method using the hydrodynamic description for a metastable fluid [14].

The nucleation of bubbles in ^{4}He at negative pressures and temperatures down to 65 mK has been studied experimentally [5]. The results are consistent with the idea that nucleation is a result of the quantum tunneling through the potential barrier below 0.6 K. The quantum nucleation of bubbles occurs at a negative pressure (P_c = –9.23 bar, which is close to the spinodal pressure, although only 0.29 bar above). For normal ^{3}He, although the observed cavitation threshold is smaller and agrees with the prediction (P_c = –3.1 bar), the results are too preliminary to assert the quantum regime of the cavitation above 40 mK.

Here, we emphasize two points that permit us to understand the disagreement between theory [9] and experiment [5]. First of all, it is the energy dissipation during the underbarrier motion of a nucleus of the stable phase in normal ^{3}He that reduces the quantum nucleation rate. The second point is that the experiments are performed near the spinodal line, and the sound velocity vanishes at $P = P_c$. In this case, the kinetic energy of a growing bubble K decreases, and the crossover temperature T^* increases [12, 13]. This may be one reason why T^* equals 0.6 K in ^{4}He [4, 5] and not 0.2 K, as predicted in [7].

2. DISSIPATION AND SOUND EMISSION IN THE THEORY OF QUANTUM CAVITATION

We discuss the rate at which the bubbles can nucleate via quantum fluctuations in the normal ^{3}He and superfluid ^{4}He at negative pressures and sufficiently low temperatures. The energy dissipation due to viscosity and the sound emission due to compressibility of a fluid are involved in the bubble growth kinetics. Due to viscosity, the quantum cavitation kinetics in ^{3}He differs qualitatively from that in ^{4}He and corresponds to the dissipative tunneling in the overdamped regime. The compressibility results in increasing the cavitation rate and is essential in both liquids, especially, for the small critical bubbles responsible for the experimentally observable rates of the quantum cavitation.

All the calculations of the cavitation rate and tensile strength in the region of the quantum tunneling regime have been performed within the framework of the Lifshits–Kagan theory [1] of the first-order phase transitions. However, this theory neglects the compressibility of the metastable liquid; in other words, the sound velocity is taken to be infinite in the liquid. Clearly, a more realistic theory of the quantum cavitation must involve the effect of the finite compressibility, especially, in the vicinity of the instability point at which the sound velocity vanishes.

To investigate quantum-mechanical tunneling between the metastable and stable states of a condensed medium and to calculate the rate at which cavities nucleate, we use the formalism based on the finite action solutions (instantons) of equations continued to the imaginary time. This approach [15], elaborated for describing quantum-mechanical tunneling in the systems with a macroscopic number of degrees of freedom, was used for incorporating the influence of energy dissipation in a metastable condensed medium on the quantum kinetics of the first-order phase transitions at low temperatures [2].

The rate of the quantum nucleation can be written as

$$\Gamma(T) = \Gamma_0(T)\exp(-S(T)/\hbar) \tag{1}$$

where the prefactor Γ_0 is the rate at which cavitation is attempted per unit volume and unit time. According to the general notions of the nucleation kinetics, the prefactor Γ_0 can be evaluated approximately as an attempt frequency ν_0 multiplied by the number of centers at which independent cavitation events can occur.

In turn, the exponent S is the critical value of the effective Euclidean action [2]:

$$S_{\text{eff}}[R_\tau] = \int_{-\beta\hbar/2}^{\beta\hbar/2} d\tau\left[U(R_\tau) + \frac{1}{2}M(R_\tau)\left(\frac{dR}{d\tau}\right)^2\right] + \frac{1}{4\pi}\int_{-\beta\hbar/2}^{\beta\hbar/2}\int d\tau d\tau' \frac{(\pi T)^2}{\hbar^2\sin^2[\pi T(\tau-\tau')/\hbar]} \times [D(R_\tau, R_{\tau'}) - C(R_\tau, R_{\tau'})], \tag{2}$$

where $\beta = T^{-1}$ is the inverse temperature. The path $R(\tau)$, which is defined in the imaginary time τ, satisfies the periodic boundary conditions $R(-\beta\hbar/2) = R(\beta\hbar/2)$. It should be emphasized that all parameters of the effective action are unambiguously associated with the corresponding parameters in the classical equation of growth:

$$U'(R) + \mu_1(R)\dot{R} + M(R) \times\left[\ddot{R} + \frac{1}{2}\frac{M'(R)}{M(R)}\dot{R}^2\right] - \mu_3(R) \times\left[\dddot{R} + \frac{3}{2}\frac{\mu_3'(R)}{\mu_3(R)}\ddot{R}\dot{R} + \frac{1}{3}\left(\frac{\mu_3''(R)}{\mu_3(R)} - \frac{\mu_3'^2(R)}{2\mu_3^2(R)}\right)\dot{R}^3\right] + \ldots = 0. \tag{3}$$

The correspondence can readily be settled by analytically continuing ($|\omega_n| \longrightarrow -i\omega$) the Euler–Lagrange equation ($\delta S_{\text{eff}}/\delta R_\tau = 0$) for the effective action to real time, which gives the classical equation of growth. The

substitution $|\omega_n| \longrightarrow -i\omega$ of the Matsubara frequencies with the real ones must be performed in the frequency representation of the corresponding equations.

In the classical equation (3) for the growth of the supercritical droplet, we assume there is a limit for the low growth rate of the droplet and expand the rate in powers of the growth rate. The growth rate-independent term corresponds to the potential energy of the bubble. The linear term in the growth rate describes the energy dissipation, the second-order term corresponds to the kinetic energy of the droplet, and the third-order term can be attributed to the sound emission with the change of the bubble volume. Thus, it is clear that the first two terms in Eq. (2) can be referred to as the potential energy $U(R_\tau)$ and the kinetic energy with the mass $M(R_\tau)$ of the bubble. The other terms are nonlocal in time and are due to the energy dissipation $D(R_\tau, R_{\tau'})$ during the bubble growth and the sound emission $C(R_\tau, R_{\tau'})$ originating from the finite compressibility. The energy dissipation is connected with the vortex $\gamma_1(R)$ [2, 13] as

$$D(R_\tau, R_{\tau'}) = (\gamma_1(R_\tau) - \gamma_1(R_{\tau'}))^2 \tag{4}$$

and $\gamma_1(R)$ is unambiguously determined by the friction coefficient

$$\mu_1(R) = (\partial\gamma_1(R)/\partial R)^2. \tag{4a}$$

We would like to make an important remark concerning the behavior of the friction coefficient $\mu_1(R)$ as a function of the bubble radius and temperature. In the hydrodynamic approximation, the bubble radius must be much larger than the mean free path $l(T)$ of excitations in the medium surrounding the bubble. Since the mean free path increases rapidly at low temperatures (in particular, $l(T) \propto 1/T^2$ for ^{3}He), the crossover from the hydrodynamic $R \gg l$ regime to the ballistic or Knudsen regime with $R \ll l$ must occur. Depending on whether the hydrodynamic or ballistic regime occurs, we arrive at the general expression for the friction coefficient $\mu_1(R)$:

$$\mu_1(R) = 16\pi\eta R f(R/l),$$

where η is the viscosity coefficient and

$$f(x) = \begin{cases} 1, & x \gg 1 \\ ax, & x \ll 1 \end{cases} \tag{4b}$$

is a dimensionless function of the ratio of the bubble radius to the mean free path of excitations in the liquid. The numerical factor a is of the order of unity, depends on the specific features of the interaction of excitations with the bubble surface, and can be calculated explicitly using the kinetic equation.

It should be noted that the friction coefficient $\mu_1(R)$ in the ballistic $R \ll l$ regime is independent of the mean free path $l(T)$ since $\eta \sim \rho c l$, where ρ is the density of the liquid. In this case, $\gamma_1(R)$ does not depend on the temperature:

$$\gamma_1(R) = \sqrt{4\pi a\eta/l}R^2. \tag{5}$$

This ballistic regime with the temperature-independent $\gamma_1(R)$ is the only possible underbarrier motion of the nucleus because the opposite case (where $R > l$) implies a large critical radius R_c within the whole range of temperatures outside a close vicinity of T_λ. The large critical radius R_c leads to a negligible decay rate of the metastable liquid and to the impossibility of recording it experimentally. Thus, for the underbarrier motion of the cavity, we can always assume the validity of Eq. (5).

The second nonlocal term in Eq. (2),

$$C(R_\tau, R_{\tau'}) = \left[\frac{\partial\gamma_3(R_\tau)}{\partial\tau} - \frac{\partial\gamma_3(R_{\tau'})}{\partial\tau'}\right]^2, \tag{6}$$

is responsible for the excitation and emission of sound waves in the course of the underbarrier growth of a bubble and $\gamma_3(R)$ is determined by the kinetic coefficient $\mu_3(R)$:

$$\mu_3(R) = (\partial\gamma_3(R)/\partial R)^2.$$

The corresponding coefficient $\mu_3(R)$ is given by

$$\mu_3(R) = \frac{4\pi\rho}{c}R^4,$$

which leads to

$$\gamma_3(R) = \frac{2}{3}\sqrt{\frac{\pi\rho}{c}}R^3. \tag{7}$$

It is interesting to note that in contrast to the term with the Ohmic dissipation $D(R_\tau, R_{\tau'})$ related to the dissipative function that is proportional to the square of the first-order time derivative, the term $C(R_\tau, R_{\tau'})$ due to the finite compressibility of the fluid medium gives a negative contribution into the effective action [Eq. (2)]. This results in enhancing the quantum nucleation rate compared with the one calculated in the framework of the Lifshits–Kagan model [1] of an incompressible fluid. Some hints for this conclusion can be seen from the fact that the finiteness of the sound velocity limits the region of the bubble environment that can be disturbed and set into motion. The size of this region is approximately equal to $\Lambda = c\tau$, where τ is the typical time of growth. In some sense, one can say that the total kinetic energy of the fluid flowing away from the expanding bubble becomes smaller than for the incompressible fluid where the perturbation induced by the formation of the bubble extends instantaneously to infinity.

The kinetic energy can be described in terms of the variable mass of the bubble:

$$M(R) = 4\pi\rho R^3. \tag{8}$$

It can be attributed to the kinetic energy of the fluid that flows away from the bubble. In the case of cavitation, the potential energy can be represented as

$$U(R) = \frac{4\pi}{3}PR^3 + 4\pi\alpha R^2, \quad (8a)$$

where α is the liquid–gas surface tension.

Equations (1) and (2) with the coefficients (4)–(8) allow us to calculate the rate of the underbarrier motion of the bubble. It should be emphasized that all parameters of the effective action are unambiguously associated with the corresponding parameters in the classical equation of growth (3).

3. THE MANIFESTATION OF COMPRESSIBILITY IN THE EXPERIMENT ON QUANTUM CAVITATION IN SUPERFLUID ^{4}HE

Recently [5], cavitation was studied in superfluid ^{4}He and normal liquid ^{3}He experimentally. The investigation of cavitation in these liquids is related to the possibility of avoiding impurities, which usually manifest themselves as centers of cavitation. On the other hand, these liquids have essentially different properties; namely, ^{4}He is a superfluid liquid and ^{3}He is a normal viscous Fermi liquid in the experimental range of temperatures from 40 to 1000 mK. The cavitation process is induced by sound pulses at a frequency ω close to 1 MHz and is focused in the center of the experimental cell. The pulses create oscillations of local pressure within several bars of the static pressure. The typical size of the acoustic focus is ~0.12 mm and the size of the experimental cell is 8 mm. The above-mentioned limiting temperature (40 mK) is connected with thermal radiation due to these sound pulses with a short duration between 30 and 70 μs and a repetition rate in the range 0.1–1 Hz.

The cavitation process is observed to be stochastic. For the invariant temperature and pressure parameters, some sound pulses of a given amplitude produce cavitation and some pulses of the same amplitude do not. Applying several sound pulses and counting the number of cavitation events, one can determine the probability Σ of cavitation as a function of the applied voltage and temperature. According to [5], it is shown that the cavitation probability in ^{4}He depends on temperature only above about 400 mK. One of the difficulties in interpreting the experiment is related to the fact that the maximum of the sound attenuation exists in the system in this temperature range. An increase of the sound attenuation results in the fact that a larger voltage is needed to produce the same pressure swing at the acoustic focus where cavitation occurs. After the correction, the cavitation voltage is found to be independent of the temperature up to 600 mK. Above this temperature, the voltage decreases as T increases, corresponding to a thermally activated nucleation. This experimental result can be interpreted as a crossover from the quantum cavitation below 600 mK to the thermally activated cavitation. It should be emphasized [5] that the stochastic behavior of the nucleation process combined with the temperature-independent behavior of the voltage at which the cavitation occurs is in contrast to the assumption that the spinodal pressure is achieved, because this pressure cannot result in the stochastic behavior of the cavitation process.

There is one more difficulty in interpreting the experiment. The temperature of the cell that is measured in the experiment [5] may be different from the temperature in the focus where the cavitation occurs. The point is that the acoustic wave is adiabatic in the first approximation; consequently, the temperature and pressure oscillate at the focus. Within the adiabaticity assumption, the temperature in the focus of the sound wave can readily be estimated [5]. At temperatures below 0.7 K, phonons make the dominant contribution into the entropy per unit mass [16]:

$$S \approx S_{\rm ph} = \frac{2\pi^2 T^3}{45\rho^3 c^3}. \quad (9)$$

In the isentropic process, the temperature is therefore proportional to the velocity of sound c. It was shown experimentally [5] that near the spinodal at $P = -9.23$ bar, the sound velocity is 74 m/s, which is lower than that at zero pressure (c = 238 m/s) by a factor of three. As a result, the local instantaneous temperature T must be reduced at the focus by the same factor with respect to the static temperature $T_{\rm stat}$ of the cell [5].

This interpretation is consistent only if the following two conditions are fulfilled. The first condition is related to the well-known fact that the nonlinear effects arise very early in an alternating field [17]. The typical field in which the nonlinear effects arise is proportional to the exponential

$$V = \tilde{V}\exp\{\omega\tau_0\}, \quad (10)$$

where ω is the sound frequency, τ_0 is a typical time of the underbarrier motion, and V is the sound amplitude. In the experiment [5], $\omega = 1$ MHz, and in the experimentally analyzed vicinity of the spinodal line, τ_0 can be estimated as $\tau_0 = 10^{-10}$ s^{-1}. Thus, in the experiment range where $\omega\tau_0 \ll 1$, the nonlinear effects can be neglected. The second condition is much more severe, meaning that T entering Eq. (9) follows local variations of the pressure in space and time in the sound wave. The conditions can be represented as

$$\omega\tau \ll 1 \text{ or } l \ll \lambda. \quad (11)$$

These conditions are essentially equivalent. The second inequality can be obtained from the first by multiplying it by the velocity of sound c. The second condition means that the sound wavelength λ must be much larger than the mean free path l. In the experiment [5], the opposite condition is fulfilled within the entire temperature range. The size of the acoustic focus is

~0.12 mm, and this distance is much less than the mean free path l_{ph} for the phonon–phonon scattering, which equals 1.3 mm at 0.7 K. According to [16], $\tau_{ph}^{-1} \approx 6 \times 10^6 T^7$ s (with T measured in K), $l_{ph} \simeq c\tau_{ph}$, with the value $c = 74$ m/s near the spinodal line used in the estimate. Moreover, l_{ph} increases as T^{-7} with a decreasing temperature and becomes about 15 mm at $T = 0.5$ K, which exceeds the size of the experimental cell of 8 mm. The other scattering processes, in particular, the phonon–roton and roton–roton scatterings, are inefficient at low temperatures for the relaxation to a local equilibrium because of freezing rotons. Thus, the local temperature T in the sound wave cannot follow the variations of the pressure in the sound wave in this range of temperatures, with the entropy S in Eq. (9) being conserved. The local temperature in the focus of the sound wave is therefore equal to the temperature outside the focus; i.e., $T^* = 0.6$ K at the crossover point from thermal to quantum behavior of the kinetics of bubble nucleation.

We are now able to compare the thermal–quantum crossover temperature T^* obtained experimentally with the calculations. We start from the simplest estimate that can be obtained from the first two terms of Eq. (2). These two terms are the potential and kinetic energy of the growing cavity and correspond to the Lifshits–Kagan analysis [1]. In the case of cavitation, the crossover temperature reduces to the following equation with the known parameters:

$$T^* = \frac{256\hbar}{405\pi\sqrt{6}}\frac{|P|^{3/2}}{\alpha\sqrt{\rho}}. \tag{12}$$

The substitution of the ^{4}He data $\rho = 0.095$ g/cm^3, $\alpha = 0.37$ erg/cm^2, and the experimental value $P = -9.5$ bar near the spinodal line gives $T^* = 0.15$ K. The estimate used is a thin-wall approximation where the bubble is assumed to have a sharp surface of radius R forming the boundary between an empty interior and the bulk liquid surrounding the bubble. A more elaborate calculation for the bubbles with a radius that is comparable with the interface thickness [9] uses the density–functional approximation for the energy of the metastable liquid and gives $T^* = 0.2$ K. The insignificant difference between these two approximations is not surprising because they both are based on the same value of the surface energy α [1, 9]. The difference between these two opposite estimates is less than the experimental value, which, as is emphasized, should be taken as $T^* = 0.6$ K instead of $T^* = 0.2$ K, as was assumed in [5]. The results are insensitive to the inclusion of the third term $D(R_\tau, R_{\tau'})$ in Eq. (2), which describes dissipation, because the dissipation in ^{4}He is negligible at low temperatures. Moreover, this leads to lowering T^* and to a deviation of its value from the experimental result $T^* = 0.6$ K. However, the term involving $C(R_\tau, R_{\tau'})$ in Eq. (2) leads to the opposite and important effect of increasing T^*. If we consider the term with $C(R_\tau, R_{\tau'})$ in Eq. (2) as a perturbation, we obtain the following expression for effective action S_{eff} in Eq. (2) at the low-temperature limit:

$$S_{eff} = \frac{5\sqrt{2}\pi^2}{16}(\alpha\rho)^{1/2}R_c^{7/2}\left(1 - \frac{4}{9c}\sqrt{\frac{2\alpha}{\rho R_c}}\right), \qquad R_c = \frac{3\alpha}{|P|}. \tag{13}$$

This expression differs from that for S_{eff} in [1] only by the factor $(1 - \dot{R}/c)$, where $\dot{R} = (2/3)^{5/2}(|P|/\rho)^{1/2}$ is the rate of the underbarrier growth of the cavity. Substituting the data $\rho = 0.095$ g/cm^3, $P = -9.5$ bar, and $c = 74$ m/s, we obtain $\dot{R}/c = 0.48$; hence, T^* increases approximately twofold and equals $T^* = 0.4$ K. We assert that the tendency of increasing T^* due to a finite compressibility of ^{4}He and the underbarrier sound emission during cavitation is the reason for the high value of T^* observed experimentally. The manifestation of the phenomenon is strongly related to a high value of the ratio $\dot{R}/c$, which approximately equals 1/2. The high value is directly related with the experimental conditions [3, 5] of the cavitation taking place in the vicinity of the spinodal pressure. For $P = 0$, this ratio is only about 0.1. In any case, more elaborate considerations should be used in analyzing the phenomenon because the leading approximation in $\dot{R}/c \ll 1$ is assumed for the derivation of Eq. (2).

4. THE MANIFESTATION OF DISSIPATION IN THE EXPERIMENT ON QUANTUM CAVITATION IN NORMAL FLUID ^{3}He

We now turn to the analysis of the experiments on quantum cavitation in liquid ^{3}He. This is a normal viscous Fermi liquid within the experimental range of temperatures from 40 to 1000 mK [4]. The simplest estimate for T^* in Eq. (12) gives $T^* = 0.09$ K for $\rho = 0.054$ g/cm^3, $\alpha = 0.16$ erg/cm^2, and the pressure $P = -3.1$ bar near the spinodal of ^{3}He. A more accurate calculation for the bubbles with radii comparable to the interface thickness uses the density–functional approximation for the energy of the metastable liquid and gives $T^* = 0.125$ K [9]. However, the crossover to the quantum behavior is not observed experimentally until $T = 0.04$ K [4].

Both the above estimates are based on the first two terms in the effective action S_{eff} in Eq. (2), which include only the potential and kinetic energies in different approximations and ignore the fact that liquid ^{3}He is viscous. The viscosity η behaves as T^{-2} with the temperature since ^{3}He is a Fermi liquid and $\eta \sim \rho v_F l_F$, where ρ is the density of ^{3}He, v_F is the Fermi velocity, and $l_F \sim v_F\tau_F$ is the mean free path. Here, τ_F is the col-

lision time for excitations in the Fermi liquid and $\tau_F \sim D\hbar\varepsilon_F/T^2$, where ε_F is the Fermi energy, $D \sim (p_F a_0/\hbar)^{-2}$ is a dimensionless coefficient, and a_0 is the scattering length [18]. Using the expression for the Fermi momentum $p_F = \hbar(3\pi^2\rho/m)^{1/3}$ and $\varepsilon_F = p_F^2/2m^*$, with m being the mass of the ^{3}He atom and m* being the effective mass such that $m^*/m = 3.08$, and substituting $\rho = 0.054$ g/cm^3 near the spinodal, we obtain $p_F/\hbar = 0.68 \times 10^8$ cm^{-1}. If we put $D \sim 0$ 0.15, we obtain $\tau_F = 1.2 \times 10^{-12}T^{-2}$ s (with T expressed in K), which differs from the value for τ_F obtained from the viscosity [16] only by a factor of 1.3 due to the difference between the density of ^{3}He near the spinodal line ($\rho = 0.054$ g/cm^3) and the density $\rho = 0.082$ g/cm^3 at pressure $P = 0$. For l_F, we have

$$l_F \simeq A/T^2,$$

$$A = \hbar^4 \frac{3\pi^2}{2}\left(\frac{m^*}{m}\right)^{-2}\frac{D\rho}{m^3}, \tag{14}$$

$$l_F \simeq 0.5 \times 10^{-8}T^{-2} \text{ cm } (T \text{ in K}).$$

We see that l_F, is about 0.5 Å at a temperature of $T = 1$ K. In this case, $l_F \ll R_c$, because $R_c \sim 10$ Å according to [1] (see also the introduction to [19]). Upon lowering the temperature, the mean free path l_F grows drastically as T^{-2} and l_F becomes ~50 Å at $T = 0.1$ K for $l_F \gg R_c$. Thus, within the temperature range 1–0.1 K, the behavior of the nucleation of bubbles varies from the hydrodynamic type to the ballistic one. We can introduce [2] the temperature T_l at which the mean free is path $l_F \simeq R_c$, and the hydrodynamic nucleation type is replaced by the ballistic one:

$$T_l = \sqrt{A/R_c}. \tag{15}$$

For cavitation in ^{3}He near the spinodal, $T_l = 0.18$ K. At this point, we go over from one type of dissipation in the system to another. Above $T > T_l$, the nucleation is governed by the hydrodynamic flow of viscous Fermi liquid (4b) (with $R_c/l_F \gg 1$) and for $T < T_l$ the ballistic propagation of excitations in the Fermi liquid occurs (4b) (with $R_c/l_F \ll 1$). For $T > T_l$, the viscosity η enters the dissipation at the bubble nucleation. Inserting the above estimate into the expression for η, we obtain from [16] $\eta = \alpha_0/T^2$, with $\alpha_0 \sim 10^{-6}$ pois. For $T < T_l$, the dissipation is governed by η/l_F.

We see that near the spinodal in ^{3}He, T_l is higher than T^*, which is about 0.1 K according to the estimate without the dissipation processes at nucleation being taken into account. The involvement of dissipation only reduces T^*. In any case, therefore, the quantum cavitation is accompanied by the dissipation of the Knudsen type. We can now compare the value of the effective action S_{eff} [Eq. (2)] in the dissipationless case (13) with the one involving the dissipation, $S_{\text{diss}} \approx \mu_1 R_c^2$. For $T < T_l$ ($R_c/l_F \ll 1$), this ratio is $q = S_{\text{kin}}/S_{\text{diss}} \sim 0.05(P/\rho)^{1/2}v_F^{-1} \sim 0.08$ for the above-mentioned values P and ρ near the spinodal of ^{3}He. This estimate means that the growth of the bubble is accompanied by a strong dissipation corresponding to the overdamped quantum regime, and the thermal–quantum crossover temperature is $T^* = \hbar U_0/S_{\text{diss}}$, with $U_0 = (16/27)\pi\alpha R_c^2$ being the height of the potential barrier. For the dissipation of the Knudsen type, we thus obtain T^* as

$$T^* = s^{-1}\frac{m^*}{3^5\pi}\frac{\hbar}{p_F}\frac{P^2}{\rho\alpha}, \tag{16}$$

where s is a coefficient that depends only on the dissipation type. In what follows, we see that for the ballistic propagation of excitations in the metastable environment, $s = s_b \approx 1.2$ and T^* becomes about 2 mK. Note that $\hbar$ does not enter into the expression for T^* in Eq. (16) because $p_F/\hbar$ depends only on the density ρ, i.e., approximately $p_F/\hbar = (3\pi^2\rho/m)^{1/3}$. The absence of $\hbar$ in T^* is related to the overdamped ballistic regime of dissipation. In this case, the dissipation is proportional to η/l, which is of the order ρv_F and is proportional to $\hbar$. Thus, $\hbar$ does not enter T^* because of the purely quantum nature of the dissipation in the Fermi liquid. A formal reduction of the dissipation ($\hbar \longrightarrow 0$) leads to a dissipationless behavior, where $\hbar$ enters again in T^*.

To determine s_b, we can use only two terms of the effective action $S_{\text{eff}}[R_\tau]$ in Eq. (2), namely, the potential energy $U(R_\tau)$ and the nonlocal dissipative term $D(R_\tau, R_{\tau'})$, because the term with the kinetic energy is small in the case of a strong dissipation and is proportional to $q \sim 0.08$. We can reduce $S_{\text{eff}}[R_\tau]$ to the dimensionless action $s_b(t)$ (see, e.g., [2]):

$$s_b[x_\tau] = \int_{-1/2t}^{1/2t} d\tau x_\tau^2(1 - x_\tau) + \frac{\pi t^2}{4}\int_{-1/2t}^{1/2t} d\tau' \frac{(x_\tau^2 - x_{\tau'}^2)^2}{\sin^2[\pi t(\tau - \tau')]}. \tag{17}$$

The numerical calculation of (17) gives $s = s_b \approx 1.2$.

5. CONCLUSIONS

We would like to emphasize the qualitative feature whereby the normal liquid ^{3}He differs from the superfluid ^{4}He [2], namely, the dissipation of energy in the course of quantum cavitation for normal ^{3}He. In the absence of dissipation, only the kinetic energy K of the motion of the metastable liquid governs the underbarrier dynamics of the growing bubble. The kinetic energy can be described in terms of the variable mass

of the bubble $M(R) = 4\pi\rho R^3$ as $K = M\dot{R}^2/2$ [1]. The underbarrier motion of the bubble in ^{4}He corresponds to the dynamic motion indicated above if we disregard phonon excitations. Anyway, this is true at low temperatures. In ^{3}He, we must also take the viscous motion of the normal Fermi liquid into account. This leads to the appearance of the term with the energy dissipation in the bubble expansion equation.

The growth rate of the bubble is determined by the interplay of the kinetic energy K and the energy dissipation. It should be emphasized that there are no free parameters in ^{3}He that can determine the relative contribution of these two terms. The kinetic energy K is of the order $M(R)\dot{R}^2/2$ and should be compared with $\mu(R)R\dot{R}$ from the energy dissipation. The ratio of K to the energy dissipation is of the order $\dot{R}/v_F \ll 1$. The last condition is connected with the approximate relation $v_F \sim c$, and v_F differs from the velocity of sound c only by a numerical coefficient of about $2/(\sqrt{3}\pi^{1/3})$. Our consideration assumes a slow growth rate of the bubble: $\dot{R} \ll c$. We would like to emphasize that the dynamics of the subbarrier motion of the bubble is governed by the energy-dissipation power rather than by the kinetic energy K. This implies the overdamped regime of quantum cavitation. Accordingly, we have the exponent $\mu(R_c)R_c^2$ in the growth rate instead of $\sqrt{2M(R_c)U(R_c)}$, as in the absence of dissipation. In addition,

$$\mu(R_c)R_c^2 \gg \sqrt{2M(R_c)U(R_c)}.$$

Thus, because of dissipation, the crossover temperature T^* decreases and becomes lower than the temperature of the ^{3}He transition into superfluidity. This manifests itself as T^* because of the lack of dissipation in the superfluid state. This is why the crossover from thermal to quantum behavior was not found [4].

For the understanding of the experiments on quantum cavitation in superfluid ^{4}He [3], it is important to incorporate compressibility and sound emission into the equation of the bubble growth. The point is that the experiments are performed near the spinodal line, and the sound velocity vanishes at $P = P_c$. In this case, the kinetic energy K decreases and T^* increases. This is the reason why T^* equals 0.6 K in ^{4}He [3, 5] and not 0.2 K, as predicted in [9].

ACKNOWLEDGMENTS

The work is supported by the Russian Foundation for Basic Research, project nos. 99-02-17289 and 98-02-16226, by the Russian Ministry for Higher Education, project no. 97-0-14.0-80, and Grant for the Promotion of Universities of Russia.

REFERENCES

1. I. M. Lifshits and Yu. Kagan, Zh. Éksp. Teor. Fiz. **62**, 385 (1972) [Sov. Phys. JETP **35**, 206 (1972)].
2. S. N. Burmistrov and L. B. Dubovskiĭ, Zh. Éksp. Teor. Fiz. **66**, 733 (1987) [Sov. Phys. JETP **66**, 414 (1987)].
3. S. Balibar, F. Caupin, P. Roche, and H. J. Maris, J. Low Temp. Phys. **113**, 459 (1998); M. S. Pettersen, S. Balibar, and H. J. Maris, Phys. Rev. B **49**, 12 062 (1994).
4. F. Caupin, P. Roche, S. Marchand, and S. Balibar, J. Low Temp. Phys. **113**, 473 (1998).
5. H. Lambare, P. Roche, G. Guthmann, *et al.*, Eur. Phys. J. B **2**, 381 (1998).
6. V. A. Akulichev and V. A. Bulanov, Akust. Zh. **20**, 817 (1974) [Sov. Phys. Acoust. **20**, 501 (1975)].
7. H. J. Maris and Q. Xiong, Phys. Rev. Lett. **63**, 1078 (1989).
8. Q. Xiong and H. J. Maris, J. Low Temp. Phys. **77**, 347 (1989); **82**, 105 (1991); M. S. Pettersen, C. Naud, S. Balibar, and H. J. Maris, Physica B (Amsterdam) **194–196**, 575 (1994).
9. H. J. Maris, J. Low Temp. Phys. **98**, 403 (1995).
10. M. Guilleumas, M. Pi, M. Barranco, *et al.*, Phys. Rev. B **47**, 9116 (1993); D. M. Jezek, M. Guilleumas, M. Pi, *et al.*, Phys. Rev. B **48**, 16 582 (1993); M. Guilleumas, M. Barranco, D. M. Jezek, *et al.*, Phys. Rev. B **54**, 16 135 (1996); D. M. Jezek, M. Guilleumas, M. Pi, and M. Barranco, Europhys. Lett. **38**, 601 (1997).
11. C. E. Campbell, R. Folk, and E. Krotscheck, J. Low Temp. Phys. **105**, 13 (1996).
12. S. E. Korshunov, Fiz. Nizk. Temp. **14**, 575 (1988) [Sov. J. Low Temp. Phys. **14**, 316 (1988)].
13. S. N. Burmistrov and L. B. Dubovskiĭ, J. Low Temp. Phys. **96** (1994); Fiz. Nizk. Temp. **23**, 527 (1997) [Low Temp. Phys. **23**, 389 (1997)].
14. D. M. Jezek, M. Pi, and M. Barranco, Phys. Rev. B **60**, 3048 (1999).
15. A. O. Caldeira and A. J. Leggett, Phys. Rev. Lett. **46**, 211 (1981); Ann. Phys. **149**, 374 (1983); A. I. Larkin and Yu. N. Ovchinnikov, Pis'ma Zh. Éksp. Teor. Fiz. **37**, 322 (1983) [JETP Lett. **37**, 382 (1983)]; Zh. Éksp. Teor. Fiz. **86**, 719 (1984) [Sov. Phys. JETP **59**, 420 (1984)].
16. I. M. Khalatnikov, *An Introduction to the Theory of Superfluidity* (Nauka, Moscow, 1965; Addison–Wesley, Redwood City, 1988); *Theory of Superfluidity* (Nauka, Moscow, 1971).
17. B. I. Ivlev and V. I. Mel'nikov, in *Modern Problems in Condensed Matter Sciences* (North-Holland, Amsterdam, 1992), Vol. 34.
18. E. M. Lifshitz and L. P. Pitaevskiĭ, in *Statistical Physics* (Nauka, Moscow, 1978; Pergamon, Oxford, 1980), Part 2.
19. S. N. Burmistrov, L. B. Dubovskii, and V. L. Tsymbalenko, J. Low Temp. Phys. **90**, 363 (1993).

Journal of Experimental and Theoretical Physics, Vol. 91, No. 4, 2000, pp. 775–785.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 4, 2000, pp. 863–876.

FLUIDS

Collapse of Vortex Lines in Hydrodynamics

E. A. Kuznetsov* and V. P. Ruban**

Landau Institute for Theoretical Physics, Russian Academy of Sciences, Moscow, 117334 Russia
**e-mail: kuznetso@itp.ac.ru*
***e-mail: ruban@itp.ac.ru*
Received May 18, 2000

Abstract—A new mechanism is proposed for collapse in hydrodynamics associated with the "breaking" of vortex lines. The collapse results in the formation of point singularities of the vorticity field, i.e., a generalized momentum curl. At the point of collapse the vorticity $|\mathbf{\Omega}|$ increases as $((t_0 - t)^{-1}$ and its spatial distribution for $t \longrightarrow t_0$ approaches quasi-two-dimensional: in the "soft" direction contraction obeys the law $l_1 \propto (t_0 - t)^{3/2}$ whereas in the other two "hard" directions it obeys $l_2 \propto (t_0 - t)^{1/2}$. It has been shown that this collapse scenario takes place in the general case for three-dimensional integrable hydrodynamics with the Hamiltonian $\mathcal{H} = \int |\mathbf{\Omega}|\, d\mathbf{r}$. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The problem of collapse in the hydrodynamics of an ideal incompressible fluid as a process involving the formation of a singularity over a finite time is one of the central problems in the theory of developed hydrodynamic turbulence. It is well known that power-type singularities produce power tails in the short-wavelength region in the turbulence spectrum and consequently collapses are important from the point of view of the Kolmogorov spectra [1]. Classical examples of these spectra are the Phillips spectrum [2] for wind waves and the Kadomtsev–Petviashvilli spectrum of acoustic turbulence [3]. In the first case "whitecaps" or surface tips appear as singularities while in the second case, these are density discontinuities (shock waves).

The problem of collapse in hydrodynamics has been around for a long time. For example, in 1981 Saffman [4] considered this problem as one of the key issues in hydrodynamics (see also [6] and the literature cited there) although it seems to us that L.F. Richardson and A.N. Kolmogorov must have understood the importance of this problem. Despite the long history of this problem, no significant understanding of the nature of collapse in hydrodynamics has yet been obtained, although a fairly large number of numerical experiments have now indicated that it does exist (as will be discussed below). As for theory, no important results have been achieved and no common viewpoint exists on the actual object, collapse in incompressible hydrodynamics: the view has been put forward that collapse is generally impossible (see, for example Section 7.8 of the book [5] and the literature cited there). The only exception in our view is a study by Zakharov in 1988 [7] (a more detailed study appeared in 1999 [8]) in which he constructed a systematic theory of the collapse of two antiparallel vortex filaments of finite but small thickness in the quasi-two-dimensional approximation where the flow is mainly two-dimensional and the dependence along the third coordinate is slow (see [9]).

Considerable progress in studying hydrodynamic collapse can be attributed to numerical modeling of the Euler equations. A considerable number of numerical experiments indicate that the vorticity $|\mathbf{\Omega}|$ goes to infinity at various points over finite time. It follows from [10–13] that $|\mathbf{\Omega}|$ increases at the point of collapse as $(t_0 - t)^{-1}$ where $t_0$is the time of collapse. According to [10, 13], the spatial scale of the collapsing distribution decreases as $(t_0 - t)^{-1/2}$. In a recent study Kerr [14] reports anisotropy of the collapsing region. An analysis of the numerical data yielded two scales, one decreasing as the square root: $l_1 \propto (t_0 - t)^{1/2}$ and the other decreasing linearly with time: $l_2 \propto t_0 - t$. It should be noted that in most known numerical experiments the initial flow possessed some symmetry or was close to symmetric resulting in the appearance of several singularities. For instance, Kerr [10] examined the evolution of two antiparallel vortex tubes when collapse is attributed to Crow instability [15] which at the nonlinear stage leads to reconnection of vortex lines even when viscosity is taken into account. Consequently, collapse was observed at two symmetric points.

In this paper, study we propose a new mechanism for the formation of collapse associated with "breaking" of the continuous distribution of vortex lines. Importantly this mechanism is not related to any symmetry of the initial distribution and collapse can occur at a single isolated point. Probably, it is this type of collapse that was recently observed in the numerical calculations [16]. The mechanism can be naturally incorporated into the classical catastrophe theory [17]. Col-

lapse is represented as a process of caustic formation for a solenoidal vector field of the vorticity. Understanding how collapse occurs directly in a Euler description is not easy. This is primarily related to a hidden symmetry of the Euler equation, i.e., the relabeling symmetry (for more details see the reviews [18, 19]). This symmetry generates a vector quantity which is conserved at each Lagrangian point, a Cauchy invariant which is expressed in terms of the velocity curl and the Jacobi matrix of the mapping from Eulerian to Lagrangian variables and for this reason is strongly nonlocal in terms of liquid velocity. The Cauchy invariant, however, is known as an invariant characterizing the frozenness of vortex lines into the fluid while all the conservation laws for vorticity: the Kelvin theorem, Ertel invariant, and the Hopf topological invariant characterizing the knotted nature of the flow, are a simple consequence of this invariant. The frozenness of the vortex lines implies that liquid particles are attached to a given vortex line and never leave it. The frozen-in property can only be destroyed as a result of viscosity, i.e., beyond the limits of ideal hydrodynamics. Thus, the next natural step in the description of vortex motion was to introduce a mixed Lagrangian–Eulerian description where the main object is the vortex line [20, 21]. Each vortex line is numbered by its (two-dimensional) Lagrangian marker and the third coordinate is the parameter determining the curve. This representation which we call the vortex line representation is the key for describing collapse in hydrodynamics as the formation of the caustic of a solenoidal field of velocity curl.

This article is set out as follows. In Section 2 we introduce the concept of vortex lines and explain its meaning. Section 3 is devoted to three-dimensional integrable hydrodynamics which we introduced in [20]. Its Hamiltonian is unusual and is expressed in terms of the modulus $|\mathbf{\Omega}|$:

$$\mathcal{H} = \int |\mathbf{\Omega}| d\mathbf{r}. \tag{1}$$

This model can be integrated using the vortex line representation and the inverse scattering method. In the vortex line representation, the Hamiltonian (1) is decomposed into the sum of the Hamiltonians of noninteracting vortex lines. The dynamics of each vortex line is described by an exactly integrable one-dimensional Landau–Lifshitz equation for a Heisenberg ferromagnet or its gauge-equivalent analog, a nonlinear Schrödinger equation. Integrable hydrodynamics is thus the hydrodynamics of free vortex filaments. For the hydrodynamics of free particles—zero-pressure hydrodynamics (see, for example, [22]) and for integrable hydrodynamics the appearance of a caustic-type singularity is typical. For zero-pressure dynamics the density in the caustic goes to infinity. The density, however, is a scalar characteristic unlike the solenoidal vector field $\mathbf{\Omega}$. This last factor imposes specific constraints on the field structure near the singularity. It is shown in Section 4 that the structure of the singularity is strongly anisotropic and acquires a pancake shape. The spatial distribution near the point of collapse at $t \longrightarrow t_0$ approaches quasi-two-dimensional: compression in the "soft" direction obeys the law $l_1 \propto (t_0 - t)^{3/2}$ while in the two other "hard" directions it obeys $l_2 \propto (t_0 - t)^{1/2}$. At the point of collapse the vorticity lies in the "pancake" plane and its value $|\mathbf{\Omega}|$ increases as $(t_0 - t)^{-1}$. This behavior is consistent with the general position. In the degenerate case considered in the last part of the article we consider the collapse of a topologically nontrivial axisymmetric vorticity distribution, known as a Hopf mapping, when any two vortex lines are singly linked. In this case, at the point of collapse the two eigenvalues of the Jacobi matrix go to zero and as a result the vorticity at the point of collapse appears to have a stronger singularity: $|\mathbf{\Omega}| \propto (t_0 - t)^{-2}$. The conclusions are devoted to a discussion of numerical experiments to observe collapse for Euler hydrodynamics and their consistency with the theory presented here.

2. VORTEX LINE REPRESENTATION

We shall analyze hydrodynamic equations

$$\frac{\partial \mathbf{\Omega}}{\partial t} = \operatorname{curl}\left[\operatorname{curl}\frac{\delta \mathcal{H}}{\delta \mathbf{\Omega}} \times \mathbf{\Omega}\right], \tag{2}$$

where $\mathcal{H}\{\mathbf{\Omega}\}$ is the Hamiltonian of the system, $\mathbf{\Omega}(\mathbf{r}, t) = \operatorname{curl} \mathbf{p}(\mathbf{r}, t)$ is the generalized vorticity field which comprise the curl of the canonical momentum $\mathbf{p}$. The vector field

$$\mathbf{v} = \operatorname{curl}(\delta \mathcal{H}/\delta \Omega) \tag{3}$$

is none other than the liquid velocity. By definition, $\operatorname{div}\mathbf{v} = 0$, i.e., we are dealing with an incompressible medium. If the Hamiltonian is the same as the kinetic energy of the fluid

$$\mathcal{H} = \int \frac{\mathbf{p}^2}{2} d\mathbf{r} = \iint \frac{\mathbf{\Omega}(\mathbf{r}_1) \cdot \mathbf{\Omega}(\mathbf{r}_2)}{8\pi |\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_1 d\mathbf{r}_2,$$

expression (3) gives the ordinary relationship $\mathbf{\Omega} = \operatorname{curl} \mathbf{v}$ between the velocity $\mathbf{v}$ and the vorticity $\mathbf{\Omega}$, and Eq. (2) is transformed into the Euler equation for the vorticity $\mathbf{\Omega}$:

$$\frac{\partial \mathbf{\Omega}}{\partial t} = \operatorname{curl}[\mathbf{v} \times \Omega], \quad \operatorname{div}\mathbf{v} = 0.$$

An important property of the Eqs. (2) is that the vorticity is frozen into the liquid so that all Lagrangian particles remain on their oun natural vortex line for $t > 0$. It is therefore natural to introduce a mixed Lagrangian–Euler description in terms of vortex lines where each vortex line is numbered by its (two-dimensional) Lagrangian label ν ascribed to some two-dimensional manifold $\mathcal{N}$, and the parameter s determining each vortex line has the meaning

of a Eulerian variable. In the vortex line representation $\mathbf{\Omega}$ has the form [20]

$$\mathbf{\Omega}(\mathbf{r}, t) = \int \Omega_0(\nu) d^2\nu \int \delta(\mathbf{r} - \mathbf{R}(s, \nu, t)) \frac{\partial \mathbf{R}}{\partial s} ds. \quad (4)$$

Here each vortex line ν corresponds to the closed curve $\mathbf{r} = \mathbf{R}(s, \nu, t)$ so that $\mathbf{R}_s = \partial\mathbf{R}/\partial s$ is its tangential vector. The quantity $\Omega_0(\nu)$, being a Lagrangian invariant, characterizes the power of each vortex line. Incidentally, without limiting the generality, this function can be assumed to be equal to unity which is achieved by suitably redefining the markers ν and replacing the orientation of the lines with the opposite in those regions of the manifold $\mathcal{N}$ where $\Omega_0(\nu) < 0$. Thus, in the following analysis we shall omit the factor $\Omega_0(\nu)$ in the corresponding formulas. For an arbitrary vortex-line topology, (4) can be generalized to give

$$\mathbf{\Omega}(\mathbf{r}, t) = \int \delta(\mathbf{r} - \mathbf{R}(\mathbf{a}, t))(\mathbf{\Omega}_0(\mathbf{a}) \cdot \nabla_{\mathbf{a}}) \mathbf{R}(\mathbf{a}, t) d^3\mathbf{a}, \quad (5)$$

where $\mathbf{\Omega}_0(\mathbf{a})$ is a Cauchy Lagrangian invariant characterizing the frozenness. In this case, the condition $\mathrm{div}_{\mathbf{a}}\mathbf{\Omega}_0(\mathbf{a}) = 0$ guarantees that the field $\mathbf{\Omega}(\mathbf{r}, t)$ is transverse: $\mathrm{div}\mathbf{\Omega}(\mathbf{r}) = 0$. In expression (5) the vector $(\mathbf{\Omega}_0(\mathbf{a})\nabla_a)\mathbf{R}(\mathbf{a}, t) = \mathbf{b}$ is the tangential vector to the vortex line at the point

$$\mathbf{r} = \mathbf{R}(\mathbf{a}, t). \quad (6)$$

The parameter s in the representation (4), for example, may be the element of the vortex line arc of the initial field $\mathbf{\Omega}_0(\mathbf{a})$ and the components of the vector $\mathbf{a}$ transverse to this line may be taken as the parameter ν.

After integrating over the variables $\mathbf{a}$ in (5), the vector $\mathbf{\Omega}(\mathbf{r}, t)$ is expressed in terms of the Jacobian of the mapping (6) $J = \det\|\partial\mathbf{R}/\partial\mathbf{a}\|$ and the Cauchy invariant $\mathbf{\Omega}_0(\mathbf{a})$:

$$\mathbf{\Omega}(\mathbf{R}) = \frac{1}{J}(\mathbf{\Omega}_0(\mathbf{a}) \cdot \nabla_{\mathbf{a}})\mathbf{R}(\mathbf{a}). \quad (7)$$

It is important to stress that in this expression the Jacobian is not assumed to be unity: $J \neq 1$. Nevertheless, this does not contradict the condition of incompressibility of the velocity field.

The equations of motion for the vortex lines are obtained after substituting (5) into the frozen-in Eq. (2) [20, 21]:

$$[\{(\mathbf{\Omega}_0(\mathbf{a}) \cdot \nabla_a)\mathbf{R}(\mathbf{a}, t)\} \times \{\mathbf{R}_t(\mathbf{a}, t) - \mathbf{v}(\mathbf{R}(\mathbf{a}, t), t)\}] = 0. \quad (8)$$

This equation of motion describes the transverse dynamics of the vortex lines since any motion along the curve does not lead to its change. In particular, this explains why no constraints attributable to the condition of incompressibility are imposed on the Jacobian J. Note that on going over from a Eulerian description to a completely Lagrangian one, the Jacobian is $J \equiv 1$. On going over to a mixed Lagrange–Eulerian description in accordance with the equations of motion (6), the motion of Lagrangian particles along the vortex lines is eliminated in the mapping (6). For this reason the Jacobian is not necessarily unity. This point is principal and as we shall see subsequently, mainly explains why collapse is possible in hydrodynamics.

Equation (8), as was shown in [20, 21], can be written in the Hamiltonian form

$$[\{(\mathbf{\Omega}_0(\mathbf{a}) \cdot \nabla_a)\mathbf{R}(\mathbf{a}, t)\} \times \mathbf{R}_t(\mathbf{a})] = \frac{\delta\mathcal{H}\{\mathbf{\Omega}\{\mathbf{R}\}\}}{\delta\mathbf{R}(\mathbf{a})}\bigg|_{\mathbf{\Omega}_0}. \quad (9)$$

This equation describes the motion of vortex filaments for systems with an arbitrary Hamiltonian, a functional which depends only on $\mathbf{\Omega}(\mathbf{r}, t)$.

It is useful to bear in mind that the expressions for such important characteristics of the system such as its momentum $\mathbf{P} = \int \mathbf{p}\, d\mathbf{r}$ and the angular moment $\mathbf{M} = \int [\mathbf{r} \times \mathbf{p}] d\mathbf{r}$ being transformed by integration by parts to a form where the vorticity $\mathbf{\Omega} = \mathrm{curl}\,\mathbf{p}$ appears instead of $\mathbf{p}$, and then rewritten in terms of vortex lines, has the following form:

$$\mathbf{P} \sim \frac{1}{2}\int [\mathbf{r} \times \Omega] d\mathbf{r} = \int_{\mathcal{N}} d^2\nu \cdot \frac{1}{2}\int [\mathbf{R} \times \mathbf{R}_s] ds, \quad (10)$$

$$\mathbf{M} \sim \frac{1}{3}\int [\mathbf{r} \times [\mathbf{r} \times \Omega]] d\mathbf{r} = \int_{\mathcal{N}} d^2\nu \cdot \frac{1}{3}\int [\mathbf{R} \times [\mathbf{R} \times \mathbf{R}_s]] ds. \quad (11)$$

The tildes in these relationships imply that the equalities are obtained to within the integrals over an infinitely distant surface. We can see here that the momentum and angular moment of the system are made up of the momenta and angular moments of each vortex filament where the momentum of a closed line is the oriented surface area stretched onto a contour.

It is readily confirmed that when $\mathbf{R}$ shifts by the constant vector $\mathbf{R}_0$,

$$\mathbf{R} \longrightarrow \mathbf{R}' = \mathbf{R}_0 + \mathbf{R},$$

the momentum $\mathbf{P}$ remains unchanged while the angular momentum $\mathbf{M}$ undergoes the well-known transformation:

$$\mathbf{M} \longrightarrow \mathbf{M}' = [\mathbf{R}_0 \times \mathbf{P}] + \mathbf{M}. \quad (12)$$

3. INTEGRABLE HYDRODYNAMICS

In this and the next two sections we shall show how and why collapse can occur in three-dimensional integrable hydrodynamics. This model was introduced in

our previous paper [20]. Its Hamiltonian is expressed in terms of the modulus $\mathbf{\Omega}(\mathbf{r}, t)$:

$$\mathcal{H} = \int |\mathbf{\Omega}(\mathbf{r})| d\mathbf{r}, \tag{13}$$

and the equations of motion are the same as the frozen-in equation (2) with the generalized velocity

$$\mathbf{v} = \operatorname{curl} \boldsymbol{\tau},$$

where $\boldsymbol{\tau} = \mathbf{\Omega}/\Omega$ is the unit vector tangential to the vortex line. Assuming that all the vortex lines are closed, selecting their labeling such that $\Omega_0(\nu) = 1$, and substituting the representation (4) into (13), we can easily see that the Hamiltonian is decomposed into the sum of the Hamiltonians of the vortex lines:[1]

$$\mathcal{H}\{\mathbf{R}\} = \int d^2\nu \int \left|\frac{\partial \mathbf{R}}{\partial s}\right| ds. \tag{14}$$

Here the integral over s is the length of a vortex filament having the index ν. The equation of motion for the vector $\mathbf{R}$ in accordance with (9) will be local with respect to all variables, it does not contain interactions with other vortices:

$$[\mathbf{R}_s \times \mathbf{R}_t] = [\boldsymbol{\tau} \times [\boldsymbol{\tau} \times \boldsymbol{\tau}_s]]. \tag{15}$$

As a result of this factor, not only are the total energy, momentum, and angular moment of the system conserved but also the corresponding geometric invariants of each vortex line: its length

$$\mathcal{H}(\nu) = \int |\mathbf{R}_s(\nu)| ds,$$

momentum

$$\mathbf{P}(\nu) = \frac{1}{2}\int [\mathbf{R}(\nu) \times \mathbf{R}_s(\nu)] ds,$$

and angular momentum

$$\mathbf{M}(\nu) = \frac{1}{3}\int [\mathbf{R}(\nu) \times [\mathbf{R}(\nu) \times \mathbf{R}_s(\nu)]] ds.$$

Equation (15) is invariant with respect to the substitutions $s \longrightarrow \tilde{s}(s, t)$. Thus, it can be solved for $\mathbf{R}_t$ to within the shift along the vortex filament, a transformation which does not change the vorticity $\mathbf{\Omega}$. This implies that in order to find $\mathbf{\Omega}$ it is sufficient to find any single solution of the following equation derived from (15)

$$|\mathbf{R}_s| \cdot \mathbf{R}_t = [\boldsymbol{\tau} \times \boldsymbol{\tau}_s] + \beta \mathbf{R}_s \tag{16}$$

for a certain value of β. The equation thus obtained for $\boldsymbol{\tau}$ as a function of the filament length l ($dl = |\mathbf{R}_s| ds$) and time t (for $\beta = 0$) reduces [20] to an integrable one-dimensional Landau–Lifshitz equation for a Heisenberg ferromagnet:

$$\frac{\partial \boldsymbol{\tau}}{\partial t} = \left[\boldsymbol{\tau} \times \frac{\partial^2 \boldsymbol{\tau}}{\partial l^2}\right], \tag{17}$$

where $\boldsymbol{\tau}$ are functions of the filament length l and time t. This equation is then gauge-equivalent to the nonlinear Schrödinger equation [23]

$$i\psi_t + \psi_{ll} + 2|\psi|^2\psi = 0 \tag{18}$$

and may be reduced to it using the Hasimoto transformation [24]:

$$\psi(l, t) = \kappa(l, t) \exp\left[i\int^l \chi(\tilde{l}, t) d\tilde{l}\right],$$

where $\kappa(l, t)$ is the curvature and $\chi(l, t)$ is the line torsion.

This system with the Hamiltonian (13) is directly related to hydrodynamics. As we know [25, 26], the local approximation for a thin vortex filament, assuming that the filament thickness is small relative to the characteristic longitudinal scale gives the Hamiltonian (4) for a single filament. This approximation essentially involves replacing the logarithmic interaction with a constant. When the thicknesses of the vortex fragments are small compared with the spacing between them, the Hamiltonian of the vortex filaments in the same approximation will be the sum of the Hamiltonians which gives the Hamiltonian (13) in the "continuous" limit.

Thus, we have a model of three-dimensional exactly integrable hydrodynamics, the hydrodynamics of free vortices. Each vortex is then a nonlinear object having its own internal dynamics. As we shall see below, singularities may appear in this model. This is caused by the intersection of vortex lines, a phenomenon similar to wave breaking in gasdynamics.

3.1. Steady-State Vortices

We shall now consider the simplest solutions of equation (15), the steady-state propagation of a closed vortex line: $\mathbf{R}_t = \mathbf{V} \equiv \text{const}$. In this case, the velocity $\mathbf{V}$ is determined from the solution of the equation

$$[\mathbf{R}_s \times \mathbf{V}] = [\boldsymbol{\tau} \times [\boldsymbol{\tau} \times \boldsymbol{\tau}_s]]. \tag{19}$$

It is easy to verify that this equation follows from the variational principle

$$\delta(\mathcal{H}(\nu) - \mathbf{V} \cdot \mathbf{P}(\nu)) = 0, \tag{20}$$

i.e., the solution (19) is a steady-state Hamiltonian point with fixed momentum $\mathbf{P}(\nu)$. Equation (19) is easily integrated. It is easy to see that (19) can be rewritten in terms of the binormal $\mathbf{b}$ and the curvature of the curve κ in the form

[1] We note that this property is common for all systems for which the Hamiltonian has the form $\mathcal{H} = \int F(\tau, (\tau\nabla)\tau, (\tau\nabla)^2\tau, \ldots)|\mathbf{\Omega}|d\mathbf{r}$. To illustrate the idea of vortex line collapse we selected the simplest example (13) from this class which also has a specific physical meaning.

$$[\boldsymbol{\tau}\times\mathbf{V}] = \kappa[\boldsymbol{\tau}\times\mathbf{b}], \tag{21}$$

whence we have

$$\mathbf{V} = \kappa\mathbf{b}. \tag{22}$$

The fact that the velocity $\mathbf{V}$ is constant in this expression implies constant curvature κ, i.e., the vortex line is a ring of radius $r = 1/\kappa$, and

$$V = 1/r. \tag{23}$$

In this case, the direction of motion of the ring is perpendicular to its plane. It is interesting to note that the exact answer for the velocity of a thin vortex ring ([27]) having the thickness $d \ll r$ agrees with (22) to within logarithmic accuracy ($\sim\log(r/d)$) which, as has been noted, specifically distinguishes this model from the Euler equation.

It is also interesting that this solution in the form of a ring is stable, and is Lyapunov stable. The momentum $\mathbf{P}$ of the vortex ring is an oriented area stretched to the vortex contour:

$$\mathbf{P} = S\mathbf{n},$$

where S is the area and $\mathbf{n}$ is its normal. Since the Hamiltonian of the filament is the same as its length, the maximum momentum or, which amounts to the same thing, the maximum area for a contour of fixed length is clearly achieved for a closed contour in the form of a circle. This is evidence of the Lyapunov stability of the solution (22) in the form of a vortex ring.

4. COLLAPSE

The solution (22), (23) can be used to write the simplest mappings $\mathbf{R} = \mathbf{R}(\nu, s, t)$.

Let us assume that all the closed vortex lines have the same orientation, for example, along the z-axis, and have the form of circles. Since, as will become clear subsequently, collapse in this model is a purely local phenomenon, in order to construct the mapping we can confine our analysis to some vortex tube which may be represented as a torus. Let us assume that the annular vortex lines are distributed continuously inside the tube. Each vortex line inside the tube will be numbered by the two-dimensional parameter ν which will cover all points in the tube cross section at $t = 0$. As the parameter s, the third independent variable, we take the arc element s: $ds = rd\phi$, where ϕ is the polar angle about the z-axis. The desired mapping can then be written using (22) in the form

$$\begin{aligned}\mathbf{R} &= \mathbf{R}_0(\nu) + r(\nu)\cos\phi\,\mathbf{e}_x \\ &+ r(\nu)\sin\phi\,\mathbf{e}_y + V(\nu)t\,\mathbf{e}_z.\end{aligned} \tag{24}$$

In this formula $\mathbf{e}_x$, $\mathbf{e}_y$, and $\mathbf{e}_z$ are the unit vectors of the corresponding axes.

For this mapping it is easy to check that the transformation Jacobian varies linearly with time:

$$J = \frac{\partial(X, Y, Z)}{\partial(\nu_1, \nu_2, s))} = J_0(\nu, s) + A(\nu, s)t. \tag{25}$$

Here $A(\nu, s)$ is a coefficient which depends linearly on the velocity derivatives with respect to ν, and J_0 is the initial value of the Jacobian.

The time dependence of J (25) implies that for each fixed pair ν and s there always exists a time $t = \tilde{t}(\nu, s)$ ($t > 0$ or $t < 0$) when the Jacobian goes to zero:

$$J(\nu, s, t) = 0.$$

We take t_0 to denote the minimum value of the function $t = \tilde{t}(\nu, s)$ for $t > 0$. Let us assume that this minimum is achieved at a certain point $\mathbf{a} = \mathbf{a}_0$ [here we denoted the point (ν_1, ν_2, s) as $\mathbf{a}$]. Quite clearly, for $t = t_0$ we have

$$\left.\frac{\partial\tilde{t}}{\partial\mathbf{a}}\right|_{a = a_0} = 0$$

or

$$\nabla_a J(\mathbf{a}, t)\big|_{a = a_0} = 0, \tag{26}$$

since

$$\nabla_a J(\mathbf{a}, t)\big|_{a = a_0} + \frac{\partial J(\mathbf{a}, t)}{\partial t}\left.\frac{\partial\tilde{t}}{\partial\mathbf{a}}\right|_{a = a_0} = 0.$$

It is also clear that for $t = t_0$ the tensor of the second derivatives of J with respect to $\mathbf{a}$

$$2\gamma_{ij} = \frac{\partial^2 J}{\partial a_i \partial a_j},$$

will be positive definite at the point $\mathbf{a} = \mathbf{a}_0$. From this we can clearly see how the Jacobian behaves in the small vicinity of $\mathbf{a} = \mathbf{a}_0$. The expansion of the Jacobian near this point (in the general situation) for $t \longrightarrow t_0$ has the form

$$J(\mathbf{a}, t) = \alpha(t_0 - t) + \gamma_{ij}\Delta a_i \Delta a_j + \ldots, \tag{27}$$

where

$$\alpha = -\left.\frac{\partial J(\mathbf{a}, t)}{\partial t}\right|_{t = t_0,\, a = a_0} > 0, \quad \Delta\mathbf{a} = \mathbf{a} - \mathbf{a}_0.$$

These are the principal terms of the Jacobian expansion.[2]

Geometrically this expansion corresponds to a fairly simple pattern. The surface $J = J(\mathbf{a}, t)$ is deformed with time so that its minimum reaches the plane $J = 0$ at $t = t_0$ at the point $\mathbf{a} = \mathbf{a}_0$, where contact takes place with the surface $J = J(\mathbf{a}, t)$. Quite clearly for smooth mappings

[2] For example, the term containing the mixed derivative with respect to time and the spatial coordinate is generally a small correction to the first term (27).

in the general situation the surfaces $J = J(\mathbf{a}, t)$ and $J = 0$ contact at an isolated point. In degenerate situations contact can occur at several points simultaneously or even on an entire curve. The degenerate case also includes the situation when two eigenvalues of the Jacobi matrix $\hat{J}$ simultaneously go to zero at the point of collapse (we shall analyze this example in the next section). The following terms of the expansion need to be retained in the appropriate directions in (27). We reiterate that all these terms do not apply to the general situation.

In accordance with (7) vanishing of the Jacobian at the point of contact implies the appearance of a singularity for the vorticity at $t = t_0$:

$$(\mathbf{\Omega}(\mathbf{r}, t)) = \frac{\Omega_0(\nu)\mathbf{R}_s}{\alpha(t_0 - t) + \gamma_{ij}\Delta a_i \Delta a_j}. \quad (28)$$

Importantly the numerator in this fraction, a vector tangential to the vortex line, does not vanish due to its geometric meaning at the point of contact (this is again the nondegenerate case). Thus, at the point of contact the vorticity goes to infinity as $(t_0 - t)^{-1}$ and the size of the collapsing distribution in the coordinates **a** narrows as $\sqrt{t_0 - t}$.

This type of collapse occurs as a result of vortex line flipping when one vortex overtakes another. In the three-dimensional nondegenerate case this always occurs first at an isolated point.

As will become clear from the following, for this type of collapse the dependence (28) for $\mathbf{\Omega}(\mathbf{r}, t)$ obtained for a particular distribution is in fact a common answer which is suitable not only for integrable hydrodynamics but can also be applied to other systems of the type (2) which allow a quasi-inertial collapse regime. What criteria must be satisfied by the Hamiltonian of a specific system so that we can confirm that quasi-inertial collapse is achieved at least under some initial conditions or such collapse is completely impossible in this system? The answer to this question is as yet unknown and offers an extensive field of activity for future investigations.

4.1. Nonsteady-State Vortices

We shall now analyze integrable hydrodynamics in more general case when the closed vortex lines are not circles. In this case, each vortex contour must be set in correspondence with a specific vortex ring. The obvious procedure is to introduce the (average) direction **n** and also the average area of the contour $S = \pi r_0^2$ using an expression for the vortex line momentum:

$$\mathbf{P} = \mathbf{n}S.$$

The position $\mathbf{R}_0$ of the ring center then varies linearly with time. The corresponding average vortex velocity $\mathbf{V}_0$ of the closed vortex line must then be directed along the momentum in order to satisfy the angular momentum conservation law. The average velocity is generally a function of the fundamental integrals of motion of the line which do not depend on the choice of origin. These are the Hamiltonian (i.e., the length L), the momentum, and the projection of the angular momentum on the direction of momentum. A λ fold increase in the size of the contour should then lead to a λ fold reduction in velocity:

$$\mathbf{V}_0 = \frac{8}{\pi^2}\frac{\mathbf{P}}{L^3}U\left(\frac{16\pi^2 P^2}{L^4}, \frac{\mathbf{M}\cdot\mathbf{P}}{L^5}\right). \quad (29)$$

The exact form of the function $U(\xi, \eta)$, where the first argument defines the measure of "crumpledness" of the vortex line and varies in the range $0 \le \xi \le 1$ and the second defines the measure of mirror asymmetry, is as yet unknown to us. In a more or less reasonable approximation which does not lead to excessive errors, we can assume that $U(\xi, \eta) \sim U(1, 0) = 1$.

After introducing the average characteristics, we can express $\mathbf{R}(\nu, s, t)$ as the sum of $\tilde{\mathbf{R}}(\nu, s, t)$ and $\delta\mathbf{r}(\nu, s, t)$:

$$\mathbf{R}(\nu, s, t) = \tilde{\mathbf{R}}(\nu, s, t) + \delta\mathbf{r}(\nu, s, t), \quad (30)$$

where the average value of $\mathbf{R} = \mathbf{R}(\nu, s, t)$ is given by

$$\tilde{\mathbf{R}}(\nu, s, t) = \mathbf{R}_0 + r_0\cos\phi'\mathbf{e}'_x + r_0\sin\phi'\mathbf{e}'_y,$$
$$\dot{\mathbf{R}}_0 = V_0\mathbf{n}, \quad (31)$$

and ϕ' is the polar angle in the plane perpendicular to the local z'-axis directed along the unit vector **n**, the relationship between r_0 and V_0 is given by formula (29). The vector function $\delta\mathbf{r}(\nu, s, t)$ describes the deviations (generally, fairly large) from the average $\tilde{\mathbf{R}}(\nu, s, t)$.

The separation of the average and oscillatory motion for each vortex contour introduced by relations (29)–(31) shows that the mapping $\mathbf{R} = \mathbf{R}(\nu, s, t)$ for each fixed value of $\mathbf{a} = (\nu, s)$ is a linearly increasing function of time with nonlinear oscillations which are described using the Landau–Lifshitz equation (17) or its gauge-equivalent analog (18). The linear dependence reflects the fact that this model is a free vortex model. Thus, collapse occurs as a result of the "incursion" of one vortex onto another where as a result of the continuity of $\mathbf{\Omega}$ between the two "outer" vortices the vorticity goes to infinity. A similar situation is obtained for the model of the formation of large-scale structures in the universe studied in [22]. This model is based on the assumption of an initially dust-like mass distribution when its behavior may be described using hydrodynamic equations for the density ρ and velocity ν at zero pressure:

$$\rho_t + \operatorname{div}\rho\mathbf{v} = 0, \quad (32)$$

$$\frac{d\mathbf{v}}{dt} = \mathbf{v}_t + (\mathbf{v}\cdot\nabla)\mathbf{v} = 0. \quad (33)$$

Integrating this system in terms of Lagrangian variables shows that all the particles are free, moving at constant velocity:

$$\mathbf{r} = \mathbf{a} + \mathbf{v}(\mathbf{a})t, \tag{34}$$

and the density ρ is expressed in terms of the initial value ρ_0 and the Jacobian of the mapping (34):

$$\rho(\mathbf{r}, t) = \frac{\rho_0(\mathbf{a})}{J}. \tag{35}$$

The appearance of large-scale (fairly high-density) structures in this model is attributable to breaking, as a result of density singularities the Jacobian J of the mapping (34) vanishes. In general, these structures have a pancake shape and may be considered as protogalaxies. A formula similar to (35), as we have seen, is obtained for the vorticity $\mathbf{\Omega}$ [see (7)]. However, there is one difference from (35) associated with the vector nature of the field $\mathbf{\Omega}$, its transverseness. This is perhaps the only difference although it is an important one.

In this problem we are interested in the structure of the singularity at $t \longrightarrow t_0$ but $t < t_0$, i.e., in a certain sense at the initial stage of collapse but not at its advanced stage which is undoubtedly of value for astrophysical applications but remains far from clear for incompressible hydrodynamics when allowance must be made for viscosity on small scales.

4.2. Singularity Structure

We shall analyze in greater detail the structure of the collapsing region in the general situation. First it becomes clear from the reasoning put forward above that the distribution for the vorticity near the singularity at $t \longrightarrow t_0$ will be defined by the previous expression (28). Second, the singularity structure will be mainly determined by the Jacobian, the denominator in (28). The numerator $(\mathbf{\Omega}_0(\mathbf{a}) \cdot \nabla_\mathbf{a})\mathbf{R}$, a vector tangential to the vortex line, should be taken at the point $\mathbf{a} = \mathbf{a}_0$ and $t = t_0$ and assumed to be constant.

In accordance with (28), the Jacobian contains a positive-definite symmetric matrix γ_{ij}. This matrix for $t < t_0$ is assumed to be nondegenerate: all its eigenvalues are positive and it can be reduced to diagonal form. It then follows directly that compression in all three principal directions in $\mathbf{a}$ space will obey the same law $l_a \propto \sqrt{t_0 - t}$. Thus, near the singular point $\mathbf{\Omega}$ will have the self-similar asymptotic form:

$$\mathbf{\Omega}(\mathbf{r}, t) = \frac{\Omega_0(\nu)\mathbf{R}_s}{(t_0 - t)(\alpha + \gamma_{ij}\eta_i\eta_j)}, \tag{36}$$

where $\eta = \Delta\mathbf{a}/\sqrt{t_0 - t}$ is a self-similar variable in $\mathbf{a}$ space. However (36) does not imply that compression will be the same in $\mathbf{r}$ space.

The fact that the Jacobian vanishes at the point of collapse implies that one of the eigenvalues of the Jacobi matrix vanishes (in the nondegenerate situation). As we can see, near the point of collapse this eigenvalue (we denote this by λ_1) is the same as the Jacobian (27) apart from a constant factor (the product of two other eigenvalues $\lambda_2\lambda_3$).

We express the Jacobi matrix $\hat{J}$ as an expansion in terms of the eigenvectors of the direct $\hat{J}|\psi^{(n)}\rangle = \lambda_n|\psi^{(n)}\rangle$ and conjugate $\langle\tilde{\psi}^{(n)}|\hat{J} = \lambda_n\langle\tilde{\psi}^{(n)}|$ spectral problems:

$$J_{ik} \equiv \frac{\partial x_k}{\partial a_i} = \sum_{n=1}^{3} \lambda_n \psi_i^{(n)} \tilde{\psi}_k^{(n)}. \tag{37}$$

Here the two sets of the eigenvectors of the direct and conjugate spectral problems are mutually orthogonal:

$$\langle\tilde{\psi}^{(n)}|\psi^{(m)}\rangle = \delta_{nm}.$$

Near the point $\mathbf{a} = \mathbf{a}_0$ the two eigenvalues $\lambda_{2,3}$ can be considered to be constant and

$$\lambda_1 \equiv \frac{J}{\lambda_2\lambda_3} = \frac{\alpha(t_0 - t) + \gamma_{ij}a_i a_j}{\lambda_2\lambda_3}.$$

Here, for simplicity we positioned the origin at the point $\mathbf{a} = \mathbf{a}_0$. The eigenvectors can be considered to be constant in this vicinity.

We then expand the vectors $\mathbf{x}$ and ∇_a in (37) in the corresponding bases, denoting the corresponding projections by X_n and A_n:

$$X_n = \langle\mathbf{x}|\psi^{(n)}\rangle, \quad \frac{\partial}{\partial A_n} = \langle\tilde{\psi}^{(n)}|\nabla_a\rangle.$$

The vector $\mathbf{a}$ is then written in terms of the projection A_n as follows:

$$a_\alpha = \sum_n \psi_\alpha^{(n)}|\tilde{\psi}^{(n)}|^2 A_n.$$

As a result, (37) is rewritten in the form

$$\frac{\partial X_1}{\partial A_1} = \tau + \Gamma_{mn}A_m A_n, \tag{38}$$

$$\frac{\partial X_2}{\partial A_2} = \lambda_2, \quad \frac{\partial X_3}{\partial A_3} = \lambda_3. \tag{39}$$

Here we have

$$\Gamma_{mn} = \gamma_{\alpha\beta}\psi_\alpha^{(n)}\psi_\beta^{(m)}|\tilde{\psi}^{(\mathbf{n})}|^2|\tilde{\psi}^{(\mathbf{m})}|^2(\lambda_2\lambda_3)^{-1},$$

$$\tau = \alpha(t_0 - t)(\lambda_2\lambda_3)^{-1}.$$

It then follows directly that in the second (X_2) and third (X_3) "hard" directions compression takes place exactly as in the auxiliary $\mathbf{a}$ space, i.e., $\propto\sqrt{\tau}$ and in the "soft" X_1 direction it takes place as $\tau^{3/2}$. Thus, in the new self-similar variables $\xi_1 = X_1/\tau^{3/2}$, $\xi_2 = X_2/\tau^{1/2}$, $\xi_3 = X_3/\tau^{1/2}$

integrating the system (39) gives a linear dependence on η for ξ_2, ξ_3 and a cubic dependence for ξ_1:

$$\xi_1 = (1 + \Gamma_{ij}\eta_i\eta_j)\eta_1 + \frac{1}{2}\Gamma_{1i}\eta_i\eta_1^2 + \frac{1}{3}\Gamma_{11}\eta_1^3, \quad i, j = 2, 3, \tag{40}$$

$$\xi_2 = \lambda_2\eta_2, \quad \xi_3 = \lambda_3\eta_3. \tag{41}$$

Relationships (40) and (41) together with (36) determine the implicit dependence of $\mathbf{\Omega}$ on the time and coordinates. The existence of two different self-similarities shows that as t approaches t_0, the spatial distribution of $\mathbf{\Omega}$ becomes severely flattened in the first direction, becoming pancake-shaped. The direction of the field $\mathbf{\Omega}$ is then determined from the compressibility condition: $\operatorname{div}\mathbf{\Omega} = 0$. In the principal order this imposes a constraint on the direction of the tangential vector $\mathbf{R}_s$ at the point of collapse:

$$(\mathbf{R}_s(\mathbf{a}_0, t) \cdot \nabla(1/J)) + \ldots = 0. \tag{42}$$

Quite clearly, the gradient from J for $t \longrightarrow t_0$ in the principal approximation is determined by the soft direction:

$$\nabla J \approx \tau^{-3/2}\frac{\partial J}{\partial \xi_1}\mathbf{e}_1.$$

The contribution from other directions is small in terms of the parameter τ.

It then follows as a result of (42) that the vector $\mathbf{R}_s$ and thus the field $\mathbf{\Omega}$ are orthogonal to the soft direction, i.e., lie in the "pancake" plane which is consistent with the transverseness of the motion of these vortex lines [see (8)].

5. EXAMPLE OF COLLAPSE IN THE DEGENERATE CASE

In the previous section we have considered collapse in the nondegenerate situation when only one eigenvalue of the Jacobi matrix vanishes at the point of contact. We shall now consider an example of collapse when two eigenvalues of the Jacobi matrix vanish simultaneously at the point of collapse. Here we are dealing with an initial vorticity distribution with a nontrivial vortex line topology with a degree of linkage of one. This special distribution is constructed using a Hopf mapping. Several equivalent methods are available to search for the corresponding field $\mathbf{\Omega}$. We shall use a method indicated in [28].

Following this study, we express the field $\mathbf{\Omega}$ in terms of the $\mathbf{n}$ field ($\mathbf{n}^2 = 1$]:

$$\Omega_\alpha(\mathbf{r}) = \frac{1}{32}\epsilon_{\alpha\beta\gamma}(\mathbf{n} \cdot [\partial_\alpha \times \partial_\gamma \mathbf{n}]), \tag{43}$$

where we shall assume that the $\mathbf{n}$ field is smooth and has the value $\mathbf{n} = \mathbf{e}$ at infinity $r \longrightarrow \infty$.

It is easy to confirm that each point $\mathbf{n} = \mathbf{n}_0$ on the unit sphere $\mathcal{S}^2$ corresponds to a vortex line in accordance with (43). Expressing the unit vector $\mathbf{n}$ in terms of the spherical angles θ and φ, we can write $\mathbf{\Omega}$ in the form

$$\mathbf{\Omega} = \frac{1}{16}[\nabla\varphi \times \nabla\cos\theta], \tag{44}$$

such that the variables φ and $\cos\theta$ appear as Clebsch variables. Thus, in this case each vortex line coincides with the intersection of two surfaces, φ = constant and $\cos\theta$ = const, i.e., the inverse image of the point on the sphere $\mathcal{R}^3$ is a closed vortex line. The Hop index N of the mapping $\mathcal{R}^3 \longrightarrow \mathcal{S}^2$ is the (integer) number of linkages of the two vortex lines, i.e., the inverse images of two points on a unit sphere. The Hopf mapping (with the index $N = 1$) is constructed using the following formula:

$$(\mathbf{n} \cdot \boldsymbol{\sigma}) = U(\mathbf{e} \cdot \boldsymbol{\sigma})U^\dagger, \quad U = \frac{1 + i(\mathbf{a} \cdot \boldsymbol{\sigma})}{1 - i(\mathbf{a} \cdot \boldsymbol{\sigma})}, \tag{45}$$

where σ are the Pauli matrices.

Then expressing the vector $\mathbf{n}$ and substituting into (43), we obtain [see (29)]

$$\mathbf{\Omega}_0(\mathbf{a}) = \frac{\mathbf{e}(1 - a^2) + 2\mathbf{a}(\mathbf{e} \cdot \mathbf{a}) + 2[\mathbf{e} \times \mathbf{a}]}{(1 + a^2)^3}. \tag{46}$$

The field lines of this field are circles, as was shown in [29], where any two lines are singly linked. For these reasons the formation of a singularity is inevitable. This field has no singular points and its absolute value only depends on $|\mathbf{a}|$. The unit vector $\boldsymbol{\tau}(\mathbf{a})$ is defined everywhere which is extremely convenient from the point of view of using our model. For $\boldsymbol{\tau}$ we have the following dependence:

$$\boldsymbol{\tau}(\mathbf{a}) = \frac{\mathbf{e}(1 - a^2) + 2\mathbf{a}(\mathbf{e} \cdot \mathbf{a}) + 2[\mathbf{e} \times \mathbf{a}]}{(1 + a^2)}. \tag{47}$$

The velocity of each ring is related to the binormal $\mathbf{b}(\nu)$ and the radius $r(\nu)$ by

$$\mathbf{V}(\nu) = \mathbf{b}(\nu)/r(\nu).$$

The radii of the rings $r(\nu)$ and their orientations $\mathbf{b}(\nu)$ are integrals of motion. Only the positions of the ring centers vary and this motion takes place at a constant velocity for each ring.

In this problem, instead of the variables ν and s, it is convenient to use the variables $\mathbf{a}$ in which the mapping $\mathbf{R}(\mathbf{a}, t)$ can be expressed in the form

$$\mathbf{R}(\mathbf{a}, t) = \mathbf{a} + t\mathbf{V}(\mathbf{a}), \tag{48}$$

where the velocity $\mathbf{V}(\mathbf{a})$ is expressed in terms of the tangential vector $\boldsymbol{\tau}$ by

$$\mathbf{V}(\mathbf{a}) = [\boldsymbol{\tau}(\mathbf{a}) \times (\boldsymbol{\tau}(\mathbf{a}) \cdot \nabla_{\mathbf{a}})\boldsymbol{\tau}(\mathbf{a})]. \quad (49)$$

We can then calculate the Jacobian

$$J(\mathbf{a}, t) = \det\left\| \hat{I} + t\left(\frac{\partial \mathbf{V}(\mathbf{a})}{\partial \mathbf{a}}\right)\right\|. \quad (50)$$

We note that the velocity of each ring is constant along the field line. Hence, the determinant of the matrix $\partial\mathbf{V}(\mathbf{a})/\partial\mathbf{a}$ determining the coefficient at t^3 in the expression for the Jacobian J is zero. As a result, J has a quadratic dependence in terms of the time t:

$$J = 1 + tc_1(\mathbf{a}) + t^2 c_2(\mathbf{a}).$$

The formation of a singularity corresponds to vanishing of this expression. The coefficients c_1 and c_2 determine the time of collapse t_0 which is the minimum positive root of the equation

$$\min_{\mathbf{a}} J(\mathbf{a}, t_0) \equiv J_{\min}(t_0) = 0.$$

Calculations using formulas (47), (49), and (50) yield the expressions

$$\mathbf{V}(\boldsymbol{a}) = \frac{-2}{(1 + a^2)^2} \times ([\mathbf{e} \times \mathbf{a}](1 - a^2) + 2\mathbf{a}(\mathbf{e} \cdot \mathbf{a}) - 2\mathbf{e}a^2), \quad (51)$$

$$J(\mathbf{a}, t) = \frac{(1 + a^2)^3 - 8ta_3(1 + a^2) + 4t^2(1 + a_3^2 - a_1^2 - a_2^2)}{(1 + a^2)^3}, \quad (52)$$

where a_3 is the projection of the vector $\mathbf{a}$ on the $\mathbf{e}$ axis.

Analyzing this expression, we can show that at $t < 1$ the minimum of the Jacobian should be sought on the symmetry axis, i.e., for $a_1 = 0$, and $a_2 = 0$. In this case, we have

$$J_{\text{axis}} = \frac{(1 - a_3^2)^2 + 4(t - a_3)^2}{(1 + a_3^2)^2}, \quad (53)$$

from which it can be seen that a singularity occurs for $t_0 = 1$, $a_3 = 1$, and the Jacobian vanishes along the quadratic asymptotic form. Thus, in this example we have $|\Omega|_{\max} \propto (t_0 - t)^{-2}$. It is also easy to confirm that the Jacobi matrix at the point of contact has two eigenvalues whose eigenvectors lie in the plane perpendicular to the $\mathbf{e}$ axis. Near the collapse the field $\boldsymbol{\Omega}$ is directed along the $\mathbf{e}$ vector. Compression in this direction is linear in time $l_3 \propto (t_0 - t)$ whereas in the perpendicular plane it is faster, obeying the law $l_{1,2} \propto (t_0 - t)^{3/2}$. As a result, the structure of the singularity is highly elongated along the axis of anisotropy.

6. CONCLUDING REMARKS

Analyzing a hydrodynamic model with the Hamiltonian $\mathcal{H} = \int |\boldsymbol{\Omega}|\, d\mathbf{r}$ we have reached the conclusion that each vortex line in this system moves independently of the others. This property makes it possible for a singularity to form in the dependence of the generalized vorticity field $\boldsymbol{\Omega}(\mathbf{r}, t)$ over a finite time from smooth initial data. A typical singularity of this type resembles an infinite thickening of the vortex lines near a certain point. Collapse therefore has a purely inertial origin in integrable hydrodynamics. Assuming that this type of collapse can occur in Euler hydrodynamics, the asymptotic form of the vorticity near the singularity, the vortex breaking point, will be defined in the nondegenerate case by formulas (28) or (36), which is clear from general concepts, i.e., the curl of the velocity goes to infinity as $(t_0 - t)^{-1}$. This dependence at the singular point is observed in almost all numerical experiments, including those cited in [6,10–14]. However, not all experiments refer to the numerical integration of the Euler equation for continuous distributions. The first numerical experiments [6] (for the development of this trend, see [12]) referred to the collapse of two antiparallel vortex filaments which, as was shown in [15] are linearly unstable with respect to transverse perturbations. A theory of collapse for thin vortex filaments as the nonlinear stage of Crow instability was developed by Zakharov [7, 8] (see also [9]). Its conclusions show good agreement with the numerical experiments down to distances comparable with the size of the filament core. In this case, the distance between the vortex filaments decreases as $\sqrt{t_0 - t}$. At shorter distances the cores of the vortex filaments lose their circular shape. They become flat and the filament attraction process is slowed [30, 31]. This tendency is observed in the Kerr numerical experiments [10] which in our view are the most advanced for reconnection problems where, unlike [6], the collapse of two antiparallel but continuously distributed vortices is considered. In addition to a natural reduction in the average distance between the distributed vortices, these experiments first revealed the formation of two singularities at two symmetric points (obviously we can only talk of singularities quite arbitrarily, in the measure of numerical possibilities). As the time of collapse is approached, an explosive increase in the maximum vorticity was observed according to the law $(t_0 - t)^{-1}$. According to recent reports by Kerr [14], an analysis of numerical data gave two scales, one of which decreases as the square root: $l_1 \propto (t_0 - t)^{3/2}$ and the other decreasing linearly with time: $l_2 \propto t_0 - t$.

In [11] a successful attempt was made to observe collapse for an initial condition possessing low symmetry. The initial vorticity was concentrated near a cylinder and it was angularly modulated so that the simplest symmetries were absent. In this experiment,

in which the best space–time resolution has been achieved so far, an isolated collapsing region was observed with increasing vorticity at the center obeying the law $(t_0 - t)^{-1}$.

Consequently the result of all existing numerical experiments for ideal hydrodynamics are consistent with the concept of collapse as the process of formation of a caustic of the nondivergent field of the velocity curl. Complete agreement is obtained with the behavior of the vorticity maximum. For the spatial structure of the collapsing region the agreement is only qualitative, as is indicated by the results of [30, 31] where compression of the vortex core was observed at the initial stage of reconnection (neglecting viscosity) and also the Kerr results. All this suggests that the collapse scenario presented here as the thickening of the vortex lines near some point is quite plausible. We reiterate that if this scenario occurs, the behavior of the vorticity near the singular point is determined by Eqs. (28) or (36). The structure of this region should be strongly anisotropic: in one of the directions perpendicular to the vorticity the compression is faster ($\propto\tau^{3/2}$) than in the other two directions ($\propto\tau^{1/2}$). The distribution becomes close to two-dimensional, highly reminiscent of tangential discontinuity. The flow rate in this region is approximated to a high degree of accuracy by the linear dependence:

$$v_\perp \sim \Omega_{\max} X_1,$$

i.e., the flow is shear flow. According to the existing classification [32], this type of collapse is weak: the energy entering the singularity (allowing for the viscosity) tends to zero as the viscosity ν decreases. It is interesting to note that the rate of energy dissipation $\sim\int \Omega^2\, d\mathbf{r}$ from the collapsing region also tends to zero for the viscosity $\nu \longrightarrow 0$.

It should also be noted that unlike the free vortex model considered above, in the actual Eulerian hydrodynamics the vortex lines interact in pairs in accordance with the Hamiltonian

$$\mathcal{H}_{\text{Euler}} = \frac{1}{8\pi}\int\int \frac{\mathbf{R}_s(\nu, s)\cdot \mathbf{R}_\xi(\mu, \xi)}{|\mathbf{R}(\nu, s) - \mathbf{R}(\mu, \xi)|} d^2\nu\, ds\, d^2\mu\, d\xi. \quad (54)$$

From the point of view of studying the problem of collapse it is fundamentally important that the interaction function (the Green's function of the Laplace operator) has a singularity for equal arguments $\mathbf{R}(\nu, s) \longrightarrow \mathbf{R}(\mu, \xi)$. If such a singularity did not exist and the interaction function was completely regular, any initial vortex line distribution topologically equivalent to some smooth field $\mathbf{\Omega}_0(\mathbf{a})$, including a very singular one, would generate a fairly smooth velocity field $\mathbf{v}(\mathbf{r})$. Thus, an arbitrary initial singularity of the generalized vorticity field could not be "disappeared" in the following times and could only be transported by the liquid flow. Since the equations of motion of an ideal medium are reversible in time, it follows that a singularity also could not form from smooth initial data within a finite time. Thus, the existence of vortex line collapse and its possible types in systems with quadratic Hamiltonians with respect to $\mathbf{\Omega}$ depends on the asymptotic form of the interaction function $G(\mathbf{r}_1, \mathbf{r}_2)$ (more accurately on it derivatives), for $\mathbf{r}_1 \longrightarrow \mathbf{r}_2$. For example, for a better understanding of the problem of collapse in hydrodynamics, it is meaningful to study those systems for which the interaction functions have the asymptotic forms $G \propto |\mathbf{r}_1 - \mathbf{r}_2|^{-q}$ where the exponent q is not necessarily equal to unity.

How does the viscosity influence the structure of the collapsing region and especially the post-collapse stage? How does this type of collapse influence the structure of the turbulent spectra? This is by no means a complete list of the most important topics requiring investigation. It is also interesting to make a numerical check of our hypothesis on the inertial character of collapse directly for ideal hydrodynamics both in terms of Euler variables and in the vortex line representation.

ACKNOWLEDGMENTS

The authors thank V.E. Zakharov, R.Z. Sagdeev, V.V. Lebedev, G. E. Fal'kovich, A. Tsinober, I. Goldhirsch, N. Zabusky, J. Herring, R. Kerr, and R. Pelz for useful discussions of many topics touched on in this study. This work was supported by the Russian Foundation for Basic Research (project no. 00-01-00929) and the Program for Support of Leading Russian Scientific Schools (grant no. 00-15-96007). One of us (V.P.R) is also grateful for financial support from the Landau Postdoctoral Scholarship Foundation (KFA, Forschungszentrum, Julich, Germany).

REFERENCES

1. A. N. Kolmogorov, Dokl. Akad. Nauk SSSR **30**, 9 (1941).
2. O. M. Phillips, J. Fluid Mech. **4**, 426 (1958); in *Windy Waves* (Inostrannaya Literatura, Moscow, 1962).
3. B. B. Kadomtsev and V. I. Petviashvili, Dokl. Akad. Nauk SSSR **208**, 794 (1973) [Sov. Phys. Dokl. **18**, 115 (1973)].
4. P. G. Saffman, J. Fluid Mech. **106**, 49 (1981).
5. U. Frisch, *Turbulence. The Legacy of A. N. Kolmogorov* (Cambridge Univ. Press, Cambridge, 1995; FASIS, Moscow, 1998).
6. A. Pumir and E. D. Siggia, in *Topological Fluid Mechanics*, Ed. by H. K. Moffatt and A. Tsinober (Cambridge Univ. Press, Cambridge, 1990), p. 469; Phys. Fluids A **4**, 1472 (1992).
7. V. E. Zakharov, Usp. Fiz. Nauk **155**, 529 (1988) [Sov. Phys. Usp. **31**, 672 (1988)].
8. V. E. Zakharov, in *Lecture Notes in Physics: Nonlinear MHD Waves and Turbulence*, Ed. by T. Passot and P. L. Sulem (Springer-Verlag, Berlin, 1999), p. 369.
9. R. Klein, A. Maida, and K. Damodaran, J. Fluid Mech. **228**, 201 (1995).
10. R. M. Kerr, Phys. Fluids A **4**, 2845 (1993).

11. R. Grauer, C. Marliani, and K. Germaschewski, Phys. Rev. Lett. **80**, 4177 (1998).
12. R. B. Pelz, Phys. Rev. E **55**, 1617 (1997).
13. O. N. Boratav and R. B. Pelz, Phys. Fluids **6**, 2757 (1994).
14. R. M. Kerr, in *Trends in Mathematics* (Birkhauser, Basel, 1999), p. 41.
15. S. C. Crow, AIAA J. **8**, 2172 (1970).
16. Ya. G. Sinai, private communication (1999).
17. V. I. Arnold, *Catastrophe Theory* (Znanie, Moscow, 1981; Springer-Verlag, Berlin, 1986); *Mathematical Methods of Classical Mechanics* (Nauka, Moscow, 1984; SpringerVerlag, New York, 1989).
18. R. Salmon, Annu. Rev. Fluid Mech. **20**, 225 (1988).
19. V. E. Zakharov and E. A. Kuznetsov, Usp. Fiz. Nauk **137**, 1137 (1997) [Phys. Dokl. **40**, 1087 (1997)].
20. E. A. Kuznetsov and V. P. Ruban, Pis'ma Zh. Éksp. Teor. Fiz. **67**, 1015 (1998) [JETP Lett. **67**, 1076 (1998)].
21. E. A. Kuznetsov and V. P. Ruban, Phys. Rev. E **61**, 831 (2000).
22. V. I. Arnold, S. F. Shandarin, and Ya. B. Zeldovich, Geophys. Astrophys. Fluid Dyn. **20**, 111 (1982).
23. V. E. Zakharov and L. A. Takhtadzhyan, Teor. Mat. Fiz. **38**, 26 (1979).
24. R. Hasimoto, J. Fluid Mech. **51**, 477 (1972).
25. L. S. Da Rios, Rend. Circ. Mat. Palerno **22**, 117 (1906).
26. R. Betchov, J. Fluid Mech. **22**, 471 (1965).
27. H. Lamb, *Hydrodynamics* (Cambridge Univ. Press, Cambridge, 1932, 6th ed.).
28. E. A. Kuznetsov and A. V. Mikhailov, Phys. Lett. A **77**, 37 (1980).
29. A. M. Kamchatnov, Zh. Éksp. Teor. Fiz. **82**, 117 (1982) [Sov. Phys. JETP **55**, 69 (1982)].
30. M. V. Melander and F. Hussain, Phys. Fluids A **1**, 633 (1989).
31. M. J. Shelley, D. J. Meiron, and S. A. Orszag, J. Fluid Mech. **246**, 613 (1993).
32. V. E. Zakharov and E. A. Kuznetsov, Zh. Éksp. Teor. Fiz. **91**, 1310 (1986) [Sov. Phys. JETP **64**, 773 (1986)].

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 4, 2000, pp. 786–790.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 4, 2000, pp. 906–910.

SOLIDS
Structure

Role of Internal Pressure in a Ferroelastic Phase Transition

S. V. Goryaĭnov and N. N. Ovsyuk*
Institute of Mineralogy and Petrography, Siberian Division Russian Academy of Sciences, Novosibirsk, 630090 Russia
**e-mail: ovsyuk@uiggm.nsc.ru*
Received May 22, 2000

Abstract—It is shown that the driving force of a ferroelastic phase transition is internal pressure. The pressure gives rise to symmetric deformation, whose energy accumulates as pressure increases and, at the phase transition point, is partially transferred into the energy of antisymmetric deformations. © *2000 MAIK "Nauka/Interperiodica".*

Ferroelastics are attracting a great deal of attention as high-temperature superconductors. They are also of interest because they have extensive applications in acoustooptic and acoustoelectric devices. Nonetheless, a great deal remains unclear in the nature and mechanisms of ferroelastic phase transitions (FPT). The phenomenological description of FPT employs the spontaneous strain tensor and artifically chosen order parameters, one of which is considered to be critical [1]. If the order parameter and the spontaneous deformation transform under symmetry operations identically, then the ferroelastic is said to be proper. For a proper FPT one component of the spontaneous strain tensor can be taken as the critical parameter [2]. The order parameters and the strain tensor are essentially determined by the displacements of the atoms accompanying a change in external conditions, and it is strange that the force giving rise to these displacements and being the true reason for the transition has still not been studied. Consequently, such an approach cannot give the correct physical description of a FPT.

The physics of a proper FPT with lowering of symmetry down to triclinic, usually followed by amorphization, as shown for the example anorthite and quartz [3, 4], is of special interest. For such FPT, characterized by two macroscopic parameters which break symmetry, there are no analytic solutions in the literature.

In this connection, in the present paper we develop a model of a proper FPT in which internal pressure, which is the driving force of the transition, is introduced. As an example illustrating this model we consider the monoclinic-triclinic proper FPT in Sr-anorthite ($(Sr, Ca)Al_2Si_2O_8$), induced by cationic exchange of Ca and Sr. This ferroelastic was chosen because the experimental values of the elastic moduli [5] and the composition dependences of the unit cell parameters [6] are available for it.

The FPT is usually described by a linear strain tensor, which has the Lagrangian form

$$e_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_{0j}} + \frac{\partial u_j}{\partial x_{0i}}\right).$$

Since the internal spontaneous deformation increases at a phase transition, the nonlinear terms may have to be taken into account in the strain tensor. When the nonlinear terms are taken into account this tensor is said to be finite and has the form

$$E_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_{0j}} + \frac{\partial u_j}{\partial x_{0i}} + \sum_k \frac{\partial u_k}{\partial x_{0i}}\frac{\partial u_k}{\partial x_{0j}}\right),$$

where the vector $\mathbf{u} = \mathbf{x} - \mathbf{x}_0$ determines the displacement of the point whose coordinate is x_0 before the deformation and x after the deformation. The indices i, j, and k correspond to Cartesian coordinates, and each one runs through the values 1, 2, and 3. For convenience in comparing with experiment, we shall express the components of the strain tensor in terms of the lattice parameters. The corresponding expressions for the components of the linear tensor are presented in explicit form in [7]. We obtain the relations for the components of the finite tensor, which can be expressed in terms of the components e_{ij} in the Voigt notations as follows:

$$E_1 = e_1 + e_1^2/2 + e_6^2/2,$$

$$E_2 = e_2 + e_2^2/2,$$

$$E_3 = e_3 + e_3^2/2 + e_4^2/2 + e_5^2/2,$$

$$E_4 = e_4 + e_2e_4,$$

$$E_5 = e_5 + e_2e_5 + e_4e_6,$$

$$E_6 = e_6 + e_2e_6.$$

1063-7761/00/9104-0786$20.00

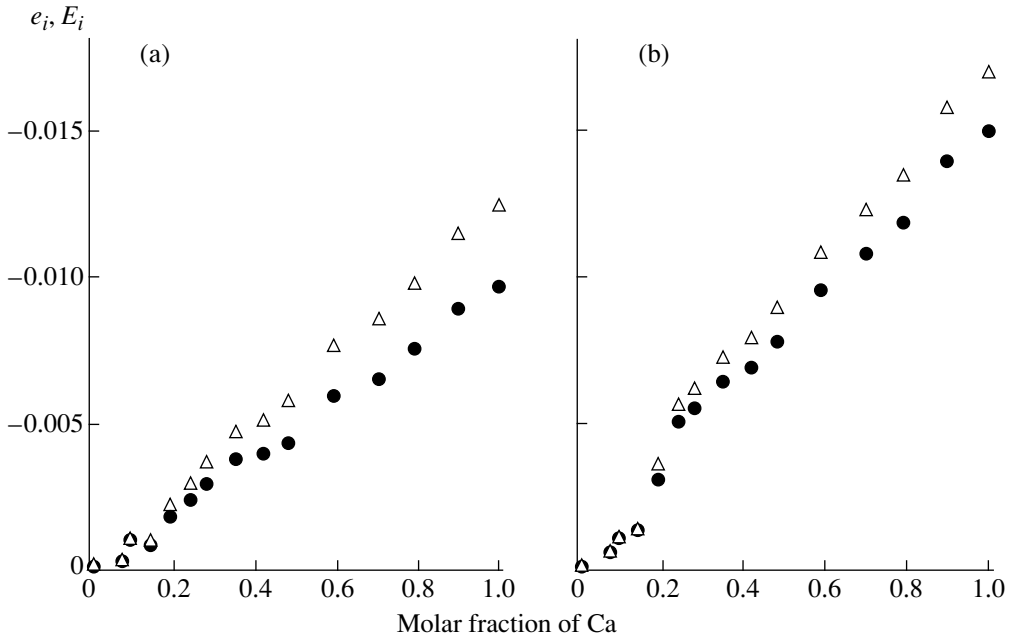

Fig. 1. Illustration of the differences between the dependences of the components of the linear e_i (triangles) and finite E_i (circles) strain tensors on the content of Ca. The curves were obtained using the unit cell parameters of Sr anorthite [6]: (a) $i = 3$; (b) $i = 5$.

Substituting into the formulas obtained the experimental lattice parameters for Sr anorthite [6] we found that there is a large difference between the corresponding components of the linear and finite strain tensors (Fig. 1). Therefore, to decrease the error the finite strain tensor must be used when calculating the FPT.

A systematic physical description of a FPT requires analysis of the displacements of the atoms and forces giving rise to these displacements. To a first approximation these forces can be expressed in terms of the isotropic internal pressure p_i. In order to close the system, the equations must also describe the internal pressure as a function of some thermodynamic parameters. To a first approximation it can be assumed that this pressure is a linear function of the cationic composition. Then the system of equations describing the FPT will have the form

$$\frac{\partial G(E_m, p_i)}{\partial E_m} = 0, \qquad p_i = A_1 N, \tag{1}$$

where $G(E_m, p_i)$ is the Gibbs potential, E_m are the components of the strain tensor, $m = 1, 2, \ldots, 6$, A_1 is a linear coefficient, and N is the molar fraction of the substituting cations.

The Gibbs potential of a crystal, to which the external pressure p_{ext} is applied, is expressed in the form

$$G = F + p_{ext}V,$$

where F is the free energy and $V \approx E_1 + E_2 + E_3$ is the relative change in the unit-cell volume. It would be logical to use the same expression for the Gibbs potential of the ferroelastic also, where the internal pressure p_i giving rise to spontaneous deformation is present:

$$G = F + p_i V. \tag{2}$$

The internal pressure can be produced for various reasons: change in the cationic composition Ca–Sr, change in the degree of ordering Al–Si, disruptions of the crystal structure on grain boundaries and on dislocations, and temperature. Making use of the term $p_i V$ in Eq. (2), we introduced the action of a thermodynamic parameter on the linear terms in the expansion of the Gibbs energy in terms of the strain tensor. The action of thermodynamic parameters on the quadratic and higher powers in the expansion of the Gibbs potential in the strain tensor is taken into account in the free energy. Thus, we assume that the proposed Gibbs potential can completely describe a proper FPT. To calculate this transition it is necessary to choose an expansion for the free energy. Since the component E_4 of the finite tensor is maximal and destroys the symmetry of the monoclinic crystal we chose, following [2], it as the critical order parameter. The expansion of the part of the free energy in terms of the critical parameter E_4, according to the Landau theory, has the form

$$F_4 = \frac{1}{2}A_4E_4^2 + \frac{1}{4}B_4E_4^4 + \frac{1}{6}C_4E_4^6, \tag{3}$$

where $A_4 = (p_{ic} - p_i)A_0$ the coefficient vanishes at the critical internal pressure p_{ic}, while all other coefficients remain constants. The component E_6 of the finite strain tensor also breaks symmetry, but since it is much

smaller than E_4 near the phase transition, we shall assume it to be noncritical, i.e., we write the expansion of the part of the free energy in terms of the parameter E_6 in the form

$$F_6 = \frac{1}{2}A_6E_6^2 + \frac{1}{4}B_6E_6^4, \quad (4)$$

where the coefficients A_6 and B_6 are constants which do not depend on the cationic composition Ca–Sr. To simplify the main behavior of the FPT in the analytic form, in order to describe the interaction of the parameters with the same symmetry, E_4 and E_6, we took only one bilinear term, dropping by the biquadratic terms,

$$F_{\text{int}} = A_{4-6}E_4E_6. \quad (5)$$

As the solution of Eqs. (1) show, the linear-quadratic interaction of symmetric components with antisymmetric components accompanying cationic exchange results only in a small additional contribution to the linear dependence of the coefficient A_4 in Eq. (3), and consequently, for simplicity, we drop it. The interaction of the symmetric components is expressed by the standard quadratic elastic strain:

$$F_{\text{sim}} = \frac{1}{2}\sum_{kl} A_{kl}E_kE_l, \quad (6)$$

where the indices k, l = 1, 2, 3, 5. Now the total free energy is

$$F = F_4 + F_6 + F_{\text{int}} + F_{\text{sim}}. \quad (7)$$

Solving the system of Eqs. (1), we find the dependences of the components of the spontaneous-strain tensor on the internal pressure with increasing Ca content. The four symmetric components are linear functions of the internal pressure:

$$\begin{aligned} E_1 &= k_1p_{\text{i}}, \quad E_2 = k_2p_{\text{i}}, \\ E_3 &= k_3p_{\text{i}}, \quad E_5 = k_5p_{\text{i}}. \end{aligned} \quad (8)$$

(here the coefficients are expressed in terms of the elastic moduli). The antisymmetric components E_4 and E_6 are zero in the monoclinic phase with $p_{\text{i}} < p_{\text{i0}}$. In the triclinic phase, with $p_{\text{i}} \geq p_{\text{i0}}$, the analytic expressions for E_4 and E_6 near a phase transition have the form

$$E_4 = -\left[-\frac{K}{2} + \left(\frac{K^2}{4} - L\right)^{1/2}\right]^{1/2}, \quad (9)$$

where $K = B_4/C_4$, $L = A_0(p_{\text{i0}} - p_{\text{i}})/C_4$, and

$$\begin{aligned} E_6 = &[-Q + (Q^2 + S^3)^{1/2}]^{1/3} \\ &+ [-Q - (Q^2 + S^3)^{1/2}]^{1/3}, \end{aligned} \quad (10)$$

where $S = A_6/3B_6$ and $Q = A_{4-6}E_4/2B_6$. Here the internal pressure at which a transition occurs,

$$p_{\text{i0}} = p_{\text{ic}} + \frac{A_{4-6}^2}{A_6A_0}, \quad (11)$$

differs from the critical pressure as a result of the interaction of the components E_4 and E_6. We attained agreement between theory and experiment by adjusting the coefficients in Eqs. (8)–(10) (Fig. 2).

Finally, the proposed model can explain the mechanism of a proper FPT, consisting in the fact that in the monoclinic phase, with increasing internal pressure, the elastic energy of the crystal accumulates as a result of the symmetric components E_i, while the elastic energy for the antisymmetric components remains zero. When the pressure p_{i0} is reached, part of the energy of the symmetric components is released into the deformation of antisymmetric components. Thus it can be said that the internal pressure is the driving force that destabilizes the crystalline structure with respect to the antisymmetric components near the point of the phase transition p_{i0}. Similar arguments were presented in [8], where the mechanical stresses in a triatomic molecule accompanying distortion of the shape of the molecule under pressure from linear to bent were calculated. It is evident on this basis that the effect of the symmetric components of the strain tensor cannot be ignored, as done previously, because they participate in the phase transition.

It is interesting to note that the internal and external pressures at which a FPT occurs differ by almost an order of magnitude. According to our model, from Eq. (11) $p_{\text{i0}} \approx 0.5$ GPa, and the external pressure for this transition is 3.2 GPa [9]. This difference is probably due to the fact that the external pressure changes the elastic moduli as a result of the anharmonicity of the interatomic potentials, while the internal pressure, in the presence of cationic exchange, acts directly on the chemical bonds and consequently changes the elastic constants more effectively.

The introduction of internal pressure elucidates a common mechanism of the influence of a change in the cationic composition Ca–Sr, Al–Si ordering and temperature on a phase transition. The replacement of Sr atoms at lattice sites by smaller Ca atoms gives rise to compression of the lattice, which is equivalent to the action of an internal pressure, which varies linearly with the Ca concentration. The effect of Al–Si ordering on a phase transition can be studied similarly. Since the Al and Si atoms also differ in size and do not occupy their positions in the presence of disorder, an additional internal pressure of the disordered structure arises. As a result, the phase transition point depends linearly on the concentration of the disordered cations Al–Si. A temperature increase results in expansion of the lattice,

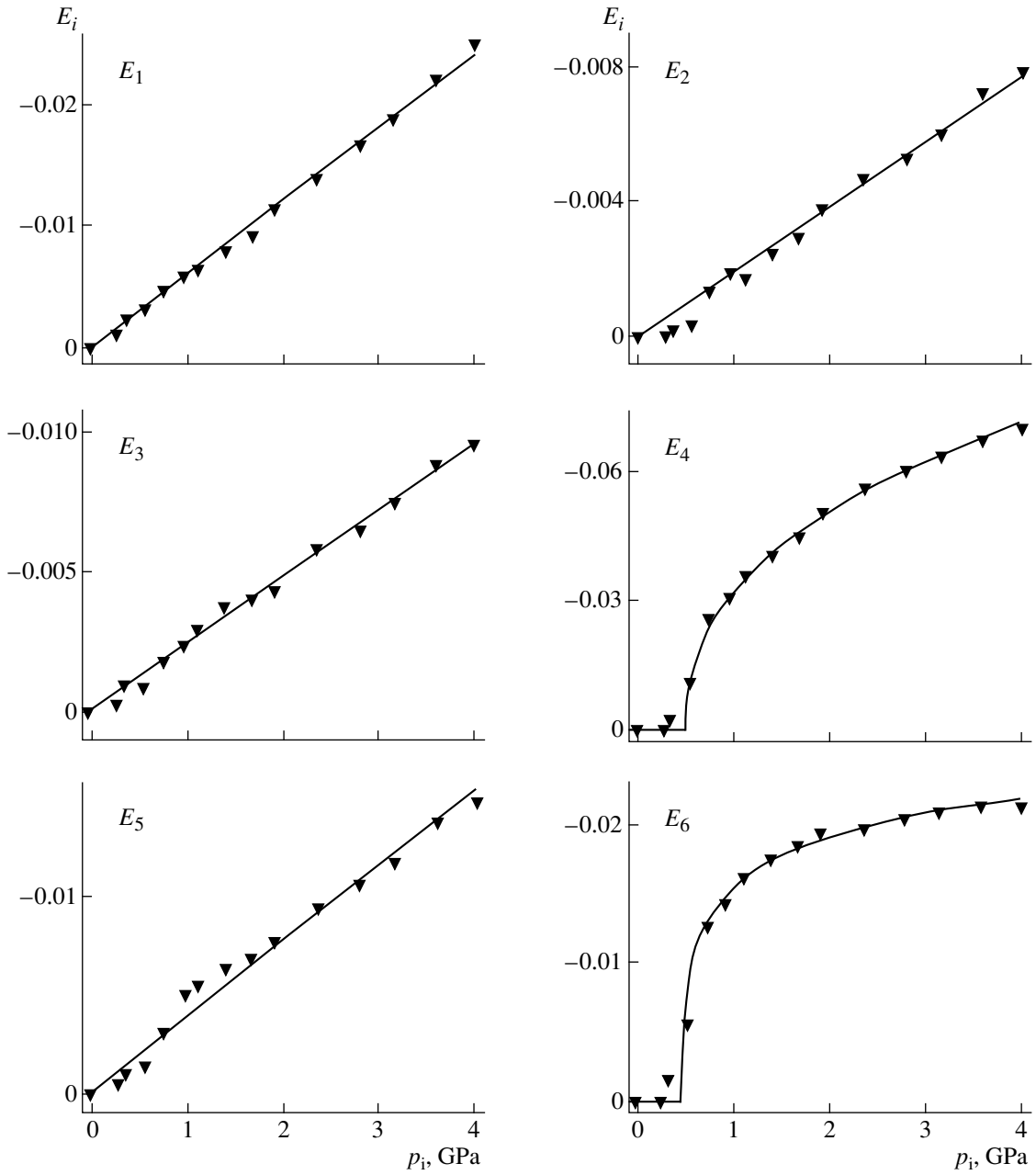

Fig. 2. Comparison of model (solid lines) and experimental (triangles) dependences of the components of strain tensors on the internal pressure. Here the internal pressure is measured from the Sr phase of the crystal, where $p_i = 0$.

which is equivalent to the action of a negative internal pressure, varying linearly with temperature. The thermal internal pressure for a α–β transition in quartz was studied in [10].

In conclusion, analytic expressions describing a proper FPT were obtained. The internal pressure was introduced in these expressions as the driving force of the transition. It was shown that the internal pressure destabilizes the crystal structure of the ferroelastic with respect to antisymmetric components of the strain tensor near the phase transition point. It was also established that the transition energy is reached as a result of the symmetric components, and consequently their contribution cannot be neglected, as done previously. Moreover, the introduction of internal pressure made it possible to represent clearly a common micromechanism by which the cationic composition Ca–Sr, Al–Si, ordering and temperature affect the phase transition.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (projects no. 98-05-65658 and 00-05-65429).

REFERENCES

1. E. K. H. Salje, *Phase Transitions in Ferroelastic and Coelastic Crystals* (Cambridge Univ. Press, Cambridge, 1990).
2. M. T. Dove and S. A. T. Redfern, Am. Mineral. **82**, 8 (1997).
3. S. A. T. Redfern, Miner. Mag. **60**, 493 (1996).
4. S. V. Goryaĭnov and N. N. Ovsyuk, Pis'ma Zh. Éksp. Teor. Fiz. **69**, 431 (1999) [JETP Lett. **69**, 467 (1999)].
5. T. V. Ryzhova, Izv. Akad. Nauk SSSR, Ser. Geofiz. **7**, 1049 (1964).
6. M. D. McGuinn and S. A. T. Redfern, Am. Mineral. **79**, 24 (1994).
7. J. L. Schlenker, G. V. Gibbs, and M. B. Boisen, Jr., Acta Cryst. A **34**, 52 (1978).
8. A. P. Mirgorodsky and M. B. Smirnov, J. Phys.: Condens. Matter **5**, 3313 (1993).
9. M. D. McGuinn and S. A. T. Redfern, Miner. Mag. **58**, 21 (1994).
10. M. B. Smirnov and A. P. Mirgorodsky, Phys. Rev. Lett. **78**, 2413 (1997).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 4, 2000, pp. 791–800.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 4, 2000, pp. 911–922.

SOLIDS
Electronic Properties

Influence of Above-Barrier and Edge States on Absorption Spectra of Shallow GaAs/AlGaAs Superlattices of Finite Length in Strong Electric Fields

V. N. Grigor'ev[a], O. V. Dorofeev[a], K. L. Litvinenko[a], V. G. Lysenko[a, *], S. I. Onishchenko[a], K. Leo[b], B. Rosam[b], and K. Kohler[c]

[a]*Institute of Microelectronic Technology and Ultra-High-Purity Materials, Russian Academy of Sciences, Chernogolovka, Moscow oblast, 142432 Russia*

[b]*Institut für Angewandte Photophysik, Technische Universität Dresden, D-01062, Dresden, Germany*

[c]*Fraunhofer-Institut für Angewandte Festkörperphysik, 79108, Freiburg, Germany*
**e-mail: lyss@ipmt-hpm.ac.ru*

Received April 21, 2000

Abstract—An investigation is made of the absorption spectra of a GaAs(82 Å)/$Al_{0.07}Ga_{0.93}As$(37 Å) superlattice at 10 K in electric fields between 0 and 60 kV/cm. By comparing the experimental absorption spectra with calculations of the energies and oscillator strengths of transitions between states of the Wannier–Stark ladder it is established that in strong electric fields above-barrier states do not form a structureless continuum but a fan of states. Intersection of electric-field-localized electron states with the fan of above-barrier states leads to broadening, splitting, and nonmonotonic changes in the intensity of an intrawell transition band. The additional absorption band observed can be attributed to intrawell transitions between states in the left and right quantum wells of the superlattice. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

In a one-dimensional isolated quantum well with fairly high walls, as a result of size quantization only a finite number (which depends on the wall height) of discrete electron states localized within the quantum well can exist. When quantum wells (an infinite number) combine to form a periodic superlattice, the discrete electron levels are transformed into subbands whose width depends on the period d of the superlattice and the transmission (material, thickness, and height) of the barriers. When a fairly strong (of the order of 10^4 V/cm) external electric field F is applied to the superlattice, the subbands are split into a fan of states

$$E_m^j(F) = E^j(0) + meFd, \quad m = 0, \pm1, \pm2, \ldots, \pm\infty, \tag{1}$$

with absolutely identical (both for electrons and for heavy and light holes) spacings between the levels eFd, forming a so-called Wannier–Stark ladder [1–3]. Figure 1a shows a schematic diagram of the left side of a superlattice in an electric field for high (dashed lines) and shallow (solid lines) barriers. The dependence of the electron energy levels on the electric field for the two lowest subbands of superlattices with high barriers is shown schematically in Fig. 1b. The wave functions $\Psi 1_m(z)$ and $\Psi 2_m(z)$ corresponding to the energy levels $E1_m(F)$ and $E2_m(F)$ in fairly strong electric fields are localized predominantly in the mth quantum well. This implies that the density of the electron wave function has a maximum in the mth quantum well (Fig. 1a) although in the nearest-neighbor wells it is not infinitely low.

Since each of the j subbands is transformed into a Wannier–Stark fan, for specific electric fields F_c the energy levels of the $E_p^i(F_c)$ and $E_k^j(F_c)$ states formed from different i and j subbands may be the same, i.e.,

$$E^i(0) + peF_cd = E^j(0) + keF_cd. \tag{2}$$

As a result of resonant interaction between electrons having the energy (2), these are localized simultaneously in the pth and kth quantum wells of the superlattice. The formation of bonding and antibonding states as a result of the mixing of wave functions of two different subbands having the same energies leads to their repulsion (Fig. 1b) [4]. Consequently, for a monotonically varying electric field instead of a linear dependence of the energies of the $E_m^i(F)$ electron states of the type (1) near resonant fields F_c we can observe numerous anticrossings (Fig. 1b) where the magnitude of the repulsion $\left|E_p^i - E_k^j\right|$ may be determined by the difference $|p - k|$ and the barrier height.

As the height of the barriers between the quantum wells decreases, the number of subbands of localized states decreases so that at a fairly low barrier height

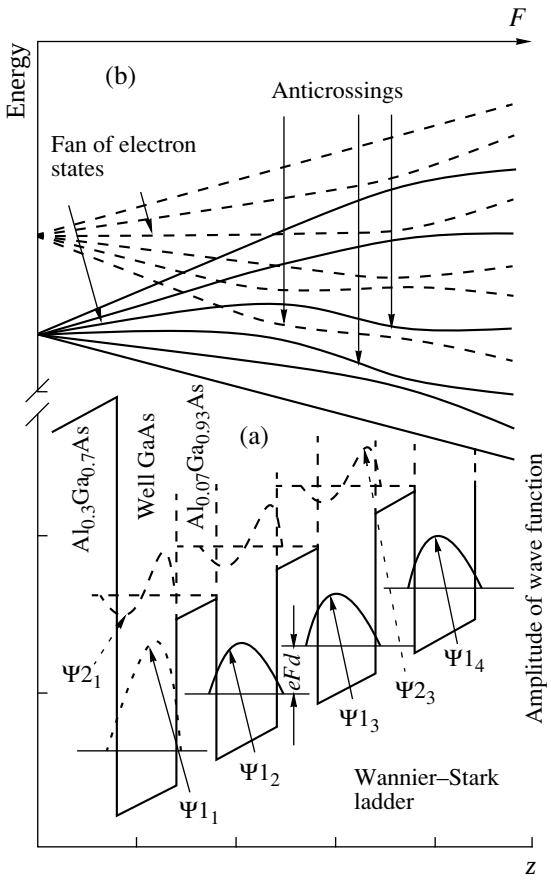

Fig. 1. (a) Dependence on the coordinate z of the potential of a GaAs/$Al_xGa_{1-x}As$ superlattice with the period d surrounded by wide $Al_{0.3}Ga_{0.7}As$ barriers in an external electric field F for small (solid lines) and large (dashed lines) x. The thin horizontal segments show the electron energy levels forming a Wannier–Stark ladder with the interlevel spacing eFd: $\Psi 1_m(z)$ and $\Psi 2_m(z)$ are the amplitudes of the wave functions of the states forming the first (solid lines) and second (dashed lines) subbands of states localized in the mth quantum well. (b) Dependence on F of the energies of the localized electron states from the first and second size-quantized subbands.

only one subband of "subbarrier" electron states may remain (this case is shown by the solid lines in Fig. 1a). The energies of the electron states of the second and subsequent subbands (shown by the dashed lines in Fig. 1) are above the barriers. The wave functions corresponding to the second and higher subbands will not be more localized in the quantum well. The qualitative difference between the wave functions of the subbarrier states and the localized wave functions is that their density in the barriers is higher than or comparable to their density in the quantum well, and the antinodes and nodes of the oscillating component of the wave function are not rigidly attached to the position of the wells and the superlattice barriers.

It is usually assumed that these subbands of "above-barrier" states form an infinite structureless continuum, which is evidently valid for an infinite periodic superlattice [5–10]. In this case, in order to facilitate analysis of the absorption spectra it is tempting to simplify their energy structure substantially by using a shallow superlattice with a barrier height such that only the states of one lowest subband are localized. In an electric field this subband would be transformed into a Wannier–Stark ladder and the ladder of above-barrier states shown in Fig. 1b would be transformed into a structureless continuum. In weak fields this continuum could lead to oscillations of the energy position and the absorption intensity as a result of Franz–Keldysh oscillations [11]. In fields $F \geq 10^4$ V/cm "anticrossings" between localized and above-barrier states should be suppressed as a result of the structureless property of their continuum.

However, it has been noted in some studies [12, 13] that in excitation spectra in shallow single quantum wells, at least some transitions involving above-barrier states form fairly narrow lines with a high oscillator strength instead of the structureless bands predicted for the continuum. The unusual position, oscillator strength, and narrow width of the corresponding lines can be attributed to the excitation of excitons formed from a particle (electron) localized in a quantum well in a size-quantized level and a free above-barrier particle of opposite polarity (hole) pulled into the well by the Coulomb attraction force. Resonant interaction between subbarrier and above-barrier states in an InGaAs/GaAs superlattice leading to singularities in the reflection spectrum was observed in [14]. In [15, 16] nonmonotonic displacements, splitting, and oscillations of the singularity amplitude were observed in the reflection spectra of a shallow GaAs/AlGaAs superlattice in electric fields between 20 and 70 kV/cm caused by anticrossings of subbarrier and above-barrier states. The results of all these studies indicate that in an electric field above-barrier states may be transformed into a fan of partially localized states. The wave functions of these states are a hybrid of localized and delocalized states: their density has a well-defined maximum in one of the quantum wells and far from this well gradually acquires the character of a delocalized wave function.

Real superlattices consist of a finite number of quantum wells $N \approx 10$–100 rather than an infinite number. States forming the lowest subband of N levels in the absence of a field are transformed with increasing field F into a set of N states each localized within only one of the N quantum wells. The lowest state is localized in the left outermost (first) well, the next states in terms of energy are localized in the second left and subsequent (positioned to the right) quantum wells such that the highest subband state is localized in the right outermost (Nth) well of the superlattice (Fig. 1a). In strong electric fields the lowest states are characterized by wave functions which decay exponentially (without oscillations) in the thick left barrier. States localized in central quantum wells in the same electric fields are not sensitive to the left outermost thick barrier so that the corresponding wave functions have an almost Bessel

wave function with an oscillating component positioned to the left of the maximum. As a result, the spacing between the energy levels at the edges and at the center of the subband differ and differ appreciably from eFd. The finite length of the superlattice leads to loss of periodicity and additional localization of states in the outermost quantum wells [17, 18]. These edge states do not have any contribution from interaction with absent wells on the left or right. The number of these states is 2–5 so that we can assume that those "equidistant" states of the Wannier–Stark ladder which are not sensitive to the influence of edge barriers are merely localized in the central $N - 5$ quantum wells of the superlattice.

2. SAMPLES AND EXPERIMENTAL RESULTS

We investigated a superlattice (Fig. 2) grown by molecular beam epitaxy on an *n*-type GaAs (001) substrate *2*, consisting of 30 GaAs(82 Å)/$Al_{0.07}Ga_{0.93}As$(37 Å) layers surrounded on the left and right by 450 and 350 nm thick $Al_{0.3}Ga_{0.7}As$ insulating layers *3* and *4* (outer barriers). A window *5* around 1 mm in diameter was etched into the substrate. This allowed focused 50 W radiation from a halogen lamp *6* (Fig. 2) to pass through the superlattice and allowed the transmission spectra to be recorded with 0.1 meV resolution using a 600 mm spectrometer *7* and an optical multichannel analyzer *8*. A dc voltage U was applied to the substrate and a semitransmitting gold contact deposited on a protective GaAs layer, creating electric fields in the range between –60 and +15 kV/cm in the superlattice. The sample *S* was attached to a sapphire substrate mounted in the ring of an optical helium cryostat *9* cooled to 8–10 K.

Figure 3 gives absorption spectra of a shallow superlattice in the exciton resonance region measured under halogen lamp illumination at temperatures below 10 K and various bias voltages U applied to the semitransmitting contact on the upper surface of the substrate. It can be seen from the figure that at positive bias voltages between +1.2 and –0.7 V the spectrum exhibits a high-intensity heavy exciton line HHX around 1 meV wide and a lower-intensity light exciton line, LHX above which extend structureless continuums. As the reverse bias voltage increases, the intensity of the exciton absorption lines decreases by two orders of magnitude and in addition to the HHX exciton line broadened by a factor of 2–3, several new lines appear. As the negative voltage increases, the spacing between the lines increases and they move apart in different directions with different derivatives, forming a fan. This behavior may be attributed to the formation of a Wannier–Stark ladder of electron $E^e(F)$, heavy $E^h(F)$, and light $E^l(F)$ hole states in accordance with Eq. (1) between which the interband optical transitions

$$\hbar\omega_{p-k} = E^e(0) - E^{h,l}(0) + (p-k)eFd, \quad p, k = 3, 4, \ldots, 27, 28, \tag{3}$$

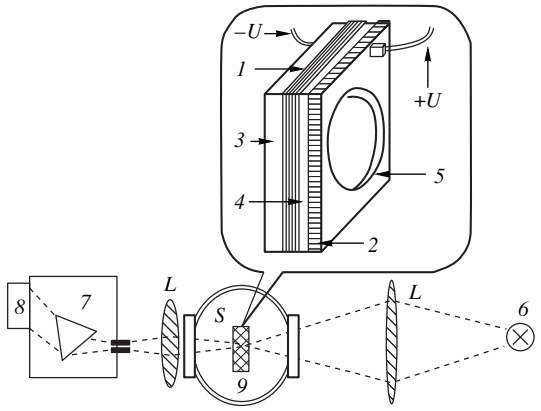

Fig. 2. Sample *S* consisting of a superlattice *1* grown by molecular beam epitaxy on a substrate *2* (with an etched window *5* up to 1 mm in diameter) between two buffer layers *3* and *4*. A transverse voltage U was applied via leads to gold contacts deposited on the substrates and coating layer. Radiation from a halogen lamp *6* passed through the sample *S* located in an optical cryostat *9* at a temperature of around 10 K was analyzed using a spectrometer *7* and a multichannel detector *8*.

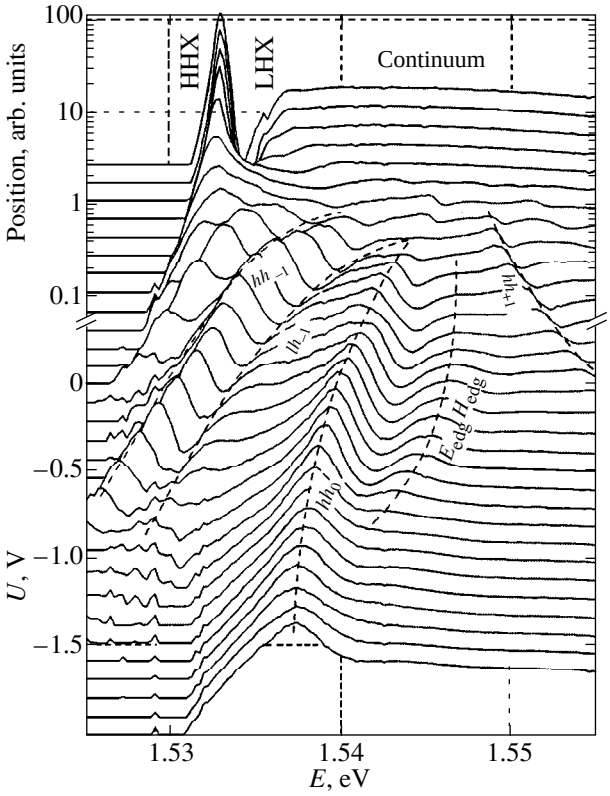

Fig. 3. Absorption spectra of a thirty-layer GaAs(82 Å)/$Al_{0.07}Ga_{0.93}As$(37 Å) superlattice at various bias voltages U in the range between +1.2 and –2.0 V and temperature of 8 K; HHX and LHX are the absorption lines of an exciton with a heavy and a light hole; hh_0 are intrawell electron–hole transitions; $hh_{-1}(lh_{-1})$ are transitions between an electron and a heavy (light) hole localized in the nearest right well; hh_{+1} are transitions between an electron and a heavy hole localized in the nearest left well; $E_{edg}H_{edg}$ are intrawell transition in the outer (first and thirtieth) wells of the superlattice.

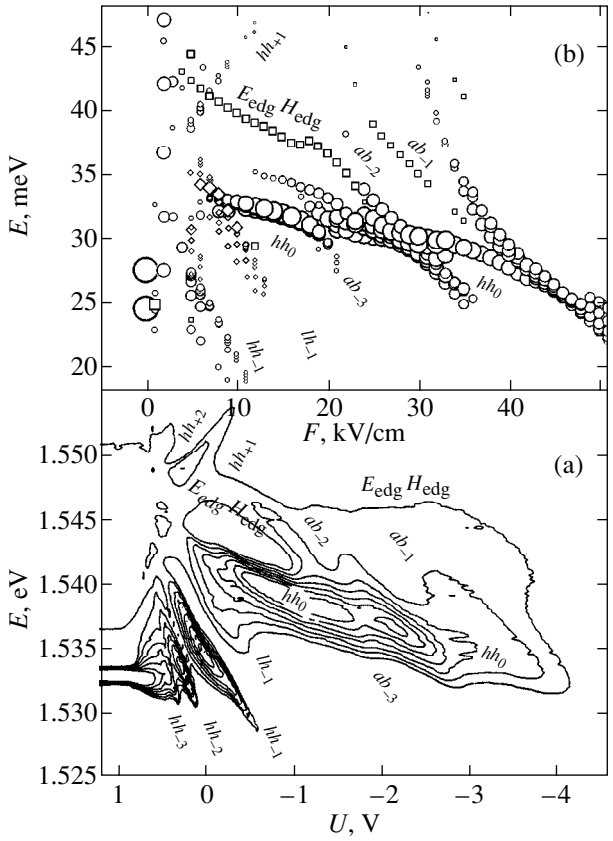

Fig. 4. (a) Contour pattern of experimentally measured absorption spectra of a thirty-layer GaAs(82 Å)/$Al_{0.07}Ga_{0.93}As$(37 Å) superlattice at bias voltages in the range between +1.2 and –4.6 V and a temperature of 8 K. (b) Theoretically calculated energies of transitions between states of electrons and heavy holes in various (apart from the outer left and right) quantum wells (circles), intrawell transitions between electrons and heavy holes in outer wells (squares), transitions between electrons and light holes localized in the nearest left quantum well (diamonds). The size of the symbols is proportional to the oscillator strength of the corresponding transitions.

lead to the formation of a fan of absorption lines. In this case, some of the observable lines may be interpreted as transitions between electrons and heavy holes localized in the neighboring quantum well on the right or left (the lines hh_{-1} and hh_{+1} in Fig. 3) or a heavy hole localized in the second quantum well on the right (the hh_{-2} line). Interpretation of the line lh_{-1} presents no significant difficulties: this corresponds to transitions between electrons and light holes localized in the neighboring quantum well to the right.

However, the field dependence of the line hh_0 which should correspond to intrawell transitions between electrons and heavy holes localized in the same well (i.e., $p = k$, p, $k = 3, 4, \ldots, 27, 28$), differs substantially from the very simple relationship (3) according to which the position of the hh_0 line should not depend on the electric field, and the transition intensity in accordance with [19] should increase monotonically in fairly high fields. Nonmonotonic variations of the energy and intensity of the intrawell transitions can be seen more clearly in Fig. 4a which shows the contour pattern of the absorption spectra of this superlattice in the spectral range between 1.525 and 1.555 eV at bias voltages between +1.2 and –4.6 V. At a forward bias between 1.2 and 0.7 V the spectrum only reveals two high-intensity lines *HHX* and *LHX* of delocalized heavy and light excitons. At biases between 0.7 and –0.7 B we observe a fan of fairly narrow lines (of the order of 1 meV) hh_{-1}, hh_{-2}, hh_{-3}, hh_{+1}, hh_{+2} corresponding to interwell transitions between electron and heavy hole states of the Wannier–Stark ladder whose position and intensity are significantly altered by Coulomb interaction between them [20–22]. It can be seen from Fig. 4a that the center of the fan is approximately situated at the point (+0.7 V; 1.545 eV, i.e., is shifted by 0.7 V as a result of the built-in field of the Schottky diode. The hh_0 line corresponding to "intrawell" transitions has a large width of ≈5 meV and is shifted toward lower energies by almost 10 meV as the field increases. The shift of the hh_0 line toward lower energies which contradicts Eq. (3) for $p - k = 0$ may be partly attributed to a "quantum-bounded Stark shift" quadratic with respect to the field, which leads to "intrawell polarization." However, the nonmonotonic variations of the intensity, half-width, and position of the hh_0 band with increasing bias voltage which are a consequence of intersections with lines of as yet unknown nature ab_{-1}, ab_{-2}, and ab_{-3}, remain unexplained. They cannot be attributed to oscillations or fluctuations of the electric field in the superlattice (for example, as a result of the formation of domains [23, 24]) since under the same conditions the other lines undergo a monotonic shift and change in intensity. In addition, the nature of the $E_{edg}H_{edg}$ line is unclear. This could be ascribed to intrawell transitions between electrons and light holes but it is then unclear why it is not shifted parallel to the hh_0 line and even intersects it.

The most probable reason for the large inhomogeneous width, nonmonotonic shift, and change in the intensity of the hh_0 line are "anticrossings" of localized states with higher electron or hole states causing its position and amplitude to change appreciably. However, we can only sufficiently reliably identify an intersection with the lh_{-1} line corresponding to recombination of an electron with a light hole localized in a neighboring well on the right. However, it is predicted that a shallow quantum well will only have one subband of localized electrons and light holes and the other subbands will be higher than the level of the barriers so that the possibility and consequences of their intersection with the hh_0 line require special theoretical analysis. The comparisons between the experimental results and corresponding calculations presented below show that the nonmonotonicity of the position, intensity, and profile of the hh_0 line may be attributed to anticrossings with above-barrier states of the shallow superlattice

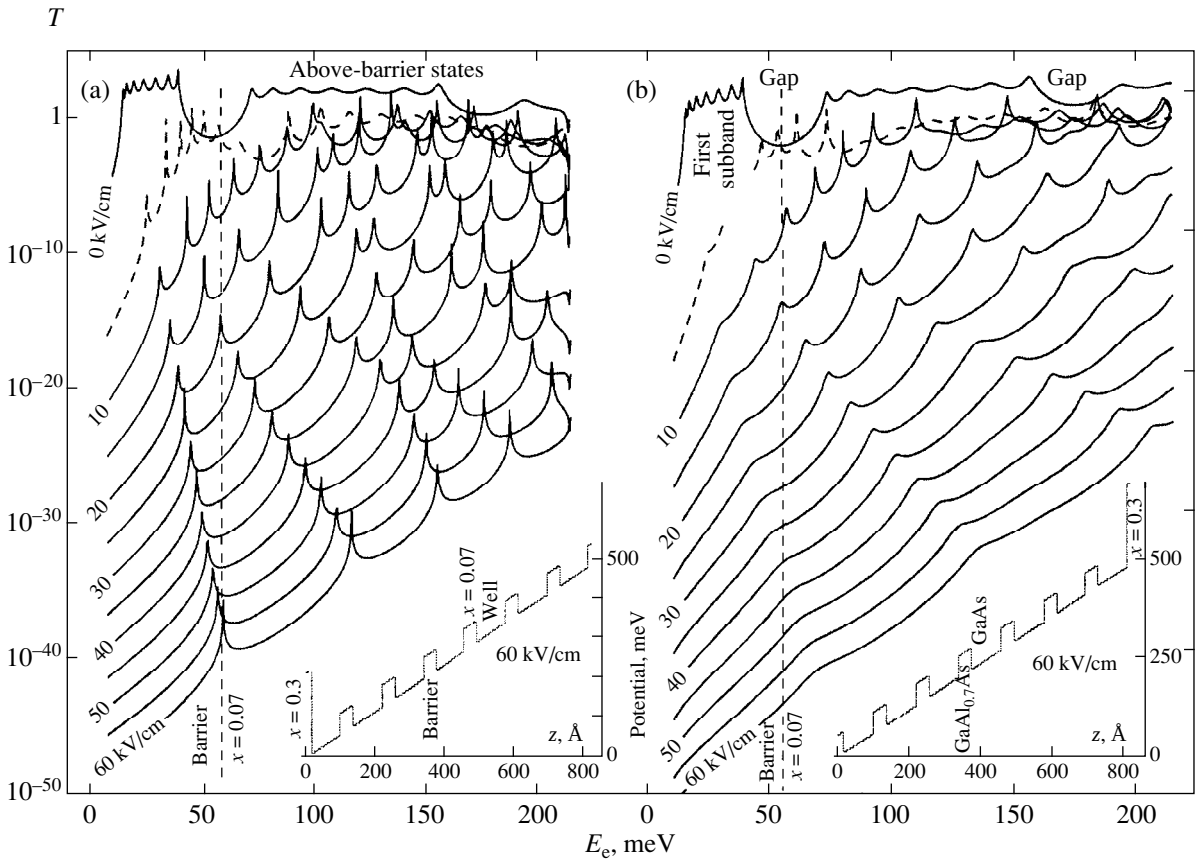

Fig. 5. Dependence of the transmission T of a seven-layer GaAs(82 Å)/$Al_{0.07}Ga_{0.93}As$(37 Å) superlattice with an insulating $Al_{0.3}Ga_{0.7}As$ barrier at its left (a) and right (b) edges on the electron energy E_e in electric fields between 0 and 60 kV/cm. The insets show the corresponding potentials for electrons in a field of 60 kV/cm. The maxima of the dependence $T(E_e)$ coincide with the allowed electron energy levels EN_m in the electric field.

while the $E_{edg}H_{edg}$ line is attributed to intrawell transitions in the outer right or left wells of the superlattice in which the size-quantized energies of the electrons and holes are shifted upward because of the contribution of high and/or wide outer insulating barriers.

3. CLASSIFICATION OF ELECTRON AND HOLE STATES IN AN ELECTRIC FIELD

It is noted in [25] that if for an electron of energy E incident on a superlattice from left to right, the latter is characterized by the transmission $T(E)$, the generalized density of electron states $\rho(E)$ in the superlattice is proportional to its transmission $T(E)$ which can be calculated by the transfer matrix method over a wide range of electric fields [26]. By analyzing the positions of the maxima of the calculated density of electron or hole states for various values of the electric field F, we can construct a fan of energy levels of electrons, heavy and light holes as a function of F and calculate the energies of the transitions between them.

The sample was simulated by asymmetric superlattices consisting of seven GaAs quantum wells 82 Å wide, separated by six $Al_{0.07}Ga_{0.93}As$ barriers 37 Å wide and 54.6 meV high for electrons or 41.1 meV for holes. One of the outer barriers coincided with the interwell barriers, another thick $Al_{0.3}Ga_{0.7}As$ barrier had a height of 214 (162) meV for the electrons (or holes) (see inset to Figs. 5a and 5b). These asymmetric structures could allow for the influence of the outer barriers on the energy structure of states localized near them. Figure 5 shows the electron transmissions of model superlattices calculated by the transfer matrix method for $Al_{0.3}Ga_{0.7}As$ barriers positioned to the left and right. The calculations allow for the Stark quantum-size effect (for transitions between electrons and holes in the same quantum well) and all anticrossings as a result of resonant overlap of the wave functions of electrons localized in different quantum wells (but neglecting Coulomb interaction between them). The maxima of the superlattice transmission $T(E)$ coincide with the allowed energy states of electrons or holes in the superlattice at the corresponding electric fields F. It can be seen from Fig. 5 that as the field increases, all the energy levels are shifted toward higher energies with different slopes so that in fields greater than 10–15 kV/cm, some of the subbarrier and above-barrier states intersect, undergoing repulsion

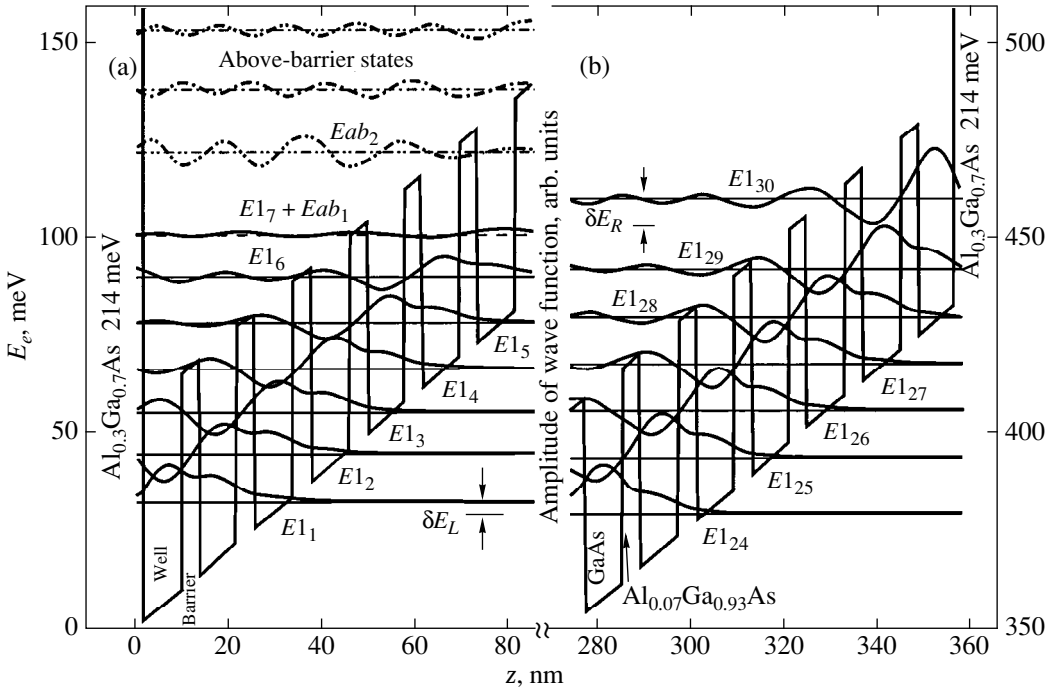

Fig. 6. Calculated electron levels (thin horizontal sections) and corresponding wave functions (thick oscillating curves) in a seven-layer GaAs(82 Å)/$Al_{0.07}Ga_{0.93}As$(37 Å) superlattice with an insulating $Al_{0.3}Ga_{0.7}As$ barrier at its left (a) and right (b) edges in an electric field of 10 kV/cm. The potential in a field of 10 kV/cm is shown by the solid broken lines. The sequence of electron levels $E1_1, E1_2, E1_3, \ldots$, transformed from the first subbarrier subband is characterized by wave functions having the maximum amplitude in the first, second, third, …, quantum wells, respectively.

(anticrossing). In fairly strong fields the energy levels may be expelled by the electric field above the levels of the outer barriers, causing substantial smearing of the maxima of the transmission curves (Fig. 5b). Nevertheless, over the entire range of fields for states sensitive to the high left-hand barrier (Fig. 5a) discrete maxima can be clearly observed not only for the subbarrier but also for the above-barrier states and their mutual anticrossing.

Figures 6 and 7 give the amplitudes of the wave functions of several electron states localized in the left and right parts of the superlattice calculated in electric fields of 10 and 30 kV/cm using the transfer matrix method. We shall first consider the profile of the electron wave functions in weak fields (10 kV/cm). It can be seen from Fig. 6 that in the lowest state the electron is localized in the outer left ($n = 1$) quantum well and its wave function $\Psi_1(z)$ has a single extremum and no node. In higher energy states the electron is localized (has a "principal" maximum of the wave function) in the nth quantum well and its wave function $\Psi_n(z)$ has $n - 1$ additional extrema of lower amplitude localized in quantum wells positioned on the left and $n - 1$ nodes localized in barriers to the left of the nth quantum well. As the energy of the states increases and the maximum of the wave function shifts toward quantum wells positioned on the right, the profile of the wave function gradually stabilizes so that at the middle or right of the superlattice the wave functions are simply shifted by a period of the superlattice d, i.e., $\Psi_n(z) = \Psi_{n-1}(z - d) = \Psi_{n+1}(z + d)$. In a superlattice consisting of 30 quantum wells 30 states with these wave functions form the lowest electron subband.

It can be seen from Fig. 6 that at the middle or right of the superlattice the energy levels of the subbarrier states are almost equidistant in accordance with Eq. (1). However, a more detailed analysis shows that the energy levels corresponding to electron localization in the outer left or right quantum well of the superlattice (edge states) are higher than that predicted by Eq. (1) by the amounts δE_L and δE_R (of the order of a few meV). This occurs because the wave functions penetrate fairly deeply into the wide high edge barriers which gives a larger contribution to the potential energy of the edge states compared with the contribution of the thin shallow interwell barriers. It will be shown below that as a result, transitions involving edge states are several meV higher than transitions between states at the center of the superlattice described by Eq. (1). Note that the equidistance of the energy levels is also destroyed because the shift of the energy levels deviates from linear with respect to the field as a result of the anticrossing of the states (see Fig. 1) although the shifts of the edge states δE_L and δE_R are of a different nature and large.

It is shown in Fig. 5a that subbands of above-barrier delocalized states (in a field of 10 kV/cm in the energy

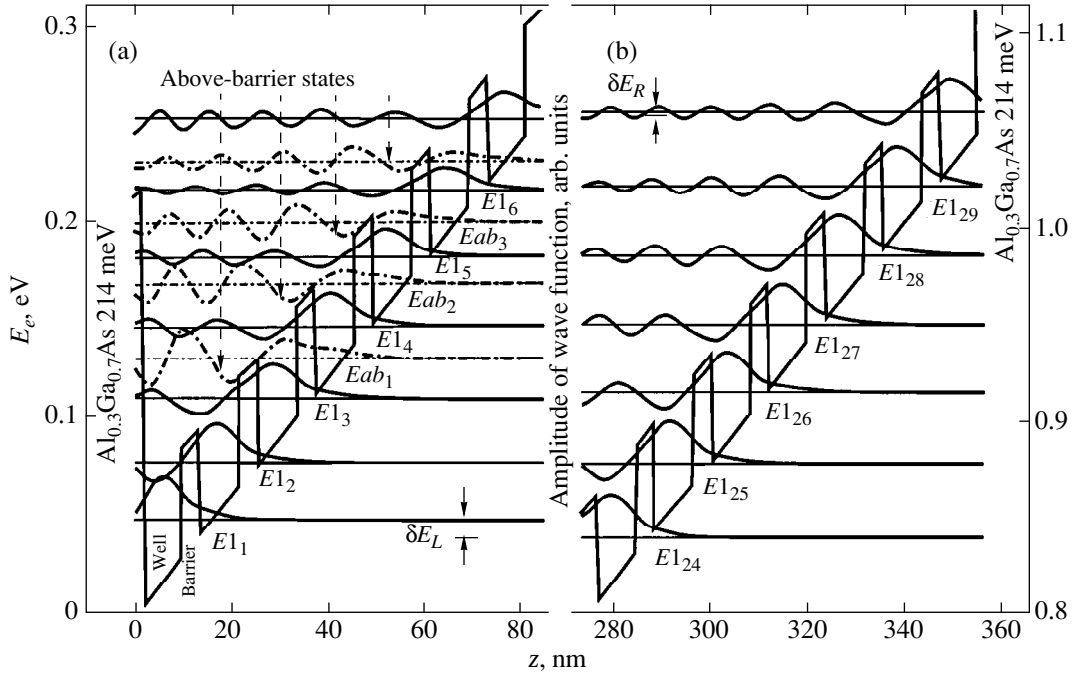

Fig. 7. Calculated electron levels and corresponding wave functions in a seven-layer GaAs(82 Å)/$Al_{0.07}Ga_{0.93}As$(37 Å) superlattice in an electric field of 30 kV/cm. The notation is the same as in Fig. 6. In a field of 30 kV/cm the localized subbarrier levels do not intersect levels of the above-barrier states so that a qualitative difference is clearly observed in the distribution of the extrema and the amplitudes of their wave functions.

range above 120 meV) are positioned above localized states. These appeared from the second or higher subbarrier size-quantized subbands which can be clearly observed in a zero electric field in Fig. 5. As the electric field increases, the positions of the energy levels and the profile of the wave functions changes radically since, in fields higher than 10 kV/cm, the above-barrier states are determined to a considerable extent by the potential of a wide triangular quantum well bounded on the left by a wide $Al_{0.3}Ga_{0.7}As$ potential barrier and below and to the right by the sawtooth bottom of the superlattice tilted by the electric field. The wave functions of the above-barrier states have an oscillating character although unlike the subbarrier states, the period of the oscillations no longer correlates with the period of the superlattice, and their extrema and nodes no longer coincide with the wells and barriers. The above-barrier electron is no longer localized in one of the quantum wells but is uniformly smeared over many wells and barriers of the superlattice. As the state energy increases the region of the superlattice over which the electron is smeared increases and as a result the amplitude of the wave function decreases. Thus, in the right part of the superlattice where the bottom of each quantum well is shifted upward by 350–400 meV by the electric field, only the wave functions of the localized states have any appreciable amplitude and the amplitude of the above-barrier states is incommensurably small. It can be seen from Fig. 6a that as a result of resonant mixing of the wave function of the $E1_7$ state localized in the seventh quantum well with the wave function of the lowest above-barrier state Eab_1, the amplitude of the resultant wave function differs substantially from the wave functions of the other states. This may lead to a reduction in the contribution of electrons localized in the seventh well to the optical transitions involving them.

The energy gap between the states of the subbarrier and above-barrier subbands which can be observed clearly in zero electric field (Fig. 5) begins to decrease with increasing field so that in fields of the order of 15 kV/cm the gap disappears completely and the above-barrier and subbarrier states become mixed (the curve corresponding to 15 kV/cm in Fig. 5a). As a result, in an electric field of 30 kV/cm (Fig. 7) at energies higher than 120 meV, states having features of above-barrier states alternate with subbarrier states localized predominantly in a single quantum well. It can be seen from Fig. 7 that the difference in the profile of the wave functions of the above-barrier and subbarrier states decreases as a result of their mixing and the main criterion becomes the position of the maximum extremum: if this is positioned at the right edge of the wave function, this is a subbarrier states whereas for other positions of the maximum we can talk of the subbarrier nature of the corresponding state. As will be demonstrated below (see Figs. 6 and 8), in some fields where anticrossing of two or more states is observed, the wave function can simultaneously have features of several types of states.

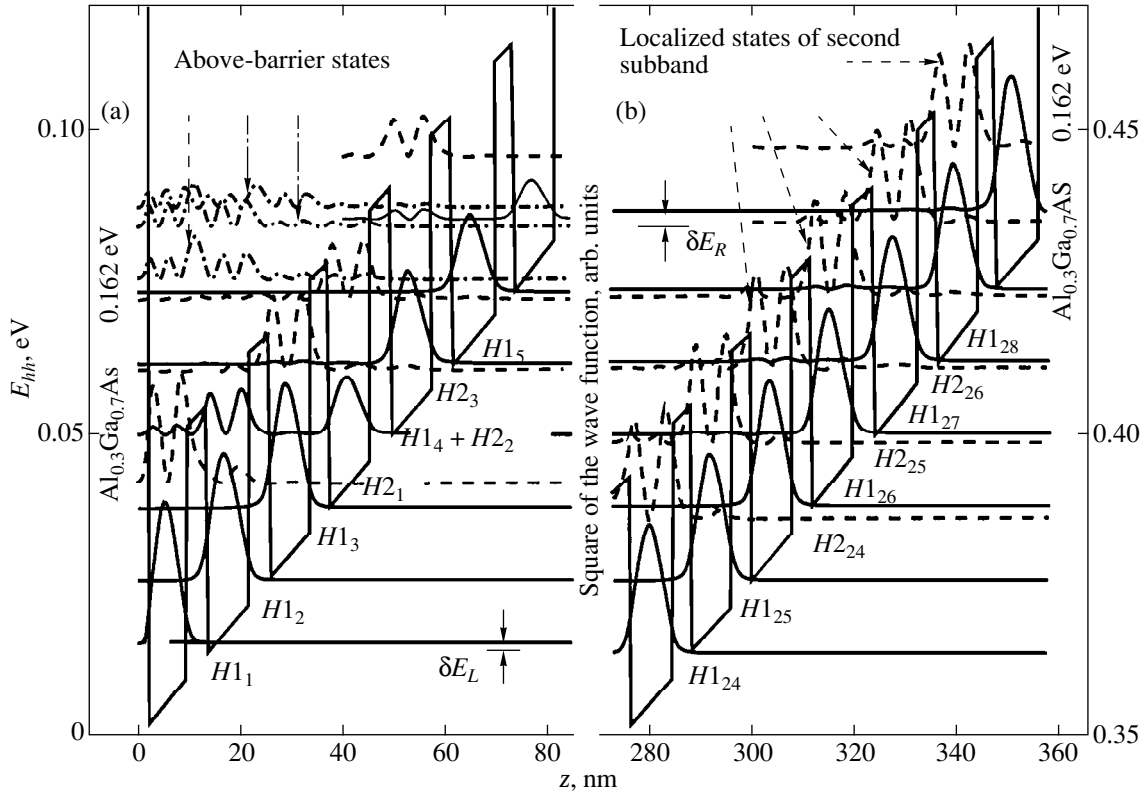

Fig. 8. Calculated energy levels of heavy holes and the corresponding squares of the wave function amplitudes in a seven-layer GaAs(82 Å)/$Al_{0.07}Ga_{0.93}As$(37 Å) superlattice with an $Al_{0.3}Ga_{0.7}As$ insulating barrier at its left (a) and right (b) edges in an electric field of 10 kV/cm. The states $H1_1$, $H1_2$, …, $H1_{28}$ are completely localized in the first, second, …, twenty-eighth quantum wells. Above are the $H2_1$, $H2_2$, …, $H2_{26}$ levels originating from the second subband of subbarrier states. Their wave functions (indicated by the dashed lines) are also localized in a single quantum well, have two extrema and a node at the center.

Figure 8 shows the energy levels and squares of the modulus of the corresponding wave functions of heavy holes in the left and right parts of the superlattice in an electric field of 10 kV/cm. A comparison with Figs. 6 and 7 shows that in the lower states heavy holes, unlike electrons, are almost completely localized in one of the quantum wells. The set of these states $H1_1$, $H1_2$, $H1_3$, and so on forms the first subband of localized subbarrier states. We recall that in zero electric field the heavy holes have not one but two subbands of localized subbarrier states. The wave functions of states from the second subband are characterized by the fact that they have two extrema and one node within a single quantum well and the probability of finding a hole outside this well is low. Typical examples of the wave functions of holes corresponding to the second subband of subbarrier states are shown by the dashed lines in Fig. 8 and are denoted by $H2_1$, $H2_3$, $H2_4$, and so on. It can be seen that in a field of 10 kV/cm states from the first and second subbands are not separated by a gap as in zero field but are mixed, although they can be classified according to the form of the wave functions. The energy levels of the holes and electrons are also almost equidistant except for the edge states localized in the outer left and right quantum wells of the superlattice shifted upward by the values of δE_L and δE_R of the order of a few meV. The states $H1_4$ and $H2_2$ from the first and second subbands are almost the same in a field of 10 kV/cm so that the wave functions describing them have characteristic features of the first and second subbands, respectively. As a result, the extrema of the wave function are approximately half those for states in the absence of anticrossing which leads to a corresponding reduction in the oscillator strengths of transitions involving the $H1_4$ and $H2_2$ states in the field F = 10 kV/cm.

In addition to subbarrier states, heavy holes also have above-barrier states whose wave functions are more uniformly smeared over many quantum wells and have extrema away from the right edge. Examples of these states and their corresponding wave functions are given by the dot-dash lines in Fig. 8a. The subbarrier states may undergo an anticrossing with the above-barrier states so that the corresponding energy levels deviate from Eq. (1) and the wave functions change their

profile and amplitude. Figure 8a shows that in the energy range 87 ± 2 meV two above-barrier and two subbarrier states undergo an anticrossing with the result that the amplitude of the combination of the wave functions of the latter is reduced significantly. The wave function of this state is a superposition of the wave functions of the above-barrier and subbarrier states where the hole is localized simultaneously in one of the quantum wells and above several barriers.

Note that the transformations of the wave functions under conditions of anticrossing of subbarrier and above-barrier states is not an exclusive characteristic of heavy holes. They may also occur in the case of anticrossings of electron states in corresponding electric fields.

Using the energy levels thus calculated and their wave functions in the range of electric fields between 0 and 60 kV/cm, we can calculate all possible combinations of intrawell and interwell transitions and the corresponding matrix elements of the transitions and thus construct a theoretical fan of transitions involving edge states, mixing of subbarrier and above-barrier states and the associated nonequidistance of their position.

Figure 4b gives the calculated energies of transitions between various states of electrons, heavy and light holes in a GaAs(82 Å)/$Al_{0.07}Ga_{0.93}As$(37 Å) superlattice in electric fields between 0 and 60 kV/cm (neglecting Coulomb interaction between electrons and holes). The overlaps of the wave functions of the states calculated at the same time were used to determine the oscillator strengths of the corresponding transitions. The size of the symbols in Fig. 4b is proportional to the estimated oscillator strengths. Figure 4b can be used to determine the contribution of various transitions to the absorption spectrum of the superlattice allowing for their intensity. Each symbol may be ascribed a transition involving different types of states in accordance with the classification given above. All the transitions are divided into three groups: the open circles give the energies and oscillator strengths of transitions between electrons and heavy holes in quantum wells located at distances of more than two periods from the outer quantum wells. Intrawell transitions in wells far from the edge form the hh_0 absorption band up to 2–3 meV wide. This nonuniform width is attributed to the nonequidistance of the distribution of electron and hole states shown in Figs. 5–8 for wells positioned at different distances from the outer barriers: the energies of the intrawell transitions at the center of the superlattice are minimal whereas in a quantum well at a distance of between 1 and 5 periods from the edge of the superlattice the transition energies increase monotonically as the edge barrier is approached. It can be seen from Fig. 4b that the inhomogenous broadening depends nonmonotonically on the electric field, increasing further in the presence of a crossing (or more accurately anticrossing) of the hh_0 band with other lines of comparable intensity (for example, in fields of 23–32 and 38–42 kV/cm). In fields of 18–20 kV/cm the hh_0 band splits in two as a result of the anticrossing of localized and above-barrier electron states.

In the range of fields between 4 and 13 kV/cm interwell transitions between electrons and holes localized in neighboring quantum wells form two lines hh_{-1} and hh_{+1} of lower intensity. It can be seen from Fig. 6 that in this range of fields above-barrier and subbarrier states do not undergo anticrossings so that the positions of the lines vary linearly with the field and the hh_{-1} and hh_{+1} lines are considerably narrower than hh_0. The line lh_{-1} shown by the diamonds is attributed to interwell transitions between electrons and light holes localized in neighboring wells on the right. It is parallel to the hh_{-1} line and shifted relative to the latter by around 9–10 meV which is equal to the difference between the energy levels of the light and heavy holes.

The line $E_{edg}H_{edg}$ shown by the squares is attributed to intrawell transitions between electrons and heavy holes localized in outer right and left wells of the superlattice. As we have discussed, the edge electron and hole states in the quantum wells nearest to the insulating barriers are shifted upward by several meV relative to the levels predicted in accordance with Eq. (1) with the result that the transitions associated with these edge states are also shifted upward by around 5–8 meV.

4. CONCLUSIONS

In a shallow superlattice bounded on both sides by high insulating barriers in a fairly strong electric field the above-barrier subbands form a fan of states instead of the assumed structureless continuum. As a result of the influence of the high edge barriers the energy level distribution of the electrons and holes localized by the electric field departs from the Wannier–Stark ladder and becomes non-equidistant. Anticrossings of the fan of non-equidistant subbarrier states with the fan of above-barrier states leads to inhomogeneous broadening of the band of intrawell transitions, and nonmonotonic variation of its width and intensity with respect to the field. Transitions involving edge states differ by 5–8 meV from transitions in wells near the center of the superlattice, are characterized by elevated oscillator strength and form an additional line $E_{edg}H_{edg}$.

ACKNOWLEDGMENTS

This work was supported financially by the Russian Foundation for Basic Research (project no. 98-02-16153) and by the program “Physics of Solid-State Nanostructures” (project no. 97-1058).

REFERENCES

1. G. H. Wannier, Phys. Rev. **117**, 432 (1969).
2. E. E. Mendez, F. Agullo-Rueda, and J. M. Hong, Phys. Rev. Lett. **60**, 2426 (1988).

3. P. Voisin, J. Bleuse, C. Bouche, *et al.*, Phys. Rev. Lett. **61**, 1639 (1988).
4. Yu. A. Bychkov and A. M. Dykhne, Zh. Éksp. Teor. Fiz. **48**, 1174 (1965) [Sov. Phys. JETP **21**, 783 (1965)].
5. H. X. Jiang, E. X. Ping, P. Zhou, *et al.*, Phys. Rev. B **41**, 12949 (1990).
6. T. Rappen, U. Peter, M. Wegener, *et al.*, Phys. Rev. B **48**, 4879 (1993).
7. P. Leisching, T. Dekorsy, H. J. Bakker, *et al.*, Phys. Rev. B **51**, 18015 (1995).
8. D. Birkedal, V. G. Lyssenko, J. M. Hvam, *et al.*, Phys. Rev. B **54**, R14250 (1996).
9. T. Dekorsey, A. M. Kim, G. C. Cho, *et al.*, Phys. Rev. Lett. **77**, 3045 (1996).
10. K. El Sayed, D. Birkedal, V. G. Lyssenko, *et al.*, Phys. Rev. B **55**, 2456 (1997).
11. K. H. Schmidt, N. Linder, G. H. Döhler, *et al.*, Phys. Rev. Lett. **72**, 2769 (1994).
12. V. D. Kulakovskiĭ, A. Forchel, K. Pieger, *et al.*, Phys. Rev. B **50**, 7467 (1994).
13. E. A. Mulyarov, N. N. Sibel'din, M. L. Skorikov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **70**, 613 (1999) [JETP Lett. **70**, 621 (1999)].
14. T. Kita, T. Inazumi, H. Nakayama, *et al.*, Phys. Rev. B **50**, 2420 (1994).
15. I. Tanaka, M. Nakayama, H. Nishimura, *et al.*, Phys. Rev. B **48**, 2787 (1993).
16. M. Nakayama, M. Ando, I. Tanaka, *et al.*, Phys. Rev. B **51**, 4236 (1995).
17. J. P. Hagon and M. Jaros, Phys. Rev. B **41**, 2900 (1990).
18. G. Bongiovanni and J. L. Staehly, Phys. Rev. B **39**, 8359 (1989).
19. P. Leisching, P. Haring Bolivar, W. Beck, *et al.*, Phys. Rev. B **50**, 14389 (1994).
20. M. M. Dignam and J. E. Sipe, Phys. Rev. B **41**, 2865 (1990).
21. M. M. Dignam and J. E. Sipe, Phys. Rev. Lett. **64**, 1797 (1990).
22. M. M. Dignam and J. E. Sipe, Phys. Rev. B **43**, 4097 (1991).
23. H. T. Grahn, H. Schneider, and K. von Klitzing, Phys. Rev. B **41**, 2890 (1990).
24. V. N. Murzin and Yu. A. Mityagin, Usp. Fiz. Nauk **169**, 464 (1999).
25. M. Ritze, N. J. M. Horing, and R. Enderlein, Phys. Rev. B **47**, 10437 (1993).
26. M. Born and E. Wolf, *Principles of Optics* (Pergamon, Oxford, 1969; Nauka, Moscow, 1973).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 4, 2000, pp. 801–816.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 4, 2000, pp. 923–940.

SOLIDS
Electronic Properties

Phenomenological Description of the Microwave Surface Impedance and Complex Conductivity of High-T_c Single Crystals¶

M. R. Trunin[a, *], Yu. A. Nefyodov[a], and H. J. Fink[b]

[a]*Institute of Solid-State Physics, Russian Academy of Sciences, Chernogolovka, Moscow oblast, 142432 Russia*
**e-mail: trunin@issp.ac.ru*

[b]*Department of Electrical and Computer Engineering, University of California, Davis, California 95616, USA*

Received May 4, 2000

Abstract—Measurements of the microwave surface impedance $Z_s(T) = R_s(T) + iX_s(T)$ and the complex conductivity $\sigma_s(T)$ in the *ab*-plane of high-quality high-T_c YBCO, BSCCO, TBCCO, and TBCO single crystals are analyzed. Experimental data of $Z_s(T)$ and $\sigma_s(T)$ are compared with calculations based on a modified two-fluid model that includes a temperature-dependent quasiparticle scattering and a unique temperature variation of the density of superconducting carriers. We describe the agreement and disagreement of our analysis with the salient features of the experimental data. We review the existing microscopic models based on unconventional symmetry types of the order parameter and on novel quasiparticle relaxation mechanisms. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

High-precision microwave measurements of the temperature dependence of the surface impedance $Z_s(T) = R_s(T) + iX_s(T)$ of high-T_c superconductors (HTSC's) considerably advance our understanding about the pairing of the superconducting electrons in these materials. In particular, in 1993, the linear *T*-dependence of the penetration depth $\lambda(T) - \lambda(0) \propto \Delta X_s(T) \propto T$ observed below 25 K in the *ab*-plane of high-quality $YBa_2Cu_3O_{6.95}$ (YBCO) single crystals [1] gave rise to productive investigations of the order parameter of HTSC's. This linear variation of $\lambda(T)$ at low *T* has by now been observed not only in orthorhombic YBCO single crystals [2–14] and films [15–18] but also in tetragonal $Bi_2Sr_2CaCu_2O_8$ (BSCCO) [19–22], $Tl_2Ba_2CuO_{6+\delta}$ (TBCO) [23, 24], and $Tl_2Ba_2CaCu_2O_{8-\delta}$ (TBCCO) [10] single crystals. This temperature dependence does not agree with the nearly isotropic superconducting gap, and it is now considered as a strong evidence for the *d*-wave pairing in these materials [25–35] in spite of the fact that the experimental data are not sensitive to the phase of the superconducting order parameter. Later research has shown that $\Delta\lambda_{ab}(T)$ can be linear at low *T* for models invoking the proximity effect between normal and superconducting layers [36] or assuming the anisotropic *s*-wave pairing [37–39]. However, none of these theories can give an explanation for the substantially different slopes of $\Delta\lambda_{ab}(T)$ at low *T* in the YBCO samples grown by different methods [40] and for certain features such as a bump [9, 11, 16, 41] or a plateau [8, 10, 12], which are observed in the intermediate temperature range $0.4\,T_c < T < 0.8\,T_c$. Models containing a mixed $(d + s)$ symmetry of the order parameter [42–56] hold some promise for a successful description of these experimental features, but this would require additional theoretical investigations.

Another important feature of the microwave response of HTSC crystals is the linear variation with temperature of the surface resistance $R_s(T)$ in the *ab*-plane at low temperatures. At frequencies about 10 GHz and below, the *T*-dependence of $R_s(T)$ in BSCCO, TBCO, and TBCCO single crystals is linear over the range $0 < T \lesssim T_c/2$ [19, 21–23]. For YBCO crystals, $\Delta R_s(T) \propto T$ for $T \lesssim T_c/3$ and $R_s(T)$ displays a broad peak with a valley at higher temperatures [4–14, 57–61]. This peak can be understood as a competition between an increase in the quasiparticle lifetime and a decrease in the quasiparticle density as the temperature is lowered. The sufficiently slow decrease in the quasiparticle density is indicative of a highly anisotropic or unconventional order parameter, resulting in a very small or vanishing energy gap, while the increase in the quasiparticle lifetime is attributed to the presence of inelastic scattering, which can be due (i) to the exchange of antiferromagnetic spin fluctuations [62], which would naturally lead to the *d*-wave pairing, or (ii) to a strong electron–phonon interaction [63–65] within the anisotropic *s*-wave pairing model [66, 67]. Moreover, there have been suggestions of unconventional states for describing the charge carriers in

¶ This article was submitted by the authors in English.

the CuO_2 planes like the marginal Fermi liquid [68, 69] and the Luttinger liquid [70, 71]. However, to fit the data of YBCO, the inelastic scattering rate must decrease with temperature much faster than predicted by any of these microscopic models. Further, the *d*-wave model with point scatterers does predict a finite low-temperature and low-frequency limit, which is independent of the concentration and the strength of the scattering centers [72]. Therefore, the latter model does not explain the very different values of the observed residual surface resistance $R_{res} \equiv R_s$ ($T \longrightarrow 0$) on different samples. Furthermore, the value of this universal surface resistance is much lower than the R_{res}values obtained from experiments. There is no microscopic theory explaining the linear temperature dependence of $\Delta R_s(T)$ up to $T_c/2$ in crystals with a nonorthorhombic structure and the shoulder of $R_s(T)$observed on YBCO [9, 11] for $T > 40$ K.

In the absence of a generally accepted microscopic theory, a modified two-fluid model for calculating $Z_s(T)$ in HTSC single crystals has been proposed independently in [73, 74] and then further developed in [8, 40, 61, 75]. Our phenomenological model has two essential features that make it different from the well-known Gorter–Casimir model [76]. The first is the introduction of the temperature dependence of the quasiparticle relaxation time $\tau(t)$ (with $t \equiv T/T_c$) described by the Gruneisen formula (electron-phonon interaction), and the second feature is the unique density of superconducting electrons $n_s(t)$, which gives rise to a linear temperature dependence of the penetration depth in the *ab*-plane at low temperatures:

$$\frac{\lambda^2(0)}{\lambda^2(t)} = \frac{n_s(t)}{n} \simeq 1 - \alpha t, \tag{1}$$

where $n = n_s + n_n$ is the total carrier density and α is a numerical parameter in our model.

The goal of this paper is to demonstrate the power of our model in describing the general and distinctive features of the surface impedance $Z_s(T)$ and the complex conductivity $\sigma_s(T)$ in the superconducting and normal states of different HTSC crystals (whose doping level corresponds to the highest T_c) at various frequencies. Section 2 describes the systematization of the $Z_s(T)$ measurements, including the analysis that is used to extract $\sigma_s(T)$ from the measured values of $Z_s(T)$. Section 3 compares experimental data of $Z_s(T)$ and $\sigma_s(T)$ over the entire temperature range with calculations based on our modified two-fluid model. In the conclusion, we compare the concepts of our model with results of microscopic theories. We hope that this can be a helpful guide for future investigations of microwave properties of HTSC's from a microscopic point of view.

2. ANALYSIS OF EXPERIMENTAL RESULTS

2.1. Surface Impedance

The surface impedance in the *ab*-plane of HTSC's, expressed in terms of the complex conductivity $\sigma_s = \sigma_1 - i\sigma_2$, obeys the local equation

$$Z_s = R_s + iX_s\left(\frac{i\omega\mu_0}{\sigma_1 - i\sigma_2}\right)^{1/2}. \tag{2}$$

The impedance components are

$$R_s = \sqrt{\frac{\omega\mu_0(\varphi^{1/2} - 1)}{2\sigma_2\varphi}}, \tag{3}$$

$$X_s = \sqrt{\frac{\omega\mu_0(\varphi^{1/2} + 1)}{2\sigma_2\varphi}}, \tag{4}$$

where $\varphi = 1 + (\sigma_1/\sigma_2)^2$. It is obvious that $R_s < X_s$ for $T < T_c$.

For temperature $T < T_c$ and with $\sigma_1 \ll \sigma_2$, Eqs. (3) and (4) reduce to

$$R_s \simeq \frac{(\omega\mu_0)^{1/2}\sigma_1}{2\sigma_2^{3/2}} = \frac{1}{2}\omega^2\mu_0^2\sigma_1\lambda^3, \qquad X_s \simeq (\omega\mu_0/\sigma_2)^{1/2} = \omega\mu_0\lambda. \tag{5}$$

The surface impedance components are measurable quantities. The real part of the surface impedance, the surface resistance R_s, is proportional to the loss of the microwave power. It is caused by the presence of "normal" carriers. In the centimeter wavelength band, typical values of the surface resistance in the *ab*-plane of HTSC single crystals are between 0.1 and 0.3 Ω above the transition temperature T_c. When T is decreased through T_c, the surface resistance abruptly drops but does not seem to approach zero as $T \longrightarrow 0$. In conventional superconductors (like Nb), $R_s(T)$ decreases exponentially upon decreasing the temperature below $T_c/2$, approaching a constant residual surface resistance R_{res} as $T \longrightarrow 0$. R_{res} is due to the presence of various defects in the surface layer of the superconductor. Therefore, it is generally accepted that lowering R_{res} leads to improving the sample quality. In high-quality HTSC's, there is no plateau in $R_s(T)$ for $T \ll T_c$. However, we extrapolate the value of $R_s(T)$ to $T = 0$ K and denote it by R_{res}. The origin of the residual surface resistance observed in HTSC crystals remains unclear. It is known that R_{res} is strongly material and sample dependent and is approximately proportional to the square of the frequency. At present, very small values ($R_{res} \approx 20$ $\mu\Omega$ at frequencies $\approx$10 GHz) are observed in YBCO single crystals [9, 14].

The imaginary part of the surface impedance, the reactance X_s, is mainly determined by the superconducting carriers and is due to nondissipative energy stored in the surface layer of the superconductor.

In the table, we summarize the main features of the temperature dependence of the surface impedance of high-quality YBCO, BSCCO, TBCO, and TBCCO single crystals whose residual surface resistance in the *ab*-plane (R_{res}) is less than one milliohm at the frequency ≈10 GHz with the $R_s(T_c)$ values about 0.1 Ω. There is good reason to believe that the temperature behavior of the electrodynamic parameters of these crystals is adequately related to the intrinsic microscopic properties of the superconducting state of HTSC.

To illustrate the data of the table, we show in Fig. 1, as an example, experimental data of $R_s(T)$ and $X_s(T)$ in the *ab*-plane of a BSCCO single crystal at 9.4 GHz [22]. In this figure, $R_s(T) = X_s(T)$ for $T \geq T_c$, which corresponds to the normal skin-effect condition. Knowing $R_s(T_c) = \sqrt{\omega\mu_0\rho(T_c)/2} \approx 0.12$ Ω, we obtain the resistivity $\rho(T_c) \approx 40$ μΩ cm. In the normal state (above T_c), the temperature dependence of $R_s(T) = X_s(T)$ is adequately described by the expression $2R_s^2(T)/\omega\mu_0 = \rho(T) = \rho_0 + bT$. For the BSCCO crystal in Fig. 1, $\rho_0 \approx 13$ μΩ cm and $b \approx 0.3$ μΩ cm/K. The insets in Fig. 1 show $R_s(T)$ and $\lambda(T) = X_s(T)/\omega\mu_0$ for $T < 0.7\ T_c$ plotted on a linear scale. The extrapolation of the low-temperature sections of these curves to $T = 0$ K yields the estimates $R_{res} \approx 0.5$ mΩ and $\lambda_{ab}(0) = 2600$ Å for this crystal.

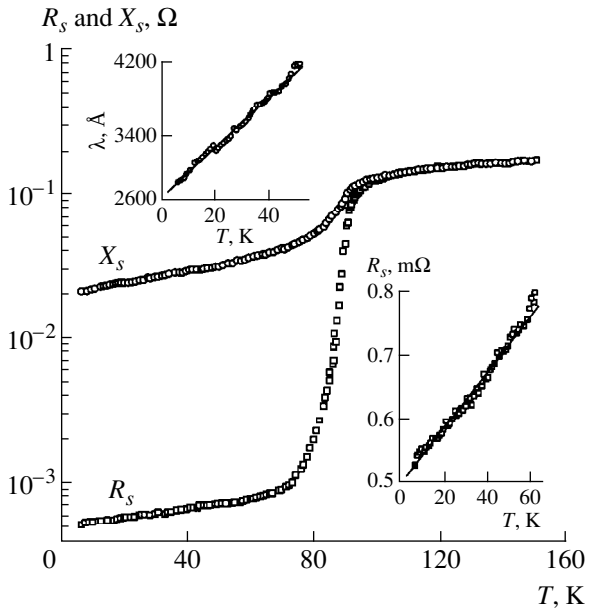

Fig. 1. Surface resistance $R_s(T)$ and reactance $X_s(T)$ in the *ab*-plane of a BSCCO single crystal at 9.4 GHz. The insets show linear plots of $\lambda(T)$ and $R_s(T)$ at low temperatures.

The experimental curves of $\Delta\lambda_{ab}(T)$ for YBCO, TBCO, and TBCCO crystals are also linear in the range $T < T_c/3$. It is important to note different slopes of the $\Delta\lambda(T) \propto T$ curves for $T \ll T_c$. In particular, in YBCO crystals fabricated by different techniques, the slopes of $\Delta\lambda_{ab}(T)$ differ by almost one order of magnitude [8, 9, 13]. The reasons for such a discrepancy are still unclear.

At frequencies about 10 GHz and below, the linear dependence $\Delta R_s(T) \propto T$ in BSCCO (Fig. 1), TBCCO, and TBCO single crystals may actually extend to the temperatures $\sim T_c/2$. This property, which is common for all HTSC crystals with the tetragonal structure, is not characteristic of YBCO. As noted previously, all microwave measurements on high-quality YBCO single crystals show a broad peak in the $R_s(T)$ curve centered near 30–40 K up to the frequencies ≈10 GHz. The peak shifts to higher temperatures and diminishes in size as the frequency is increased. In higher quality YBCO crystals, the peak amplitude increases and $R_s(T)$ reaches its maximum at a lower temperature [14].

The underlying origin of this YBCO feature has remained unclear. The simplest idea is that the absence of this peak in crystals with tetragonal structure might be caused by their "poor" quality, as is the case with YBCO doped with Zn [2, 4, 58]. However, this conclusion is probably incorrect because (i) there is a sufficiently large set of experimental data indicating that $R_s(T)$ is a linear function of T for BSCCO, TBCO, and TBCCO, and (ii) the peak in $R_s(T)$ was also detected in the YBCO crystals [7, 10, 60] with the parameters R_{res} and $\rho(T_c)$ that would characterize the quality of these crystal as being "poor" compared to those of, for example, TBCCO [10] or BSCCO [21]. Results for the latter crystals are shown in Fig. 2. The more probable cause of the peak, however, is the presence of an additional component in the YBCO orthorhombic structure, namely, the CuO chains that lead to a mixed ($d + s$) symmetry of the order parameter in YBCO. The electrons of the chains form an additional band, contributing to the observed *T*-dependence of $Z_s(T)$. This contribution seems to result in another distinctive feature of YBCO, namely, a plateau or a bump (see table) on the $\lambda_{ab}(T)$ curve, which has been observed in high-quality YBCO single crystals [8–12] and films [16, 41]. However, recent measurements of $\Delta\lambda_{ab}(T)$ in the YBCO crystals [14] grown in a high-purity $BaZrO_3$ crucible do not show such features in the intermediate temperature range. The authors of [14] argue that the disagreement with the results of [9] arises from a problem related with the surface of the crystal. The last observation still lacks a convincing explanation.

Finally, another feature in the *T*-dependence of the impedance of high-quality YBCO crystals was detected: a noticeable increase of $R_s(T)$ with an increasing temperature (a shoulder) at temperatures larger than the peak (30 K). It turns out that this shoulder is reproducible in the experiments [9, 11]. Similarly, an explanation of this observation is lacking.

Surface impedance $Z_s(T) = R_s(T) + iX_s(T)$ in the *ab*-plane of high-T_c single crystals at frequencies ~10 GHz

HTSC	Superconducting state, $T < T_c$			Normal state $1.5T_c > T \geq T_c$
	low temperatures $4\ \text{K} < T \ll T_c$	intermediate temperatures $T \sim T_c/2$	$T \longrightarrow T_c$	
Orthorhombic structure YBCO $T_c \approx 92$ K	$\Delta R_s(T) \propto T, \Delta X_s(T) \propto T$ at $T \leq T_c/4$; Essentially different slopes of $\Delta\lambda(T) \propto T$ [1–14]	Broad peak in $R_s(T)$ at $25 < T < 45$ K [4–14, 57–60] Peculiarities: 1. Shoulder [9, 11] in $R_s(T)$ at $T > 40$ K; 2. Bump [9] or plateau [8, 10] on the curves of $X_s(T)$ at $50 < T < 80$ K	Different slope of $\lambda(T)$ [3–14]	Normal skin-effect
Tetragonal structure BSCCO $T_c \approx 90$ K [19–22] TBCO $T_c \approx 80$ K [23, 24] TBCCO $T_c \approx 110$ K [10, 12]	$\Delta R_s(T) \propto T, T \leq T_c/2$ $\Delta X_s(T) = \omega\mu_0\Delta\lambda(T) \propto T, T \leq T_c/3$		Rapid growth of $R_s(T)$ and $X_s(T)$	$R_s(T) = X_s(T) = \sqrt{\omega\mu_0\rho(T)/2}$ $\Delta\rho(T) \propto T$

2.2. Complex Conductivity

Equations (2)–(4) allows us to express the real and imaginary parts of the complex conductivity $\sigma_s = \sigma_1 - i\sigma_2$ in terms of R_s and X_s as

$$\sigma_1 = \frac{2\omega\mu_0 R_s X_s}{(R_s^2 + X_s^2)^2}, \quad \sigma_2 = \frac{\omega\mu_0(X_s^2 - R_s^2)}{(R_s^2 + X_s^2)^2}. \qquad (6)$$

Above the superconducting transition temperature, the mean free path l of current carriers is shorter than the skin depth δ_n in the normal state (for $T \geq T_c$, $l \ll \delta_n$), which corresponds to the conditions of the normal skin effect. Equations (2)–(4) and (6) also apply to the normal state of HTSC's, where $R_n(T) = X_n(T) = \sqrt{\omega\mu_0/2\sigma_n(T)}$ with $\sigma_n \equiv \sigma_1$ ($T > T_c$) and $\sigma_2 \ll \sigma_1$ at microwave frequencies.

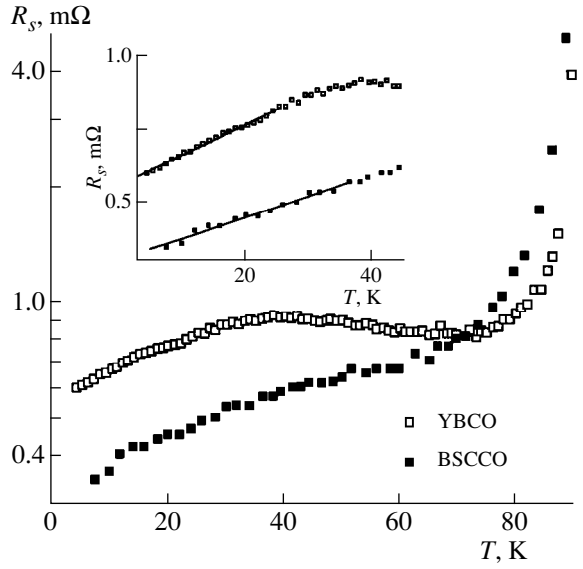

Fig. 2. Comparison of the temperature dependence of the surface resistance $R_s(T)$ of BSCCO and YBCO single crystals at 14.4 GHz. Experimental data are taken from [21] (BSCCO at 14.4 GHz) and [8] (YBCO at 9.4 GHz, scaled by ω^2 to 14.4 GHz). The inset shows the linear T-dependence of R_s at low T for both materials and a broad peak of $R_s(T)$ for YBCO.

The components $\sigma_1(T)$ and $\sigma_2(T)$ are not measured directly but can be derived from measurements of $R_s(T)$ and $X_s(T)$ using Eq. (6).

2.2.1. Low-temperature region ($T \ll T_c$). When $R_s(T) \ll X_s(T)$, Eq. (6) reduces to

$$\sigma_1(T) = \frac{2\omega\mu_0 R_s(T)}{X_s^3(T)}, \quad \sigma_2(T) = \frac{\omega\mu_0}{X_s^2(T)}. \qquad (7)$$

It then follows from Eq. (7) that for low and intermediate temperatures, $\sigma_1/\sigma_2 = 2R_s/X_s \ll 1$. The increments $\Delta\sigma_1(T)$ and $\Delta\sigma_2(T)$ depend on the increments $\Delta R_s(T)$ and $\Delta X_s(T)$ relative to the respective quantity:

$$\Delta\sigma_1 \propto \left(\frac{\Delta R_s}{R_s} - 3\frac{\Delta X_s}{X_s}\right), \quad \Delta\sigma_2 \propto -\frac{\Delta X_s}{X_s}. \qquad (8)$$

It follows from Eq. (8) that the dominant changes of $\sigma_2(T)$ are determined mainly by the function $X_s(T) = \omega\mu_0\lambda(T)$, which reflects the T-dependence of the magnetic field penetration depth.

The T-dependence of the real part of the conductivity, $\sigma_1(T)$, is determined by the competition between the increments $\Delta R_s/R_s$ and $\Delta X_s/X_s$.

In conventional superconductors, the quantity $X_s(T)$ ($\gg R_s$) is practically T-independent ($\Delta X_s \approx 0$) at temperatures $T \leq T_c/2$, and $R_s(T)$ decreases exponentially and approaches the residual surface resistance R_{res} as $T \longrightarrow 0$. By subtracting R_{res} from the measured $R_s(T)$, we obtain, using Eqs. (7) and (8), the temperature dependence of

$\sigma_1(T)$ predicted by the BCS theory: $\sigma_1 = 0$ at $T = 0$ and, for $T \le T_c/2$, $\Delta\sigma_1(T)$ shows an exponentially slow growth with an increasing temperature. We note that the smallest value of R_{res} detected in pure Nb is at least two orders of magnitude smaller than the smallest value of R_{res} measured in YBCO. The extremely small values of the surface resistance in Eq. (8) indicate that the increment $\Delta\sigma_1(T)$ is always positive in classical superconductors ($\Delta\sigma_1(T) > 0$), at least in the temperature interval $T < 0.8\ T_c$, before the maximum of the BCS coherence peak is reached.

For HTSC single crystals, the T-dependence of $\Delta\sigma_1(T)$ is radically different from that predicted by the theories of the microwave response of conventional superconductors. For $T < T_c$, the increments $\Delta R_s(T)$ and $\Delta X_s(T)$ in HTSC's are not small; in addition, $\Delta X_s(T) \gg \Delta R_s(T)$. Although $R_s(T) < X_s(T)$, $\Delta R_s/R_s$ is not necessarily greater than $3\Delta X_s/X_s$ in Eq. (8) or positive at all temperatures. When that occurs, $\sigma_1(T)$ increases with a decreasing temperature. The function $\sigma_1(T)$ is maximum at some $T = T_{max}$, and then $\sigma_1(T)$ becomes smaller with a decreasing temperature. $\sigma_1(T)$ has a peak if the value of R_{res} is sufficiently small as $T \longrightarrow 0$:

$$R_{res} < \frac{X_s(0)}{3}\frac{\Delta R_s(T)}{\Delta X_s(T)}. \qquad (9)$$

If inequality (9) is satisfied, T_{max} is a finite temperature, while for R_{res} being equal to the right-hand side of (9), T_{max} shifts to 0 K. If R_{res} is such that (9) is not satisfied, $\sigma_1(T)$ decreases at low temperatures as the temperature is increased, which is quite different from that observed in conventional superconductors.

Thus, the shape of $\sigma_1(T)$ for $T \ll T_c$ depends on the value of the residual surface resistance R_{res}, whose origin and accurate value are unknown. For this reason, the shapes of the $\sigma_1(T)$ curves are not determined unambiguously for $T \le T_c/2$, unlike the functions $R_s(T)$ and $X_s(T)$, which are directly measured in experiments.

If we linearly extrapolate $R_s(T)$ to $T = 0$ and attribute the resulting value $R_s(0)$ to the residual surface resistance ($R_s(0) = R_{res}$) and then substitute the temperature-dependent difference $R_s(T)$–R_{res} into the numerator of the first expression in Eq. (7), the result is that the $\sigma_1(T)$ curve has a broad peak for HTSC materials. Near $T = 0$, $\sigma_1(T)$ increases linearly with T from zero, reaches a maximum at T_{max}, and then decreases to $\sigma(T_c)$. This procedure, however, ignores the possibility of intrinsic residual losses. Therefore, some authors (see, e.g., [14, 21, 59]) associate residual losses in HTSC single crystals with a residual normal electron fluid. This implies that the source of the residual loss is in the bulk of the sample, although it is probably not intrinsic. If this contribution is excluded from the complex conductivity of the superconductor, one obtains $\sigma_1(T = 0) \longrightarrow 0$, as can be seen in Fig. 3 from the measurements taken at 13.4, 22.7, and 75.3 GHz in [14]. The peak of $\sigma_1(T)$ shifts to higher temperatures, and its size diminishes as the experimental frequency is increased. In YBCO single crystals, the temperature T_{max} at which the maximum of σ_1 occurs is close to the temperature at which the peak of $R_s(T)$ occurs.

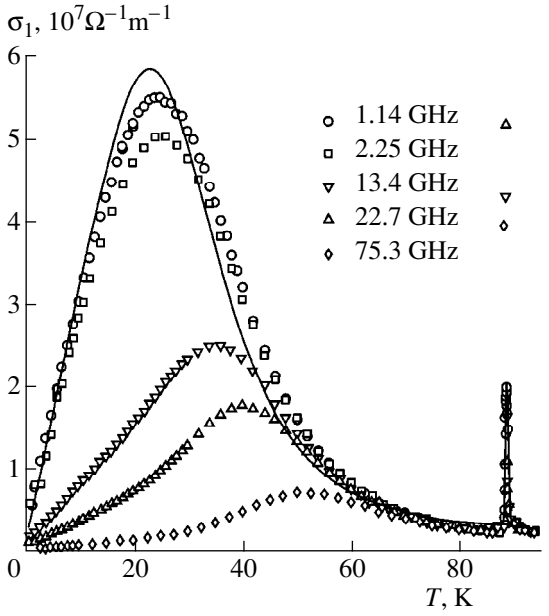

Fig. 3. Real part of the conductivity $\sigma_1(T)$ of YBCO single crystal at different frequencies [14]. The data (symbols) are the courtesy of the Vancouver group (D.A. Bonn). Solid line is the T-dependence of $\sigma_1(T)$ at 1.14 GHz calculated using the modified two-fluid model and taking the inhomogeneous broadening of the superconducting transition into account ($\delta T_c = 0.4$ K in Eq. (21), see Section 3.3).

Finally, one can procure $\sigma_1(T)$ from measurements of $R_s(T)$ and $X_s(T)$ for $T > 0$ without any concern about R_{res}. In this case, $\sigma_1(0)$ is not determined uniquely. Whether $\sigma_1(T)$ has a peak depends on the validity of condition (9). The curves at 1 and 2 GHz in Fig. 3 were obtained using Eq. (6) without subtracting any residual losses.

2.2.2. Temperatures close to T_c ($T \longrightarrow T_c$). Equations (7) and (8) do not apply near T_c. In this temperature range, it is necessary to use the general local relations (2)–(4) and (6).

The conductivity $\sigma_2(T)$ in the *ab*-plane of HTSC crystals abruptly drops to very small values in the normal state. The expression $[T_c/\sigma_2(0)]d\sigma_2(T)/dT$ at $T = T_c$ that defines the slope of $\lambda^2(0)/\lambda^2(T)$ at $T = T_c$ varies between –2 and –4 for different crystals.

The real part of the conductivity, $\sigma_1(T)$, does not show a coherence peak near $T = 0.85T_c$, as predicted by BCS. Usually, the real part of the conductivity of HTSC single crystals has a narrow peak near T_c, which increases with decreasing the frequency [21, 23, 24].

The width of the narrow peak of $\sigma_1(T)$ coincides with the width of the $R_s(T)$ transition at microwave frequencies. A possible explanation of the sharp peak just below T_c is an inhomogeneous broadening of the superconducting transition [77–79] or the fluctuation effects [24, 80, 81].

3. MODIFIED TWO-FLUID MODEL

As was shown in [65], high T_c values ($T_c \sim 100$ K), the temperature dependence of the resistivity, the frequency dependence of the momentum relaxation time, and other properties of the normal state in optimally doped HTSC's are well described within the framework of the Fermi-liquid approach involving strong electron–phonon coupling (SC) [63]. The SC model also explains some of the features of the superconducting state of HTSC's. It follows from the Eliashberg theory that the distinctive property of superconductors with strong coupling is that the gap in the spectrum of electronic excitations is smeared. Strictly speaking, there is no gap whatsoever at $T \neq 0$ [82, 83]. This leads to breaking of Cooper pairs, smearing of the peak in the density of states at $\hbar\omega = \Delta(T)$ due to the inelastic scattering of electrons by thermally excited phonons, and suppression of coherence effects. As a result, the amplitude of the coherence peak decreases and, according to [84, 85], virtually disappears at frequencies around 10 GHz if the electron–phonon coupling constant exceeds unity. Moreover, the quasiparticle generation mechanism is radically different from that of the BCS model. The quasiparticles are generated without jumps across the energy gap and can be in states with all energies down to $\hbar\omega = 0$. These states can be classified as gapless, and the quasiparticles can be treated [65] as normal current carriers in the two-fluid model. Thus, it is not surprising that an important consequence of the SC model is the nonexponential behavior of $R_s(T)$ [86] and $\lambda(T)$ [87]. Power-law temperature dependences were also predicted by the two-fluid Gorter–Casimir (GC) model [76]; near T_c, they proved to be quite close to the results of calculations performed in the SC model. In particular, the curves $\lambda^2(0)/\lambda^2(T)$ calculated by the SC model [88–91] proved to be sufficiently close to the function $n_s(t)/n = 1 - n_n(t)/n = 1 - t^4$ in the GC model. At $T = T_c$, the slopes of these curves are in agreement with those measured with different YBCO single crystals and are equal to –3 [4] or –4 [5, 8, 10]. The experimental fact that there is no BCS coherence peak in the conductivity of HTSC crystals indicates the necessity of taking the strong coupling effects near T_c into account and the feasibility of interpreting HTSC properties at microwave frequencies in terms of a two-fluid model.

Complex conductivity σ_s is a basic property of superconductors. In accordance with the GC model [76], the expressions for the components of $\sigma_s = \sigma_1 - i\sigma_2$ are

$$\sigma_1 = \frac{n_n e^2 \tau}{m}\left[\frac{1}{1+(\omega\tau)^2}\right],$$
$$\sigma_2 = \frac{n_s e^2}{m\omega}\left[1 + \frac{n_n}{n_s}\frac{(\omega\tau)^2}{1+(\omega\tau)^2}\right]. \quad (10)$$

At temperatures $T \leq T_c$, the total carrier concentration is $n = n_s + n_n$, where $n_{s,n}$ are the fractions of the superconducting and normal carrier densities (both have the same charge e and effective mass m). The real part σ_1 is determined solely by the normal component, while both components (normal and superconducting) contribute to the imaginary part σ_2. In the GC model, the relaxation time τ of normal carriers is independent of the temperature. This is acceptable if we assume that the behavior of normal carriers in superconductors is similar to that of normal carriers in normal metals at low temperatures. Scattering of electrons at very low temperatures is due to impurities and is independent of the temperature. Therefore, the temperature dependence of the real part of the conductivity (10) in the GC model is determined entirely by the function $n_n(T)$ with $n_s(T) = n - n_n(T)$ only.

For sufficiently low frequencies $(\omega\tau)^2 \ll 1$, the expressions of the conductivity components in Eq. (10) transform into simple relations

$$\sigma_1 = \frac{e^2\tau}{m}n_n, \quad \sigma_2 = \frac{e^2}{m\omega}n_s = \frac{1}{\mu_0\omega\lambda^2}, \quad (11)$$

where $\lambda = \sqrt{m/\mu_0 n_s e^2}$ is the London penetration depth of a static magnetic field.

Penetration of alternating fields into superconductors is controlled by the frequency-dependent skin depth. Using complex conductivity (11), one obtains the complex skin depth δ_s by generalizing the corresponding expression for a normal conductor:

$$\delta_s = \frac{\sqrt{2}\lambda}{\sqrt{\omega\tau(n_n/n_s) - i}}. \quad (12)$$

With an increasing angular frequency ω, the skin depth $\mathrm{Re}\,\delta_s$ decreases; therefore, the London penetration depth λ gives the upper bound for the penetration of the electromagnetic field into a superconductor. In the GC model, the λ value diverges near T_c as $\lambda(t) = \lambda/(2\sqrt{1-t})$ and the function $\sigma_2(t)/\sigma_2(0) = 4(1-t)$ tends linearly to zero at $T = T_c$ with a slope equal to –4. At the same time, at $T = T_c$, the skin depth $\mathrm{Re}\,\delta_s$, defined by Eq. (12), crosses over to the skin depth δ_n for a normal conductor.

3.1. Scattering and Surface Resistance of HTSC Single Crystals

In conventional superconductivity, one assumes that the mean free path does not vary with the temperature below T_c. In a normal metal at much higher temperatures than the corresponding T_c of a conventional superconductor, the electron scattering rate is proportional to T [92]. Since the transition temperatures of HTSC's are much larger than those of conventional superconductors, it stands to reason that temperature affects the electron scattering rate of the quasiparticles of HTSC's below T_c but is limited to a constant rate at low temperatures. Therefore, if a two-fluid model is to be successful in explaining transport properties of HTSC's, it is natural to include a temperature variation of τ into that model.

To obtain an expression for $\tau(T)$, we rely on the analogy between the "normal fluid" component in the superconducting state and charge carriers in a normal metal. According to Mathissen's rule, the reciprocal relaxation time at temperatures below the Debye temperature Θ is

$$\frac{1}{\tau} = \frac{1}{\tau_{imp}} + \frac{1}{\tau_{e-ph}}. \tag{13}$$

The first term on the right-hand side is due to impurity scattering and is independent of temperature, and the second is due to the electron–phonon scattering and is proportional to T^5.

From Eq. (13), we express $\tau(T)$ as

$$\frac{1}{\tau(t)} = \frac{1}{\tau(T_c)}\frac{\beta + t^5}{\beta + 1}, \tag{14}$$

where β is a numerical parameter ($\beta \approx \tau(T_c)/\tau(0)$), provided this ratio is much less than unity. It must be pointed out, however, that this approximation is not always satisfied.

Equation (14) corresponds to the low-temperature limit of the Bloch–Grüneisen formula, which includes impurity scattering and can be presented over a wide temperature range by the expression

$$\frac{1}{\tau(t)} = \frac{1}{\tau(T_c)}\frac{\beta + t^5 \mathcal{J}_5(\kappa/t)\mathcal{J}_5(\kappa)}{\beta + 1},$$
$$\mathcal{J}_5(\kappa/t) = \int_0^{\kappa/t} \frac{z^5 e^z dz}{(e^z - 1)^2}, \tag{15}$$

where $\kappa = \Theta/T_c$. For $T < \Theta/10$ ($\kappa > 10t$), Eq. (15) approaches the form of Eq. (14). For $T > \Theta/5$ ($\kappa < 5t$), we obtain from Eq. (15) the linear T-dependence $1/\tau(t) \propto t$. Examples of $1/\tau(t)$ for different parameters of β, κ, and $\tau(T_c)$ are shown in Fig. 4.

For $\omega\tau(T_c) \ll 1$, which is normally satisfied at microwave frequencies in HTSC's, the parameter

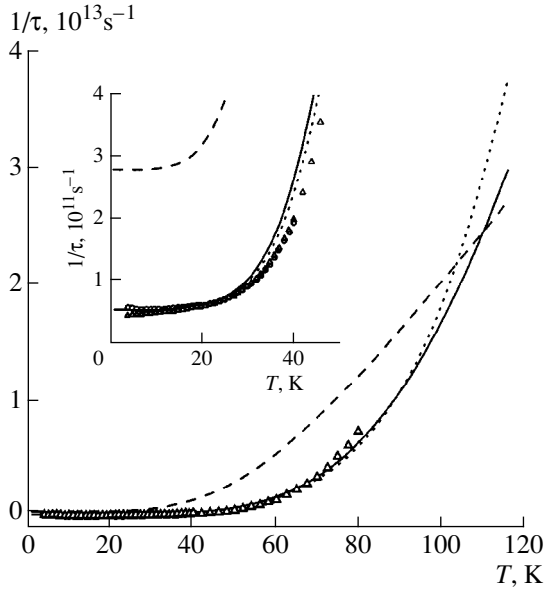

Fig. 4. Scattering rate of quasiparticles calculated form Eq. (14), dotted line: $\beta = 0.005$, and Eq. (15), solid line: $\beta = 0.005$, $\kappa = 9$; dashed line: $\beta = 0.02$, $\kappa = 4$. The triangles are calculated from $1/\tau = [1 - \lambda^2(0)/\lambda^2(T)]/[\mu_0\sigma_1(T)\lambda^2(0)]$, with $\sigma_1(T)$ and $\lambda(T)$ at 1.14 GHz and $\lambda(0) = 1600$ Å in the *ab*-plane, with the currents parallel to the *a*-direction of the YBCO crystal [14]. The inset shows the low-temperature parts of the curves. The circles are from Fig. 8 of [14].

$\omega\tau(T_c)$ is obtained from measurements of $R_s(T_c)$ and $X_s(0)$:

$$\omega\tau(T_c) = \frac{X_s^2(0)}{2R_s^2(T_c)} = \frac{\sigma_1(T_c)}{\sigma_2(0)}. \tag{16}$$

At frequencies ≈10 GHz, the value of $\omega\tau$ for the best HTSC crystals is of the order of 10^{-3} at $T = T_c$ and remains less than unity at all temperatures $T < T_c$, as is discussed in what follows. In the two-fluid model, therefore, the expressions of the conductivity components in Eq. (10) turn into the simple form (11).

All experimental data of $R_s(T)$ for high-quality HTSC single crystals can be elucidated by our two-fluid model with $\tau(T)$ given by Eqs. (14) or (15).

Measurements of $R_s(T)$ of YBCO single crystals at frequencies of order or less than 10 GHz are analyzed first. The values of $\sigma_2(T)/\sigma_2(0) = \lambda^2(0)/\lambda^2(T) = n_s(T)/n$ measured in the same experiments and those of $\sigma_1(T)/\sigma(T_c)$ obtained from Eq. (11) are substituted into Eq. (3). We then use the relation $n_n(T)/n = 1 - \sigma_2(T)/\sigma_2(0)$, which is obtained from the experimental data, and take $\tau(T)$ from Eqs. (14) or (15).

Setting $\beta = 0.005$ and $\kappa = 9$ in Eq. (15) and taking the experimental values $\sigma_2(T)/\sigma_2(0)$ from Fig. 11 (see below) and $\omega\tau(T_c) = 7.5 \times 10^{-4}$ at 1.14 GHz, we find from Eqs. (11) and (3) the T-dependence of $R_s(T)$, shown by the curves in Fig. 5. These curves match the

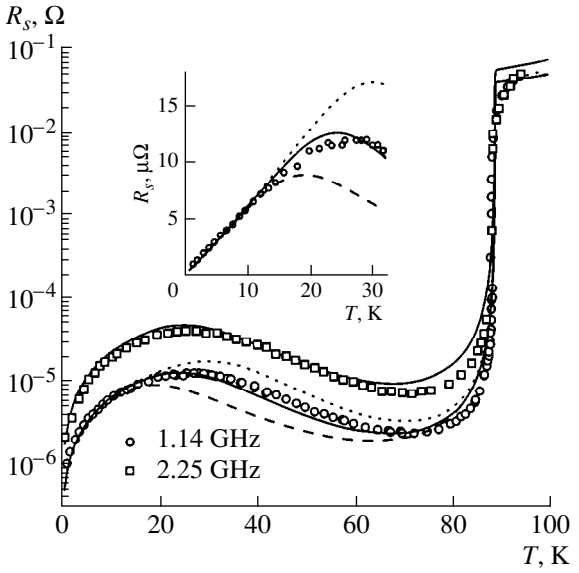

Fig. 5. Experimental $R_s(T)$ data of YBCO single crystal [14] at 1.14 GHz (circles) and 2.25 GHz (squares). Solid curves are calculations using Eqs. (3), (11), and (14). The dashed curves are calculated at 1.14 GHz with the term t^5 replaced by t^4 in the numerator of Eq. (14); the dotted curves, with t^6. The inset shows a linear plot of $R_s(T)$ at low temperatures at 1.14 GHz.

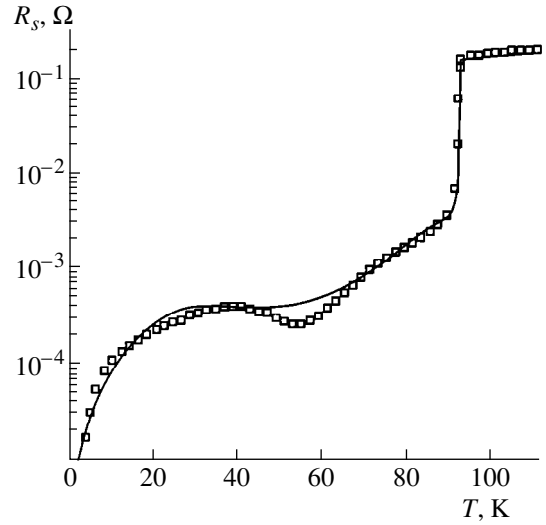

Fig. 6. Comparison between the calculated (solid line) and measured (squares) surface resistance $R_s(T)$ of YBCO single crystal at 10 GHz. Experimental data are from [9].

data of [14] over the entire temperature range. The same result is obtained using Eq. (14) instead of Eq. (15), with $\beta = 0.005$. For $\kappa \gg 1$ and $T \lesssim T_c$, Eqs. (14) and (15) are identical.

It follows from Eqs. (5) and (11) that for $\alpha t \ll 1$ [see Eq. (1)], a rough estimate of the temperature t_m at which $R_s(T)$ is maximum is obtained from the relation $\beta \simeq 4t_m^5$. The value of $\tau(0)$ is found from the slopes dR_s/dT and dX_s/dT of the experimental data of $R_s(T)$ and $X_s(T)$ as $T \longrightarrow 0$ ($\omega\tau(0) < 1$):

$$\omega\tau(0) = \left.\frac{dR_s}{dX_s}\right|_{T \to 0}. \qquad (17)$$

With Eqs. (16) and (17), the parameter $\beta \approx \tau(T_c)/\tau(0)$ is determined from the surface impedance data. As β increases, the maximum and minimum of $R_s(T)$ change into an inflection point with horizontal tangent; for larger β, the maximum of $R_s(T)$ disappears completely [74].

The linear growth of $R_s(T)$ with T at low temperatures (inset in Fig. 5) is a direct consequence of the linear change of $\lambda(T)$ near $T = 0$, which is proportional to the coefficient α in Eq. (1), and is the result of a constant scattering rate at low temperatures, as shown in Fig. 4.

The dashed and dotted curves shown in Fig. 5 are the calculated $R_s(T)$ values at 1.14 GHz, with t^5 replaced by t^4 (dashed curve) and by t^6 (dotted curve) in Eq. (14). The best fit of the experimental data is $1/\tau(t) \propto t^5$. Moreover, Eq. (15) enables us to incorporate the shoulder of $R_s(T)$ obtained with YBCO single crystals in [9, 11]. This is shown in Fig. 6, which contains measurements (squares) of $R_s(T)$ at 10 GHz taken from [9] and calculations (solid line) of $R_s(T)$ based on Eqs. (11) and (3) with $\omega\tau(T_c) = 4 \times 10^{-3}$, $\sigma_2(T)/\sigma_2(0)$ obtained from the same experimental data [9], $\beta = 0.02$, and $\kappa = 4$ in Eq. (15).

The calculated curves in Figs. 5 and 6 are very close to the experimental data and display the following common and unique features of $R_s(T)$ for $T < T_c$ and $\omega\tau < 1$ of high-quality YBCO single crystals fabricated by different methods: (i) the linear temperature dependence of the surface resistance, $\Delta R_s(T) \propto T$, caused by the linear variation of $\Delta X_s(T) \propto \Delta\lambda_{ab}(T) \propto T$ at temperatures $T \ll T_c$ and by the limit $\tau(T) \longrightarrow$ const at low temperatures; (ii) the broad peak of $R_s(T)$ in the intermediate temperature range due to the rapid decrease of the relaxation time $\tau(T) \propto T^{-5}$ with an increasing temperature; and (iii) the increase in $R_s(T)$ in the range $T_c/2 < T < T_c$ (Fig. 6) caused by the crossover from T^{-5} to T^{-1} of $\tau(T)$ in Eq. (15), which occurs in Fig. 6 at a lower temperature than in Fig. 5. The behavior of $1/\tau(T)$ for these two YBCO crystals is shown in Fig. 4.

Up to this point, we did not take the residual surface resistance R_{res} of the samples into account. In the YBCO crystals, whose data are plotted in Figs. 5 and 6 scaled to the same frequency (10 GHz), the resistance is $R_{res} < 50$ μΩ. $R_{res}/R(T_c) < 10^{-3}$ is so small that R_{res} was neglected even at $T \ll T_c$. In most HTSC crystals that were investigated, however, $R_{res}/R(T_c) > 10^{-3}$ (see, e.g., Figs. 1 and 2). Therefore, it is important that R_{res} is

added to the calculated $R_s(T)$ values when comparing the latter with the experimental data.

Figure 7 compares the measured $R_s(T)$ and $X_s(T)$ of BSCCO, plotted in Fig. 1, with the results of calculations based on Eqs. (3) and (4). In this case, we added the constant $R_{res} = 0.5$ mΩ to the calculated values of $R_s(T)$. The calculations are based on measurements of $\sigma_2(T)/\sigma_2(0)$ obtained in the same experiment and are plotted in the inset to Fig. 13 (see below), with the parameters $\omega\tau(T_c) = 0.9 \times 10^{-2}$, $\beta = 2$ and $\kappa = 3$ in Eq. (15). It is clear that the agreement between the calculated and experimental curves is good throughout the temperature interval $5 \le T \le 120$ K.

Another reason for including R_{res} is the ratio $R_{res}/R(T_c) \propto \omega^{3/2}$. Figure 8 is based on the experimental data of BSCCO single crystal measured in [21] at three frequencies: 14.4 ($\omega\tau(T_c) = 0.7 \times 10^{-2}$), 24.6, and 34.7 GHz. The solid curves are the calculations at these frequencies obtained from Eqs. (11) and (3) using $\tau(T)$ from Eq. (15) with $\beta = 0.1$ and $\kappa = 4$. The comparison procedure is different from that discussed above for YBCO crystals because $R_{res} \propto \omega^2$ is added to the calculated $R_s(T)$ values. The inset of Fig. 8 shows a linear plot of the measured and calculated surface resistance at low temperatures. We emphasize that at temperatures below $T_c/2$, the value of $\Delta R_s(T)$ is proportional to T.

In the millimeter and shorter wavelength bands, the condition $\omega\tau < 1$ may not be satisfied in the superconducting state of the purest HTSC single crystals because of the fast growth of $\tau(T)$ with decreasing $T < T_c$. In analyzing the experimental data of $Z_s(T)$ and $\sigma_s(T)$, therefore, it is natural to not only take R_{res} into account but also to use the more general Eq. (10) of the two-fluid model to replace Eq. (11). The $R_s(T)$ data of [14] at the frequencies of 13.4, 22.7, and 75.3 GHz are shown in Fig. 9 with the calculated $R_s(T)$ values (obtained on the same YBCO crystal that was used in Fig. 5). We used $\tau(T_c)/\tau(0) \approx \beta = 5 \times 10^{-3}$ in Eq. (14) for all curves shown in Fig. 9 (the same as previously used in Fig. 5) and added $R_{res} = 0.3$ mΩ to $R_s(T)$ [Eq. (3)] at 75.3 GHz only. The conductivity components $\sigma_1(T)$ and $\sigma_2(T)$ involved in Eq. (3) are obtained from the experimental data of $\sigma_2(T)/\sigma_2(0)$ at 1.14 GHz [14] (shown in Fig. 11) and from Eq. (10).

Figure 10 shows another example. The experimental $R_s(T)$ data (squares) of a TBCO single crystal ($T_c = 78.5$ K) [23] are compared with the results of calculations based on Eqs. (3), (10), and (15). The curve representing the theoretical values $R_s(T) + R_{res}$ is plotted using $\beta = 0.1$, $\kappa = 5.5$, $\omega\tau(T_c) = 1.7 \times 10^{-2}$, $R_{res} = 0.8$ mΩ, and with $\sigma_2(T)/\sigma_2(0)$ shown in the inset (circles) of Fig. 10.

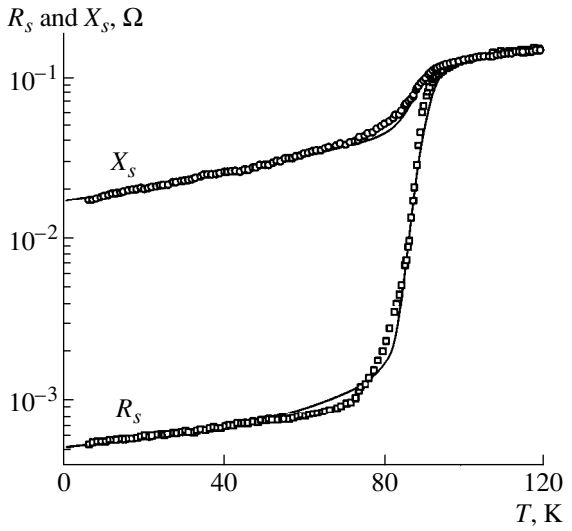

Fig. 7. Comparison between the calculated (solid lines) and measured surface impedance (symbols) of BSCCO single crystal (see Fig. 1). A constant $R_{res} = 0.5$ mΩ is added to the values of $R_s(T)$obtained from Eq. (3).

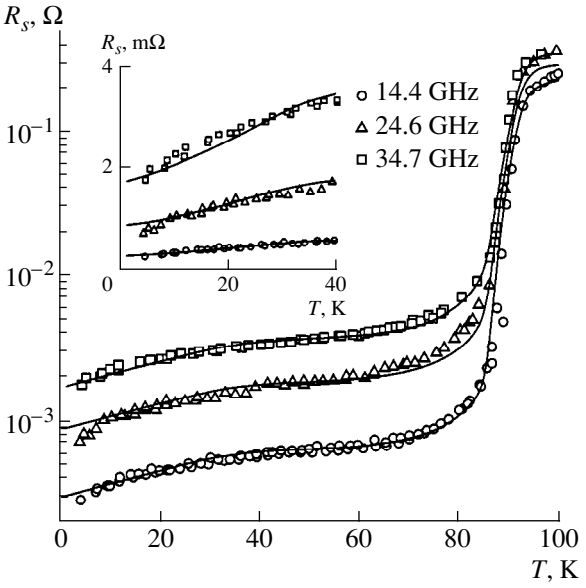

Fig. 8. Experimental data of the BSCCO single crystal [21] at various frequencies: 14.4 GHz (circles), 24.6 GHz (triangles), and 34.7 GHz (squares). The solid curves are the calculated $[R_s(T) + R_{res}]$-functions, with the respective R_{res} value of 0.29, 0.85 and 1.7 mΩ. The inset shows the linear temperature dependences of the surface resistance at low temperatures.

3.2. Temperature Dependence of the Superconducting Electron Density

In the previous section, we accentuated that the modified two-fluid model describes all features of the surface resistance $R_s(T)$ of different HTSC's over a wide frequency range with only one parameter (k). This

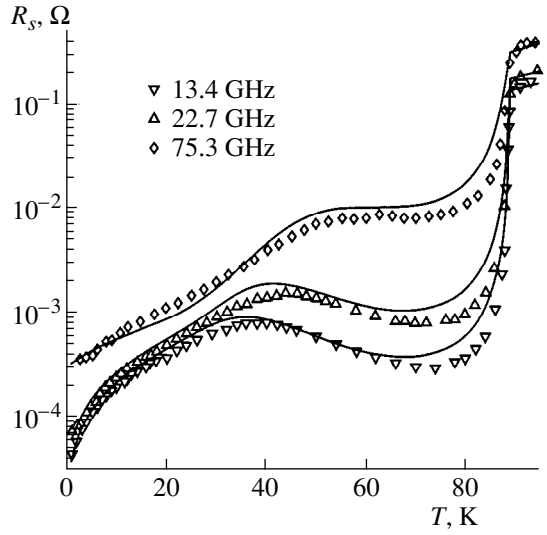

Fig. 9. Comparison between the calculated (lines) and measured [14] (symbols) surface resistance $R_s(T)$ of the YBCO single crystal at 13.4, 22.7, and 75.3 GHz. We assumed $R_{res} = 0.3$ mΩ for 75.3 GHz, zero for the other frequencies.

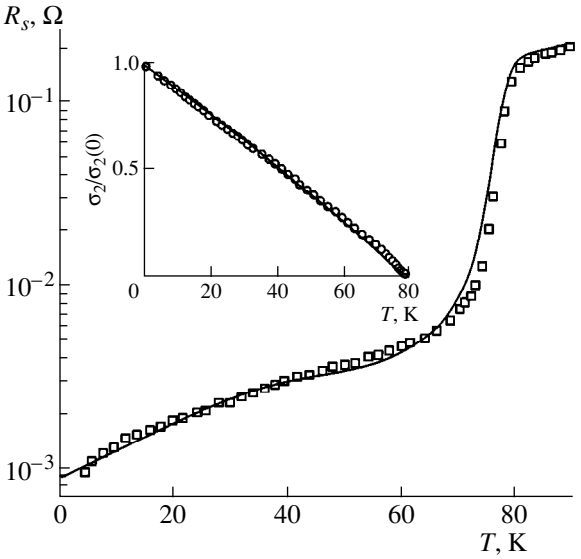

Fig. 10. Surface resistance $R_s(T)$ of a TBCO single crystal at 24.8 GHz taken from [23]. Solid curve is the calculated $[R_s(T) + R_{res}]$-function with $R_{res} = 0.8$ mΩ. The inset shows the measured [23] (circles) and calculated results of $\sigma_2(T)/\sigma_2(0)$ (solid line) using Eq. (18) with $\alpha = 0.9$.

was done using the measured (known from the same experiment) T-dependences of the superconducting electron density $\lambda_{ab}^{-2}(T)$. However, we think that our phenomenological model would be incomplete unless simple formulas are available that correctly describe the measurements of $\Delta\lambda_{ab}(T)$. Figures 10 (the inset), 11, and 12 show $\sigma_2(T)/\sigma_2(0) = \lambda^2(0)/\lambda^2(T) = n_s(T)/n$ in the *ab*-plane of TBCO, YBCO, and BSCCO single crystals from [23, 14, 21], respectively. All of these quantities change linearly with temperature at low temperatures and can be approximated by the function [73]

$$n_s/n = (1-t)^{\alpha}, \qquad (18)$$

where α is a numerical parameter. For $t \ll 1$, Eq. (1) follows from Eq. (18). For the cited experiments, the values of α fall into the range $0.4 < \alpha \le 0.9$. Near T_c, we obtain $\lambda(t) \propto n_s(t)^{-1/2} \propto (1-t)^{-\alpha/2}$, which is also in reasonably good agreement with the experimental data. However, Eq. (18) yields an infinite value of the derivative $d\sigma_2(t)/dt \propto (1-t)^{\alpha-1}$ at $t = 1$ for $\alpha < 1$.

An approximation for $n_s(t)/n$ proposed in [75] is close to Eq. (18),

$$n_s/n = 1 - \alpha t - (1-\alpha)t^6, \qquad (19)$$

and is shown by solid lines in Figs. 11 and 12. Equation (19) insures that the slope of $\lambda^2(0)/\lambda^2(t)\big|_{T_c} = 5\alpha - 6$ at T_c is finite and negative for $\alpha < 1.2$.

However, the above functions for $n_s(t)$ in their simplest forms (18) and (19) do not account for all features in $\lambda^2(0)/\lambda^2(T)$ detected recently in YBCO crystals (see table) in the intermediate temperature range [8–11]. Moreover, the slope of these curves at $T \ll T_c$ requires that $\alpha > 1$ in Eq. (18), which would lead to a zero slope of the $\sigma_2(T)/\sigma_2(0)$ curve $T = T_c$. Therefore, we have added an additional empirical term to the right-hand

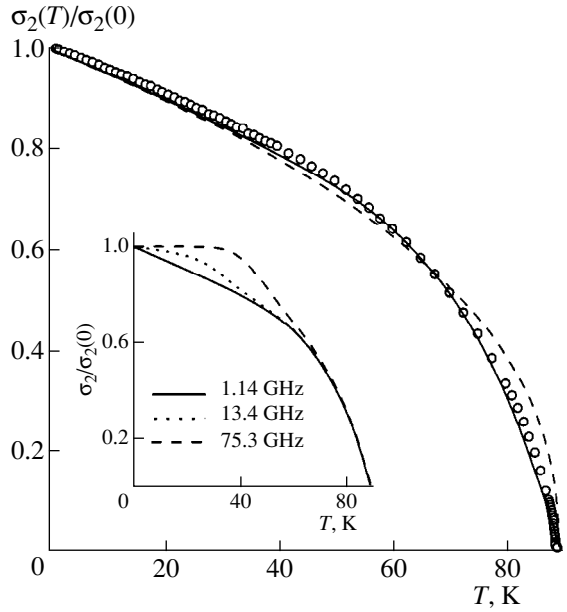

Fig. 11. Plots of Eq. (18) (dashed line, $\alpha = 0.42$) and Eq. (19) (solid line, $\alpha = 0.47$), showing the fit to the empirical $\sigma_2(T)/\sigma_2(0)$. The experimental data (circles) are from [14] at 1.14 GHz. The inset shows the temperature dependences of $\sigma_2(T)/\sigma_2(0)$ at various frequencies calculated from Eqs. (10), (19) and (14).

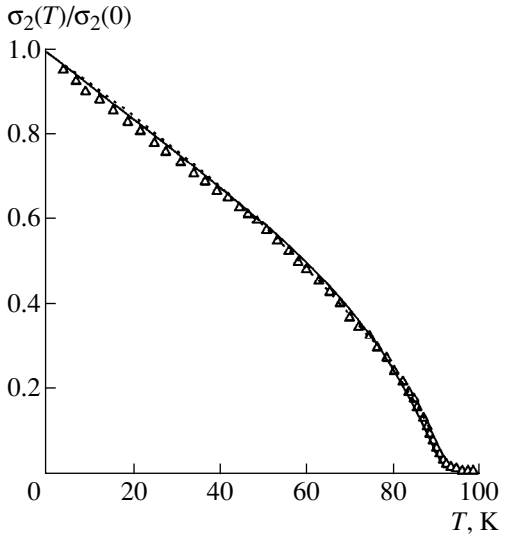

Fig. 12. Comparison between the calculated (solid curve: Eq. (19), $\alpha = 0.74$; dotted line: Eq. (18), $\alpha = 0.7$) and measured [21] (symbols) of the $\sigma_2(T)/\sigma_2(0)$ values of BSCCO single crystal [75].

side of Eq. (18) without violating the particle conservation condition $n_s + n_n = n$:

$$n_s/n = (1-t)^{\alpha}(1-\delta) + \delta(1-t^{4/\delta}), \qquad (20)$$

where $0 < \delta < 1$ is the weight factor. For $\delta \ll 1$ and $\alpha > 1$, the dominant contribution to $\sigma_2(T)$ throughout the relevant temperature range is still due to the first term on the right-hand side of Eq. (20), while the second is responsible for the finite slope of $\sigma_2(T)/\sigma_2(0)$ at $T = T_c$, which is equal to –4 in accordance with the GC model. As δ increases, the second term on the right-hand side of Eq. (20) becomes more essential. The experimental curve of $\sigma_2(T)/\sigma_2(0)$ derived from the $R_s(T)$ and $X_s(T)$ measurements of the YBCO crystal in [8] is properly described by Eq. (20) with $\delta = 0.5$ and $\alpha = 5.5$ (Fig. 13). This calculation reflects the characteristic features of the experimental data, namely, the linear section of n_s and the positive second derivative ($\alpha > 1$) in the low-temperature region, the plateau in the intermediate temperature range, and the correct value of the slope near T_c.

Using Eq. (20) with $\alpha = 2$ and $\delta = 0.2$, we can also describe the T-dependence of $\sigma_2(T)$ of BSCCO crystals (Figs. 1 and 7), plotted in the inset to Fig. 13.

3.3. The Real Part of the Conductivity

Since the measurements and calculations of $R_s(T)$, $X_s(T)$, and $\sigma_2(T)$ are in good agreement and consistent with $\sigma_1(T)$ in the range $T < T_c$, the modified two-fluid model can be a powerful tool in describing the electrodynamic properties of HTSC's. The only feature that has not been investigated by this model is the behavior of $Z_s(T)$ and $\sigma_s(T)$ in the temperature range near T_c. A spectacular display is the narrow peak in the real part of the conductivity (see Fig. 3).

The narrow peak near T_c can be described by an effective medium model [79, 93] that takes the inhomogeneous broadening of the superconducting transition into account. We assume that different regions of a given specimen experience transitions to the superconducting state at different temperatures within the T-range δT_c. If the dimension of each of these regions is smaller than

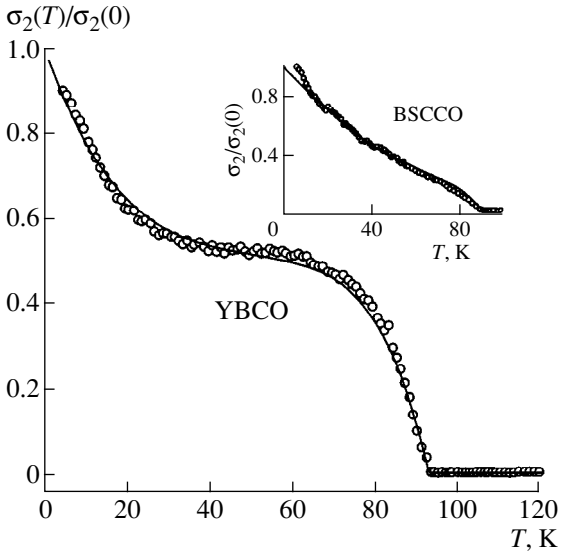

Fig. 13. Comparison between the calculated (solid line) and measured (circles) values of $\sigma_2(T)/\sigma_2(0)$ of the YBCO single crystal [8]. The inset shows the measured and calculated values with Eq. (20) used for the temperature dependences of $\sigma_2(T)/\sigma_2(0)$ of the BSCCO crystal shown in Fig. 1.

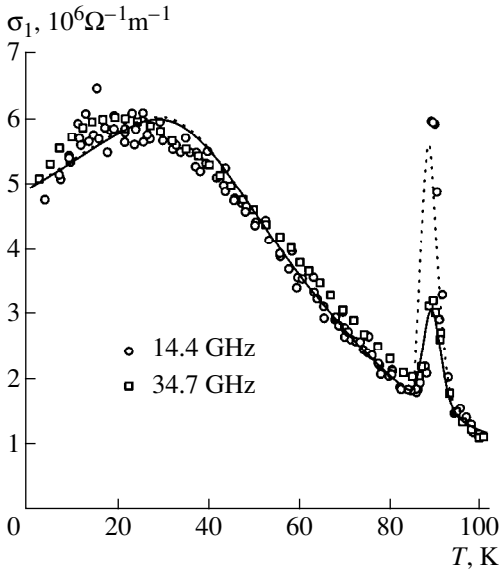

Fig. 14. Experimental data of $\sigma_1(T)$ at 14.4 and 34.7 GHz of the BSCCO single crystal [21] (symbols) and the calculations of $\sigma_1(T)$ (lines) using Eqs. (14), (21), and (6), with sample inhomogeneities taken into account ($\delta T_c = 2$ K).

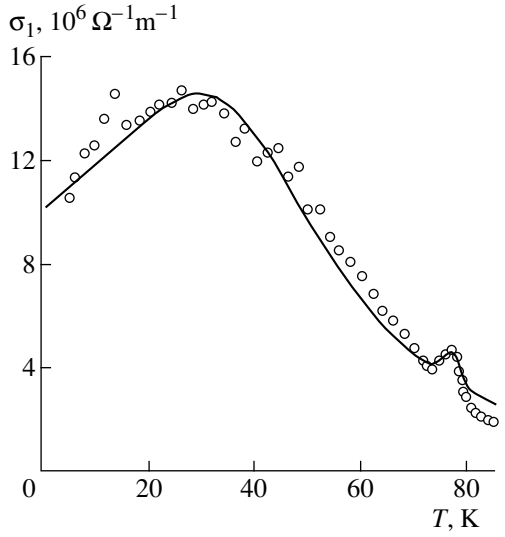

Fig. 15. Comparison of the experimental T-dependence of $\sigma_1(T)$ (open circles) of the TBCO single crystal at 24.8 GHz [23] with the one calculated using the modified two-fluid model (solid line) and taking the inhomogeneous broadening of the superconducting transition into account ($\delta T_c = 2.5$ K in Eq. (21)).

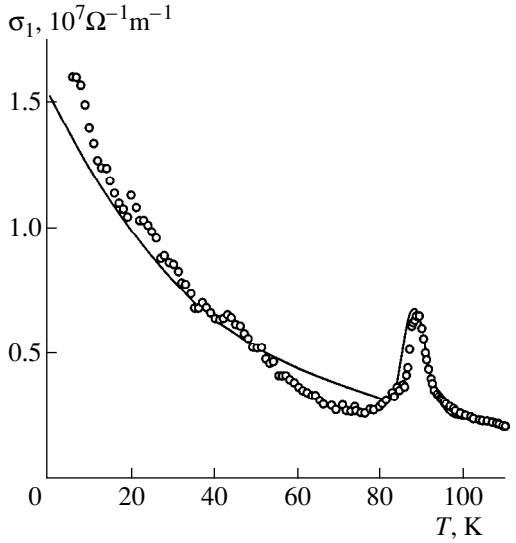

Fig. 16. The conductivity $\sigma_1(T)$ of the BSCCO single crystal at 9.4 GHz extracted from the surface impedance measurements of Fig. 1 and the calculation based on the modified two-fluid model, which takes the inhomogeneous broadening of the superconducting transition into account ($\delta T_c = 4.5$ K). $\sigma_1(T)$ does not have a broad peak at low temperatures in this particular case.

the magnetic field penetration depth (microscopic-scale disorder), the distribution of the microwave currents over the sample is uniform and the calculation of the effective impedance Z_{eff} of the sample reduces to two operations: adding the impedances Z_s of all regions in the specimen (with different T_c) that are connected in series along a current path and averaging over the sample volume. As the result, we obtain

$$Z_s^{\text{eff}}(T) = R_s^{\text{eff}}(T) + iX_s^{\text{eff}}(T) = \int_{\delta T_c} Z_s(T, T_c) f(T_c) dT_c, \tag{21}$$

where the distribution function $f(T_c)$ is such that the fraction of the sample volume with critical temperatures in the range $T_c < T < T_c + dT_c$ equals $f(T_c)dT_c$. In the simplest case, $f(T_c)$ is a Gaussian function. In the experiments of [14], the width of the superconducting resistive transition was approximately 0.4 K, which we equate to the width of the Gaussian distribution $f(T_c)$. Using general relations (6) with the effective impedance components obtained from Eq. (21), we calculate $\sigma_1^{\text{eff}}(T)$ near T_c and plot it together with the experimental data for YBCO in Fig. 3. The overall agreement is good.

In the framework of the discussed approach, $\sigma_1^{\text{eff}}(T)$ displays a narrow peak at $T^* = T_c - \delta T_c$. It is easy to check that the relative peak amplitude is approximately equal to

$$\frac{\sigma_1(T^*) - \sigma(T_c)}{\sigma(T_c)} \approx \begin{cases} \gamma & \text{if } \gamma > 1 \\ \gamma^2 & \text{if } \gamma < 0.1, \end{cases} \tag{22}$$

where $\gamma = \delta T_c/[T_c \omega \tau(T_c)]$, implying that the peak decreases with the decrease of the superconducting resistive transition width. Usually, experiments yield $\gamma > 1$ (e.g., the data of [14] give $\gamma \approx 7$ at 1.14 GHz); therefore, the peak amplitude should be inversely proportional to the frequency.

In calculating the $\sigma_1(T)$ curves for other specimens, we also applied the above procedure to incorporate corrections caused by the inhomogeneous broadening of the superconducting transition. We adjusted the previous calculations of $R_s(T)$ (Figs. 7, 8, and 10) and $\sigma_2(T)$ (Figs. 10, 12, and the inset to Fig. 13) by substituting the resulting $Z_s^{\text{eff}}(T)$ into Eq. (6) for the conductivity σ_1. The resulting curves for BSCCO and TBCO are shown in Figs. 14–16.

4. DISCUSSION AND CONCLUSION

We have presented a summary of measurements of the surface impedance $Z_s(T) = R_s(T) + iX_s(T)$ of high-quality YBCO, BSCCO, TBCO, and TBCCO crystals in the superconducting and normal states (table). For

frequencies $\lesssim$10 GHz, the common features of all these materials are the linear temperature dependence of the surface resistance ($\Delta R_s(T) \propto T$) and that of the surface reactance ($\Delta X_s(T) \propto \Delta\lambda_{ab}(T) \propto T$) at temperatures $T \ll T_c$; their rapid growth as $T \longrightarrow T_c$; and their behavior in the normal state corresponding to a linear T-dependence of $\Delta\rho_{ab}(T)$, with $R_s(T) = X_s(T) = \sqrt{\omega\mu_0\rho(T)/2}$. There are differences between the T-dependence of $Z_s(T)$ in BSCCO, TBCCO, or TBCO single crystals with tetragonal lattices and in YBCO crystals with an orthorhombic lattice. The linear resistivity region extends to near $T_c/2$ for the tetragonal materials and terminates near or below $T < T_c/3$ for YBCO. At higher temperatures, $R_s(T)$ of YBCO has a broad peak. In addition, the $\lambda_{ab}(T)$ curves of some YBCO single crystals have unusual features in the intermediate temperature range.

We describe all of the above features of $Z_s(T)$ and $\sigma_s(T) = \sigma_1(T) - i\sigma_2(T) = i\omega\mu_0/Z_s^2(T)$ of high-quality HTSC crystals by generalizing the well-known GC two-fluid model as follows.

(i) We introduce a temperature dependence of the relaxation time of the quasiparticles in accordance with the Bloch–Grüneisen law. We find that the $R_s(T)$ curves in different HTSC crystals are well described using Eqs. (14) or (15) for $1/\tau(T)$. In the latter equation, there is only one fitting parameter, $\kappa = \Theta/T_c$, while the other parameter $\beta = \tau(T_c)/\tau(0) \ll 1$ can be estimated directly from the experimental data with the help of Eqs. (16) and (17). The absence of the broad peak of $R_s(T)$ in tetragonal HTSC single crystals is due to a less rapid increase of $\tau(T)$ with decreasing the temperature. In other words, the value of β is smaller for YBCO crystals than for BSCCO, TBCO, or TBCCO. For the latter crystals, the residual losses R_{res} are usually large and they have to be taken into account.

(ii) We replace the well-known temperature dependence of the density of superconducting carriers in the GC model ($n_s = n(1 - t^4)$) by one of the functions in Eqs. (18), (19) or (20). All of these functions change linearly with temperature for $t \ll 1$ (see Eq. (1)). This permits one to extract the common and distinctive features of $X_s(T)$ and $\sigma_2(T)$ from different HTSC crystals.

It also follows from the equations of the modified two-fluid model that at low temperatures ($t \ll 1$) and low frequencies ($\omega\tau(0) < 1$), all curves of $Z_s(T)$ and $\sigma_s(T)$ have linear regions: $\sigma_1 \propto \alpha t/\beta$, since $n_n/n \approx \alpha t$ and $\tau \approx \tau(0) \approx \tau(T_c)/\beta$. Furthermore, $\Delta\sigma_2 \propto -\alpha t$. In accordance with Eq. (5), we then have $R_s \propto \alpha t/\beta$ and $\Delta X_s \propto \Delta\lambda \propto \alpha t/2$. As the temperature increases, the curve of $\sigma_1(t)$ passes through a maximum at $t \lesssim 0.5$ if the inequality (9) is valid. This peak is due to the superposition of two competing effects, namely, the decrease in the number of normal carriers as the temperature decreases (for $t < 1$) and the increase in the relaxation time (which saturates at $t \sim \beta^{1/5}$) where the impurity scattering starts to dominate. The features in the $X_s(T)$ and $\sigma_2(T)$ curves for YBCO single crystals in the intermediate temperature range (plateau [8] or bump [9]) can also be described within the framework of our modified two-fluid model if we take into account the modification of $n_s(t)$ described by Eq. (20) with $0 < \delta \leq 0.5$. In HTSC single crystals, the narrow peak in the real part of the conductivity $\sigma_1(T)$ occurring near T_c can be explained in terms of an effective medium model, where the strong electron–phonon coupling of the quasiparticles and the inhomogeneous broadening of the superconducting transition are taken into account.

It is natural to compare the tenets of our phenomenological model with the results of microscopic theories. As was shown in [40] and [61], the simple formula (1), which defines the linear low-temperature dependence of the magnetic field penetration depth in the *ab*-plane of HTSC crystals, is consistent with the *d*-wave model [25–27] in the strong (unitary) scattering limit [31]. Besides, there is nothing foreign in introducing the function $1/\tau(T) \propto T^5$ for the purpose of characterizing scattering in the superconducting state of HTSC. A similar temperature dependence of the relaxation rate of quasiparticles follows from the SC model if the phonon corrections to the electromagnetic vertex are taken into account [94].

In the framework of our modified two-fluid model, the linear low-temperature dependence of the real part of the conductivity $\sigma_1(T)$ is consistent with a constant scattering rate, as it is in a normal metal. While the assumption of a Drude form of the conductivity is supported by the *d*-wave microscopic analysis [31], it was shown that in the usual impurity scattering models, pair correlations lead to a strong temperature dependence of the scattering time (neglecting vertex corrections), namely, $\tau(T) \propto T$ in the unitary limit or $\tau(T) \propto 1/T$ in the Born limit. An attempt to resolve this problem in [16] by choosing an intermediate scattering rate has not provided satisfactory results yet. Very recently, the authors of [95] and [96] argued that the experimental observation $\sigma_1(T) \propto T$ could be explained by the generalized Drude formula $\sigma_1(T) \propto n_{qp}(T)\tau(T)$ if the quasiparticle density varies as $n_{qp}(T) \propto T$ (as indeed happens for the *d*-wave pairing) and if the effective quasiparticle scattering time $\tau(T)$ saturates at low T. Various possible physical mechanisms of the temperature and energy dependence of τ are discussed in [95, 96]: scattering from the "holes" of the order parameter at impurity sites and scattering from extended defects. These mechanisms may provide the required saturation of $\tau(T)$ at low T. As was discussed recently in [97], the vertex corrections can also modify the low-temperature conductivity. However, the temperature dependence has not been investigated yet.

Nevertheless, the microscopic models aimed at investigating the microwave response using a pure d-wave order parameter symmetry cannot account for the linear section of the $R_s(T)$ curves extending to $T_c/2$ (at the frequencies 10 GHz and below) in tetragonal HTSC

single crystals, for the observation of radically different values of the slopes of $\sigma_2(T)$ for $T \ll T_c$ (corresponding to $\alpha > 1$ in Eq. (20)) observed on YBCO crystals [8–12], and for unusual features of $\sigma_2(T)$ in the intermediate temperatures range.

Recently, observations of unusual microwave properties of HTSC materials have caught the attention of a number of researchers [43–47, 55, 56]. These observations are tentatively attributed to the mixed ($d + s$) wave symmetry of the order parameter. Most studies deal with the low-temperature variation of the London penetration depth and its relation to an order parameter of mixed symmetry. In particular, it was shown in [55] that the low-temperature properties of $\lambda(T)$ can be used to distinguish between a pure *d*-wave order parameter and one with the ($s + id$) symmetry, having a small subdominant *s*-wave contribution in systems connected with a tetragonal lattice. Moreover, additions of impurities suppress the *d*-wave symmetric part to the benefit of the *s*-wave part. As a result, a variety of low-temperature dependences of $\lambda(T)$ can occur for various impurity concentrations. This allows one, in principle, to determine whether the order parameter of a superconductor with an orthorhombic lattice pertains to the ($s + id$) or the ($s + d$) symmetry [53]. In [46], the ($d + s$) model was generalized by taking the normal state anisotropy into account. This is the realistic approach to high-T_c cuprates with an orthorhombic distortion, since recent microwave conductivity data suggest that a substantial part of the *ab*-anisotropy of $\lambda(T)$ is a normal state effect. It was shown that such an anisotropy affects not only the *ab*-anisotropy of the transport coefficients but also the density of states and other thermodynamic quantities. The possible temperature variation of the penetration depth $\lambda(T)$ was analyzed recently in [56] in the framework of the ($d + s$) model of hybrid pairing. The slope of $\Delta\lambda(T) \propto T$ for $T \ll T_c$ and its dependence on the Δ_s/Δ_d admixture in the gap function were analyzed quantitatively, with the impurity scattering taken into account. However, the quantitative comparison of the latter calculation with the experimental data has not been performed yet. More interesting discoveries in this field of research can be expected in the near future.

ACKNOWLEDGMENTS

We are greatly indebted to N. Bontemps, V.T. Dolgopolov, V.F. Gantmakher, A.A. Golubov, L.M. Fisher, E.G. Maksimov, and V.P. Mineev for many helpful discussions. H.J. Fink thanks D.A. Bonn for permission to use the original data points of [14] in our figures. The research of Yu.A. Nefyodov and M.R. Trunin was supported by the Russian Foundation for Basic Research (projects no. 00-02-17053 and 00-02-04021) and Scientific Council on Superconductivity (project 96060), and in part by the Program for Russian–Dutch Research Cooperation (NWO).

REFERENCES

1. W. N. Hardy, D. A. Bonn, D. C. Morgan, *et al.*, Phys. Rev. Lett. **70**, 3999 (1993).
2. D. Achkir, M. Poirier, D. A. Bonn, *et al.*, Phys. Rev. B **48**, 13184 (1993).
3. S. Kamal, D. A. Bonn, N. Goldenfeld, *et al.*, Phys. Rev. Lett. **73**, 1845 (1994).
4. D. A. Bonn, S. Kamal, K. Zhang, *et al.*, Phys. Rev. B **50**, 4051 (1994).
5. J. Mao, D. H. Wu, J. L. Peng, *et al.*, Phys. Rev. B **51**, 3316 (1995).
6. D. A. Bonn, S. Kamal, K. Zhang, *et al.*, J. Phys. Chem. Solids **56**, 1941 (1995).
7. T. Jacobs, S. Sridhar, C. T. Rieck, *et al.*, J. Phys. Chem. Solids **56**, 1945 (1995).
8. M. R. Trunin, A. A. Zhukov, G. A. Emel'chenko, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **65**, 893 (1997) [JETP Lett. **65**, 938 (1997)].
9. H. Srikanth, B. A. Willemsen, T. Jacobs, *et al.*, Phys. Rev. B **55**, R14733 (1997).
10. A. A. Zhukov, M. R. Trunin, A. T. Sokolov, *et al.*, Zh. Éksp. Teor. Fiz. **112**, 2210 (1997) [JETP **85**, 1211 (1997)].
11. H. Srikanth, Z. Zhai, S. Sridhar, *et al.*, Phys. Rev. B **57**, 7986 (1998).
12. M. R. Trunin, A. A. Zhukov, and A. T. Sokolov, J. Phys. Chem. Solids **59**, 2125 (1998).
13. S. Kamal, R. Liang, A. Hosseini, *et al.*, Phys. Rev. B **58**, R8933 (1998).
14. A. Hosseini, R. Harris, S. Kamal, *et al.*, Phys. Rev. B **60**, 1349 (1999).
15. L. A. de Vaulchier, J. P. Vieren, Y. Guldner, *et al.*, Europhys. Lett. **33**, 153 (1996).
16. S. Hensen, G. Müller, C. T. Rieck, *et al.*, Phys. Rev. B **56**, 6237 (1997).
17. S. Djordjevich, L. A. de Vaulchier, N. Bontemps, *et al.*, Eur. Phys. J. B **5**, 847 (1998).
18. E. Farber, G. Deutscher, J. P. Contour, *et al.*, Eur. Phys. J. B **5**, 159 (1998).
19. T. Jacobs, S. Sridhar, Q. Li, *et al.*, Phys. Rev. Lett. **75**, 4516 (1995).
20. T. Shibauchi, N. Katase, T. Tamegai, *et al.*, Physica C (Amsterdam) **264**, 227 (1996).
21. S.-F. Lee, D. C. Morgan, R. J. Ormeno, *et al.*, Phys. Rev. Lett. **77**, 735 (1996).
22. D. V. Shovkun, M. R. Trunin, A. A. Zhukov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **71**, 132 (2000) [JETP Lett. **71**, 92 (2000)].
23. D. M. Broun, D. C. Morgan, R. J. Ormeno, *et al.*, Phys. Rev. B **56**, R11443 (1997).
24. J. R. Waldram, D. M. Broun, D. C. Morgan, *et al.*, Phys. Rev. B **59**, 1528 (1999).
25. A. Millis, H. Monien, and D. Pines, Phys. Rev. B **42**, 167 (1990).

26. H. Monien, P. Monthoux, and D. Pines, Phys. Rev. B **43**, 275 (1991).
27. P. Monthoux, A. Balatsky, and D. Pines, Phys. Rev. B **46**, 14803 (1992).
28. P. J. Hirschfeld and N. Goldenfeld, Phys. Rev. B **48**, 4219 (1993).
29. J. P. Carbotte and C. Jiang, Phys. Rev. B **48**, 4231 (1993).
30. H. Won and K. Maki, Phys. Rev. B **49**, 1397 (1994).
31. P. J. Hirschfeld, W. O. Putikka, and D. J. Scalapino, Phys. Rev. Lett. **71**, 3705 (1993); Phys. Rev. B **50**, 4051 (1994).
32. D. J. Scalapino, Phys. Rep. **250**, 329 (1995).
33. J. Annett, N. Goldenfeld, and A. Leggett, in *Physical Properties of High Temperature Superconductors V*, Ed. by D. M. Ginsberg (World Scientific, Singapore, 1996).
34. K. Maki and H. Won, J. Phys. I **6**, 2317 (1996).
35. Yu. A. Izyumov, Usp. Fiz. Nauk **167**, 465 (1997) [Phys. Usp. **40**, 445 (1997)]; Usp. Fiz. Nauk **169**, 225 (1999).
36. R. A. Klemm and S. H. Liu, Phys. Rev. Lett. **74**, 2343 (1995).
37. V. Z. Kresin and S. A. Wolf, Phys. Rev. B **41**, 4278 (1990); **46**, 6458 (1992); **51**, 1229 (1995).
38. A. A. Golubov, M. R. Trunin, A. A. Zhukov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **62**, 477 (1995) [JETP Lett. **62**, 496 (1995)].
39. S. D. Adrian, M. E. Reeves, S. A. Wolf, *et al.*, Phys. Rev. B **51**, 6800 (1995).
40. M. R. Trunin, Usp. Fiz. Nauk **168**, 931 (1998) [Phys. Usp. **41**, 843 (1998)].
41. N. Klein, N. Tellmann, H. Schulz, *et al.*, Phys. Rev. Lett. **71**, 3355 (1993).
42. Q. P. Li, E. C. Koltenbah, and R. Joynt, Phys. Rev. B **48**, 437 (1993).
43. R. Combescot and X. Leyronas, Phys. Rev. Lett. **75**, 3732 (1995).
44. C. O'Donovan and J. P. Carbotte, Phys. Rev. B **52**, 4568 (1995); Phys. Rev. B 55, 8520 (1997).
45. H. Kim and E. J. Nicol, Phys. Rev. B **52**, 13576 (1995).
46. M. T. Beal-Monod and K. Maki, Phys. Rev. B **55**, 11730 (1997); Phys. Rev. B **53**, 5775 (1996); Phys. Rev. B **55**, 1194 (1997).
47. S. V. Pokrovsky and V. L. Pokrovsky, Phys. Rev. B **54**, 13275 (1996).
48. K. A. Musaelian, J. Betouras, A. V. Chubukov, *et al.*, Phys. Rev. B **53**, 3598 (1996).
49. Y. Ren, J. Xu, and C. S. Ting, Phys. Rev. B **53**, 2249 (1996).
50. A. A. Shapoval, Pis'ma Zh. Éksp. Teor. Fiz. **64**, 570 (1996) [JETP Lett. **64**, 625 (1996)].
51. M. Liu, D. Y. Xing, and Z. D. Wang, Phys. Rev. B **55**, 3181 (1997).
52. E. A. Pashitskiĭ and V. I. Pentegov, Zh. Éksp. Teor. Fiz. **111**, 298 (1997) [JETP **84**, 164 (1997)].
53. M. T. Beal-Monod, Phys. Rev. B **58**, 8830 (1998); Physica C (Amsterdam) **298**, 59 (1998).
54. I. Schürrer, E. Schachinger, and J. P. Carbotte, Physica C (Amsterdam) **303**, 287 (1998).
55. R. Modre, I. Schürrer, and E. Schachinger, Phys. Rev. B **57**, 5496 (1998).
56. Yu. A. Nefyodov, A. A. Golubov, M. R. Trunin, *et al.*, Physica B (Amsterdam) **284** (2000) (in press).
57. D. A. Bonn, P. Dosanjh, R. Liang, *et al.*, Phys. Rev. Lett. **68**, 2390 (1992).
58. K. Zhang, D. A. Bonn, R. Liang, *et al.*, Appl. Phys. Lett. **62**, 3019 (1993).
59. D. A. Bonn, R. Liang, T. M. Riseman, *et al.*, Phys. Rev. B **47**, 11314 (1993).
60. H. Kitano, T. Shibauchi, K. Uchinokura, *et al.*, Phys. Rev. B **51**, 1401 (1995).
61. M. R. Trunin, J. Supercond. **11**, 381 (1998).
62. S. M. Quinlan, D. J. Scalapino, and N. Bulut, Phys. Rev. B **49**, 1470 (1994).
63. G. M. Eliashberg, Zh. Éksp. Teor. Fiz. **38**, 966 (1960) [Sov. Phys. JETP **11**, 696 (1960)]; Pis'ma Zh. Éksp. Teor. Fiz. **48**, 275 (1988) [JETP Lett. **48**, 305 (1988)].
64. W. E. Pickett, J. Supercond. **4**, 397 (1991).
65. V. L. Ginzburg and E. G. Maksimov, Sverkhprovodimost': Fiz., Khim., Tekh. **5**, 1543 (1992).
66. A. A. Golubov, M. R. Trunin, A. A. Zhukov, *et al.*, J. Phys. I **6**, 2275 (1996).
67. A. Bille and K. Scharnberg, J. Phys. Chem. Solids **59**, 2110 (1998).
68. C. M. Varma, P. B. Littlewood, S. Schmitt-Rink, *et al.*, Phys. Rev. Lett. **63**, 1996 (1989).
69. E. Abrahams, J. Phys. I **6**, 2191 (1996).
70. P. W. Anderson, Physica C (Amsterdam) **185–189**, 11 (1991).
71. P. W. Anderson, *Theory of Superconductivity in the High-T_c Cuprates* (Princeton Univ. Press, Princeton, 1997).
72. P. A. Lee, Phys. Rev. Lett. **71**, 1887 (1993).
73. M. R. Trunin, A. A. Zhukov, G. E. Tsydynzhapov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **64**, 783 (1996) [JETP Lett. **64**, 832 (1996)].
74. H. J. Fink, Phys. Rev. B **58**, 9415 (1998); **61**, 6346 (2000).
75. H. J. Fink and M. R. Trunin, Physica B (Amsterdam) **284**, 923 (2000).
76. C. S. Gorter and H. Casimir, Phys. Z. **35**, 963 (1934).
77. N. E. Glass and W. F. Hall, Phys. Rev. B **44**, 4495 (1991).
78. H. K. Olsson and R. H. Koch, Phys. Rev. Lett. **68**, 2406 (1992).
79. A. A. Golubov, M. R. Trunin, S. V. Shulga, *et al.*, Physica C (Amsterdam) **213**, 139 (1993).
80. M. L. Horbach, W. van Saarlos, and D. A. Huse, Phys. Rev. Lett. **67**, 3464 (1991).

81. S. M. Anlage, J. Mao, J. C. Booth, *et al.*, Phys. Rev. B **53**, 2792 (1996).

82. G. M. Eliashberg, Zh. Éksp. Teor. Fiz. **39**, 1437 (1960) [Sov. Phys. JETP **12**, 1000 (1961)].

83. A. E. Karakozov, E. G. Maksimov, and S. A. Mashkov, Zh. Éksp. Teor. Fiz. **68**, 1937 (1975) [Sov. Phys. JETP **41**, 971 (1975)].

84. F. Marsiglio, Phys. Rev. B **44**, 5373 (1991).

85. A. E. Karakozov, E. G. Maksimov, and A. A. Mikhaĭlovskiĭ, Zh. Éksp. Teor. Fiz. **102**, 132 (1992) [Sov. Phys. JETP **75**, 70 (1992)].

86. O. V. Dolgov, E. G. Maksimov, A. E. Karakozov, *et al.*, Solid State Commun. **89**, 827 (1994).

87. G. V. Klimovich, A. V. Rylyakov, and G. M. Eliashberg, Pis'ma Zh. Éksp. Teor. Fiz. **53**, 381 (1991) [JETP Lett. **53**, 399 (1991)].

88. A. A. Mikhailovsky, S.V. Shulga, A. E. Karakozov, *et al.*, Solid State Commun. **80**, 511 (1991).

89. R. T. Collins, Z. Schlesinger, F. Holtzberg, *et al.*, Phys. Rev. B **43**, 8701 (1991).

90. J. Rammer, Europhys. Lett. **5**, 77 (1988).

91. A. Andreone, C. Cantoni, A. Cassinese, *et al.*, Phys. Rev. B **56**, 7874 (1997).

92. E. G. Maksimov, D. Yu. Savrasov, and S. Yu. Savrasov, Usp. Fiz. Nauk **167**, 354 (1997).

93. M. R. Trunin, A. A. Zhukov, and A. T. Sokolov, Zh. Éksp. Teor. Fiz. **111**, 696 (1997) [JETP **84**, 383 (1997)].

94. G. M. Eliashberg, G. V. Klimovich, and A. V. Rylyakov, J. Supercond. **4**, 393 (1991).

95. M. H. Hettler and P. J. Hirschfeld, Phys. Rev. B **61**, 11313 (2000).

96. A. J. Berlinsky, D. A. Bonn, R. Harris, *et al.*, Phys. Rev. B **61**, 9088 (2000).

97. A. C. Durst and P. A. Lee, cond-mat/9908182.

Journal of Experimental and Theoretical Physics, Vol. 91, No. 4, 2000, pp. 817–831.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 4, 2000, pp. 941–958.

SOLIDS
Electronic Properties

Coulomb Pairing of Like-Charged Particles with Negative Effective Mass in High-Temperature Superconductors

V. I. Belyavskiĭ[a], V. V. Kapaev[b, *], and Yu. V. Kopaev[b]

[a]*Voronezh State Pedagogical University, Voronezh, 394043 Russia*
[b]*Lebedev Physical Institute, Russian Academy of Sciences, Moscow, 117924 Russia*
**e-mail: kopaev@sci.lebedev.ru*

Received May 18, 2000

Abstract—Quasi-steady states of pairs of like-charged quasi-particles can be formed because the electronic structure of compounds exhibiting high-temperature superconductivity has various important characteristics: a quasi-two-dimensional electron spectrum, clearly defined nesting of constant-energy lines, and the presence of a logarithmic singularity of the density of states in the immediate vicinity of the Fermi level. Thus, a situation is achieved where, in an extensive region of the Brillouin zone adjacent to the Fermi level, the principal values of the tensor of the reciprocal effective masses have opposite signs and differ appreciably in absolute value. As a result, the nature of the Coulomb correlation interaction between charge carriers of the same sign (holes in *p*-cuprates) varies: effective attraction may predominate, leading to the formation of long-lived states of relative motion of quasi-particles which form a pair having a quasi-momentum approximately equal to twice the Fermi quasi-momentum typical of this direction (focused pairs). Assuming that the correlation interaction is short-range (screened Coulomb interaction attenuated by filling of states inside the Fermi contour), we determine the energies and envelope functions of the relative motion of a hole pair which correspond to the density-of-states maxima of the pairs attributable to these quasi-steady states. The dependence of these quantities on the polar angle in the plane of the conducting layer reflects the symmetry of the electronic structure of the compound in the normal state and is generally consistent with a mixture of states assigned to *s* and *d* types of orbital symmetry. The quasi-steady state as a function of the doping level of the system agrees qualitatively with the concentration dependence of the temperature for the appearance of a pseudogap observed in *p*-cuprates at below-optimum doping levels. An estimate of the pair concentration above which a gain in correlation energy occurs gives a value which corresponds to the onset of effective pair overlap (for which the characteristic spatial scale is a few or a few tens of interatomic distances).

1. INTRODUCTION

The high-temperature superconductivity (HTSC) of a fairly broad class of compounds has not yet received a generally accepted interpretation. It is usually assumed that the properties of HTSC compounds in the superconducting and normal states are mainly determined by two characteristics: the quasi-two-dimensional nature of the electronic structure and the strong electron–electron interaction. It can be confirmed that the HTSC state is produced by pair electron correlations; however the reason for the appearance of bound states of electron or hole pairs in HTSC compounds remains unclear. Electron–phonon interaction hole pairing, which forms the basis of the Bardeen–Cooper–Schrieffer (BCS) theory and imposes a substantial constraint on the superconducting transition temperature, can hardly be responsible for HTSC since in some HTSC compounds the binding energy is an order of magnitude higher than obtained in principle from the BCS theory, so that the average size of a Coulomb pair is comparable with the average distance between the pairs. The formation of such compact composite quasi-particles composed of two fermions cannot be specifically attributed merely to electron–phonon interaction under any reasonable assumptions on the intensity of this interaction. There is thus reason to assume that the formation of electron or hole pairs in HTSC compounds is directly attributable to the specific characteristics of their electron subsystem.

An extensive and the most well-studied group of cuprate HTSC compounds mainly exhibits hole type (*p*-type) conductivity ($La_{2-x}Sr_xCuO_4$, $YBa_2Cu_3O_7$, $Bi_2Sr_2CaCu_2O_8$, and many others); some HTSC cuprates (such as $Nd_{2-x}Ce_xCuO_4$) have electron (*n*-type) conductivity. All other known HTSC compounds also generally have hole-type conductivity. The type of conductivity and carrier concentration are associated with deviations from stoichiometric composition (stoichiometric HTSC compounds are dielectrics) and the presence of dopants. These two factors determine the doping level of a particular compound.

All known HTSC compounds have a layered crystal structure; moreover fairly convincing experimental facts indicate that in cuprate compounds the HTSC

state occurs as a result of carrier pairing in CuO_2 layers [1]. Arguments have also been put forward to indicate that the superconducting transition temperature in *p*-type HTSC cuprates only depends on the hole concentration in the CuO_2 layers [2] whereas the charge screening in the conducting layers is determined by the total carrier concentration. The layered structure of HTSC cuprates allows these to be considered as natural superlattices in which the lateral conductivity is determined by the properties of the CuO_2 layers which function as unique quantum wells separated by barrier layers [3]. These quantum wells can be, for example, isolated CuO_2 layers in $La_{2-x}Sr_xCuO_4$ and $Nd_{2-x}Ce_xCuO_4$ compounds; double CuO_2 layers separated by a layer containing Y atoms in $YBa_2Cu_3O_7$; triple CuO_2 layers separated by layers containing Ca atoms in $Bi_2Sr_2CaCu_2O_8$. Thus, layers containing La, Ce, Ba, and Bi atoms play the role of barriers and reservoirs at the same time and when these are doped, for example, with Sr atoms in $La_{2-x}Sr_xCuO_4$, Ce in $Nd_{2-x}Ce_xCuO_4$, and oxygen in $YBa_2Cu_3O_7$ and $Bi_2Sr_2CaCu_2O_8$ carriers enter the CuO_2 layers. The relative influence of the layers on each other is small [1] since HTSC compounds can be considered to be quasi-two-dimensional electron systems. However, the atomic composition of the layers positioned between the CuO_2 layers has an appreciable influence on the superconducting transition temperature [4]. The introduction of dopants into *p*-type cuprate compounds, for example, Sr into an $La_{2-x}Sr_xCuO_4$ crystal or the formation of oxygen vacancies, for example, by the removal of oxygen from a $YBa_2Cu_3O_7$ crystal is accompanied by transitions of electrons from the *d*-shell of the Cu atoms to restore broken chemical bonds (to divalent Sr which has replaced trivalent La in $La_{2-x}Sr_xCuO_4$ or to the chemical bond of a Ba atom in $YBa_2Cu_3O_{7-x}$ which is unsaturated as a result of the formation of an oxygen vacancy). In this case, the hole localized mainly at the copper atom and the negative charge of the restored chemical bond are positioned on the straight, almost parallel, *c* axis of the structure. This configuration minimizes the Madelung energy of the crystal and is therefore a stabilizing factor which significantly compensates for the increase in elastic energy associated with the defect structure of the reservoir layers.

The migration of electrons in HTSC cuprates is in fact limited by the CuO_2 layers; in the direction of the *c* axis perpendicular to these layers the normal-state conductivity is substantially lower than that parallel to the layers. The constant-energy surfaces of the electron dispersion law in zones containing the Fermi surface are thus slightly corrugated cylinders whose axis coincides with the *c* axis [5]. The smallness of the matrix elements of the single-electron Hamiltonian linking the atoms in neighboring CuO_2 layers compared with the matrix elements between the atoms inside the layer means that the electron subsystem of HTSC cuprates can be considered to be almost two-dimensional (2D). The cross section of the constant-energy surfaces with the plane perpendicular to the *c* axis determines a family of constant-energy lines (isolines) in the 2D Brillouin zone (in compounds having tetragonal symmetry this is a square of side length $2\pi/a$ where a is the period of the crystal lattice in the plane of the CuO_2 layer). One of the isolines, the Fermi contour, corresponds to the Fermi surface.

The distribution of isolines in the 2D Brillouin zone is determined by the interaction of atoms in the CuO_2 layers with each other and with atoms belonging to neighboring layers. In *p*-type HTSC compounds with an optimum doping level (i.e., for that carrier concentration which gives the maximum superconducting transition temperature T_c) the Fermi contour determined experimentally [6–10] using angular-resolved photoemission spectroscopy (ARPES) at $T > T_c$ is a square with rounded corners. A change in doping level corresponds to a transition of the Fermi contour from one isoline to another, which means that ARPES can be used to study the character of the isolines in a fairly large region of the Brillouin zone. If the doping level is below the optimum (underdoped HTSC compounds), over a wide range of concentrations at temperatures above a certain value $T^* > T_c$ the isoline corresponding to the Fermi contour is also a square with rounded corners. The Fermi contour also has this profile when the carrier concentration is above the optimum (overdoped HTSC compounds); in this case, the Fermi contour remains in the form of a square with rounded corners as far as the temperature T_c, at which a superconducting gap appears in the spectrum of elementary excitations. Thus, the Fermi contour of *p*-type HTSC cuprates has relatively long almost rectilinear sections which are parallel because of the crystal symmetry (Fermi contour nesting). In this context it should be noted that in *n*-type HTSC cuprates such as $Nd_{2-x}Ce_xCuO_4$ no nesting of the Fermi contour occurs, the contour is close to a circle [8,9]. If the doping level x (for example, the atomic concentration of dopant or oxygen vacancies) is such that it corresponds to the superconducting phase at low temperatures ($T < T_c$), then at $T > T_c$ the HTSC compound behaves as a metal and the area of the region of the Brillouin zone inside the Fermi contour is determined by the atomic concentration $1 + x$ in accordance with the Luttinger theorem [6, 9, 10].

Another important characteristic of HTSC cuprates [11] is that in *p*-type compounds a Van Hove logarithmic singularity is situated in the immediate vicinity of the Fermi surface, whereas in *n*-type compounds it is substantially lower than the Fermi level [11]. This singularity corresponds to the saddle point of the electron dispersion law at which the electron (or hole) energy as a function of the 2D quasi-momentum passes through a maximum in one direction and a minimum in the other, perpendicular, direction. Thus, in a certain vicinity of the saddle point the principal values of the 2D tensor of the reciprocal effective masses have different signs and

also have large absolute values, i.e., in *p*-type HTSC cuprates this vicinity occupies a fairly extensive part of the 2D Brillouin zone.

The explanation of the nature of pairing and the establishment of the symmetry of the order parameter are closely interrelated and form an extremely important stage in constructing the theory of high-temperature superconductivity. Unfortunately, numerous experimental data for various HTSC cuprates which can be used as the basis to assess the symmetry of the order parameter cannot always be satisfactorily matched [12]. There is some experimental evidence that in *p*-type HTSC cuprates the pairing possibly corresponds to $d_{x^2-y^2}$ orbital symmetry [13]. Experimental data relating to *n*-type HTSC compounds (specifically $Nd_{2-x}Ce_xCuO_4$) tend to suggest *s*-wave pairing or, more accurately, indicate the absence of any zeros of the order parameter [14]. However, an STM/STS study of the excitation spectra in $Nd_{2-x}Ce_xCuO_4$ single crystals specifically indicates anisotropy of the order parameter and this anisotropy is characterized by *d*-wave symmetry [15]. An analysis of various experimental data [12] indicates with a fairly high degree of certainty that the order parameter is a spin singlet and may be approximately represented as a linear combination of orbital *s*- and *d*-wave functions. Similar anisotropy of the order parameter in HTSC cuprates may be indirectly evidenced by experimental data on the Raman electron spectra of the HTSC crystals $Bi_2Sr_2CaCu_2O_{8+\delta}$ and $Tl_2Ba_2CuO_{6+\delta}$ [16] and also tunnel spectroscopy data for $Nd_{1.85}Ce_{0.15}CuO_{4-\delta}$ [5]. The angular dependence of the superconducting gap obtained in [5] $\Delta(\vartheta) = \Delta_0 + \Delta_1\cos 4\vartheta$, where $\Delta_0 \approx 2.2$ meV, $\Delta_1 \approx 1.5$ meV agrees with the conclusion reached in [13] on the basis of measurements of the angular dependence of the upper critical field: $\Delta(\vartheta)$ has four maxima and four minima in the range of variation of the direction of the quasi-wave 2D vector $0 \le \vartheta < 2\pi$.

Much experimental data [17] for a large number of very different underdoped HTSC compounds indicate that a gap exists (a so-called pseudogap since there is still a certain number of excitations within this gap) in the spectrum of elementary excitations at $T_c < T < T^*$. In overdoped HTSC compounds this pseudogap is evidently absent. There is a definite link between the pseudogap and the superconducting gap. For instance, NMR experiments carried out for a large group of HTSC cuprates [17] show that the pseudogap and superconducting gap in these compounds have the same symmetry and energy scale and for various cuprates exhibit a universal dependence on the doping level. The fact that the width of the pseudogap and the superconducting transition temperature for an optimally doped crystal have the same energy scale, unrelated to the antiferromagnetic exchange energy or the Fermi energy, is interpreted as meaning that this is the same energy scale which forms the basis of the pairing interaction [17]. The formation of a pseudogap in underdoped $Bi_2Sr_2CaCu_2O_{8+\delta}$ HTSC single crystals as the temperature decreased from T^* to T_c was observed using ARPES [18]. The pseudogap on the Fermi contour appears at $T = T^*$ at four points at the centers of the sides of a square with rounded corners and as the temperature decreases further, it extends toward the corners. Thus, the Fermi contour is discontinuous and comprises four arcs (rounded edges of a square) which are not interconnected and gradually decrease with decreasing temperature. At $T = T_c$ the Fermi contour disappears completely and is replaced by a superconducting gap, and this gap evidently has zeros (or is minimal) at those points where the Fermi contour last disappeared. No phase transition associated with the presence of a pseudogap is observed. At present there is no common and generally accepted viewpoint on the nature of the pseudogap. In particular, it is assumed that at $T = T^*$ Cooper pairs appear and the phase coherence between them appears, not when they are formed but at a substantially lower temperature $T = T_c$ [19, 20]. There is reason to assume [21] that the formation of the pseudogap can be attributed to the subsequent (with decreasing temperature) formation of a superconducting state rather than a dielectric (antiferromagnetic) one since the problem of identifying the nature of the pseudogap in HTSC compounds is undoubtedly closely related to the problem of understanding the hole pairing mechanism in *p*-type HTSC cuprates.

The observed anisotropy of the pseudogap and the order parameter in HTSC cuprates is apparently also related to the symmetry of the Fermi contour and the characteristics of the electronic structure near the Fermi level. Despite the fact that the electrons in HTSC cuprates are strongly correlated, from a certain doping level the carriers can provide fairly effective screening so that the electronic structure of these compounds at energies close to the Fermi level can be fairly accurately described using a band scheme [11]. The results of band calculations [22] broadly show satisfactory agreement with the existing experimental data, reproducing the observed profile of the Fermi contour in *p*- and *n*-type cuprate HTSC compounds. The simplest approximation of nearest (in the CuO_2) plane neighbors in a strong coupling scheme leads to ideal (at half filling) nesting of the Fermi contour, which has the form of a square oriented in the $\langle 110\rangle$ directions. However, the real interactions in the CuO_2 layer are such that in addition to nearest neighbors, it is fundamentally important to allow for interactions in the next few coordination spheres. For example, as a result of allowing for interaction between atoms in the first, second, and third coordination spheres with quite specific relationships between the corresponding matrix elements of the single-electron Hamiltonian [23], the Fermi contour acquires the form of a square with rounded corners with the sides oriented in the $\langle 100\rangle$ directions. This profile of the Fermi contour is consistent with the experimental

data obtained using ARPES for a whole range of *p*-type HTSC cuprates. In addition, it is also consistent with the fact that the vector determining the antiferromagnetic structure of stoichiometric HTSC cuprates is also directed along ⟨100⟩ rather than ⟨110⟩ as is deduced from the nearest neighbor approximation. The physical reason for this rotation of the Fermi contour through the angle $\pi/4$ and the corresponding rearrangement of the isolines is clearly interaction between atoms of the CuO_2 layer and atoms of the reservoir layers. In fact, the chemical bonds between nearest neighbors in these layers are oriented in the ⟨110⟩ directions whereas in the CuO_2 layer, the chemical bonds are oriented in the ⟨100⟩ direction so that by creating a strained state in the CuO_2 layers, the reservoirs in fact fix the electronic structure in the bands associated with the atoms of the CuO_2 layer.

A change in the sign of one of the effective masses has the result that when pair correlations are taken into account, the Coulomb interelectron interaction ensures effective attraction of electrons or holes [24]. It thus becomes possible in principle for quasi-steady states of like-charged carrier pairs to form. These states may be considered as various composite quasiparticles which alter the character of the statistics in the system. In particular, if the characteristic spatial scale of the correlation in the pair is comparable with or smaller than the average distance between the pairs and the lifetime of the quasi-steady state is fairly long, these composite quasiparticles can be considered to be Bose particles to a rough approximation, for example to estimate the ground-state energy.

In the present study, taking into account the real electronic structure of cuprate HTSC compounds we investigate the pair correlations in the electron subsystem, which are manifest as quasi-steady states of hole pairs, i.e., we solve the problem of two quasiparticles similar to the Cooper problem.

2. CORRELATION ENERGY

Since the characteristic spatial scale of the quasi-steady states is in any case appreciably greater than the interatomic distance, it is natural to use the method of envelope functions to describe correlation effects. In this case, the electron dispersion law $\varepsilon(k)$ should be considered to be a known function of the 2D quasi-momentum k and an electron in a crystal can be considered to be a free quasiparticle having a complex dispersion law. The envelope function of this quasiparticle is a plane wave $|k\alpha\rangle$; here the spin quantum number α has values of $\pm 1/2$ (subsequently denoted as $\uparrow$ and $\downarrow$).

Using the electron creation operators $\hat{c}^+_{k\alpha}$ in the single-particle states $k\alpha$, the state of a system of N particles can be obtained from the vacuum state $|\mathrm{vac}\rangle$

$$|\ \rangle = \prod_{k\alpha} \hat{c}^+_{k\alpha}|\mathrm{vac}\rangle, \qquad (2.1)$$

where the product contains exactly N electron creation operators in various given single-particle states. In particular, (2.1) is the wave function of the ground state of the system in the absence of interaction between electrons if the electron quasi-momenta in the product (2.1) fill the inner region of the 2D Brillouin zone bounded by the Fermi contour.

As a result of the pair short-range interaction $U(r)$ between the electrons, the contribution of the pair correlations to the system energy is a determining factor compared with the contribution of higher-order correlations [25, 26]. In order to make approximate allowance for the pair correlations, we define a basis constructed from the envelope functions of electron pairs [27]. We shall assume that the state formed by $N/2$ singlet-coupled electron pairs is the ground state. We shall also assume that the Schrödinger equation with the potential $U(r)$ is solved for any pair of electrons. In the orbital part of the envelope function of the pair which, depending on the spin state of the pair, is either symmetric (for singlet pairing) or antisymmetric (for triplet pairing) with respect to permutation of the electron position vectors, we can explicitly isolate a cofactor corresponding to the motion of the center of inertia of the pair:

$$\Psi(r_1, r_2) = \frac{1}{\sqrt{S}}\varphi_{Kk}(r)\exp(iKR) \equiv |Kk\rangle. \qquad (2.2)$$

Here $K = k_1 + k_2$ and $R = (r_1 + r_2)/2$ are the quasi-momentum and position vector of the center of inertia, respectively, and $k = (k_1 - k_2)/2$ and $r = r_1 - r_2$ are the quasi-momentum and position vector of the relative motion of the electrons forming the pair, respectively. In the case of singlet pairing considered below the envelope function of the relative motion is $\varphi_{Kk}(-r) = \varphi_{Kk}(r)$. The vector k can be considered to be the initial quasi-momentum of a particle having a reduced effective mass scattered at the potential $U(r)$.

We express (2.2) in the form

$$|Kk\rangle = \sum_q \left|\frac{K}{2} + q, \frac{K}{2} - q\right\rangle \tilde{\varphi}_K(k, q). \qquad (2.3)$$

Here $|k_1k_2\rangle$ is the product of two plane waves having the quasi-momenta $k_1 = K/2 + q$ and $k_2 = K/2 - q$ without any factors which depend on the spin variables

$$\tilde{\varphi}_K(k, q) = \frac{1}{\sqrt{S}}\int d^2r\,\varphi_{Kk}(r)\exp(-iqr), \qquad (2.4)$$

S is the normalization area in the plane of the conducting layer. The Fourier transform $\tilde{\varphi}_K(k, q)$ of the envelope function of the relative motion of the pair electrons should possess specific symmetry with respect to permutation of the initial (k) and final (q) quasi-momenta: $\tilde{\varphi}_K(k, q) = \tilde{\varphi}^*_K(k, q)$. The functions $\tilde{\varphi}_K(k, q)$ form a complete orthonormalized system.

The case of Cooper pairs having optimal zero total momentum and a large momentum of relative motion

will not be considered, even though this pairing is not sensitive to the profile of the Fermi surface. This pairing only occurs in the presence of effective attraction.

When the Fermi contour has rectilinear sections, an alternative structure of electron or hole pairs with a high overall momentum (of the order of the distance between the parallel sections of the Fermi contour) is obtained. In the presence of effective attraction, the formation of such a pair, like a Cooper pair, corresponds to an energy minimum. However, this pairing can also occur under conditions of repulsive interaction when a negative light component of the effective mass appears on a large part of the Fermi contour. It is shown below that this case is achieved in cuprate high-temperature superconductors. The formation of such an isolated pair corresponds to an energy maximum. At concentrations of these pairs above a certain critical level, their formation becomes energetically favorable as a result of suppression of the correlations between these boson pairs compared with the correlations between unpaired fermions.

We shall determine the creation operator of an electron pair [27] having the quantum numbers K and k:

$$|Kk\rangle = \hat{A}^+_{Kk}|\mathrm{vac}\rangle,$$
$$\hat{A}^+_{Kk} = \sum_q \tilde{\varphi}_K(k, q)\hat{c}^+_{K/2+q\uparrow}\hat{c}^+_{K/2-q\downarrow}, \tag{2.5}$$

and we denote by $\{F\}$ the region of k-space bounded by the Fermi contour. Then, if the vectors $k_1 = K/2 + k$ and $k_2 = K/2 - k$ belong to $\{F\}$, the wave function of the system of noninteracting electrons can be expressed in the form

$$|F\rangle = \prod_{k_1k_2}\hat{c}^+_{k_1\uparrow}\hat{c}^+_{k_2\downarrow}|\mathrm{vac}\rangle$$
$$= \prod_{Kk}\hat{c}^+_{K/2+k\uparrow}\hat{c}^+_{K/2-k\downarrow}|\mathrm{vac}\rangle. \tag{2.6}$$

Expression (2.6) contains the vectors k_1 and k_2 corresponding to all the different points of the set $\{F\}$. The vectors $K/2 \pm k$ can be obtained in principle by various methods. For instance, the vector $k_1 = K/2 + k$ may appear in pairs with different K. Thus only those cofactors corresponding to all different k_1 and k_2 should be retained in the product over Kk (formally this conditions is automatically satisfied in (2.6)). Thus, for each, generally speaking, vector K a certain set of points should be excluded from the corresponding region of definition of the vector k in order to ensure that the number of electrons in the system is constant.

We write the wave function of the ground state of the system of interacting electrons in the form

$$|0\rangle = \prod_{Kk}\hat{A}^+_{Kk}|\mathrm{vac}\rangle \tag{2.7}$$

provided that $K/2 \pm k \in \{F\}$. Since the summation variable q in the definition of the pair creation operator is in no way bounded, a contribution in (2.7) is made by single-electron states belonging to the inner and outer regions relative to the Fermi contour.

When no scattering takes place at the potential $U(r)$ (which in, particular, corresponds to the single-electron approximation), it is clear that

$$\varphi_{Kk}(r) \longrightarrow \frac{1}{\sqrt{S}}\exp(ikr), \quad \tilde{\varphi}_K(k, q) \longrightarrow \delta_{qk} \tag{2.8}$$

for any K so that the ground state is the same as (2.6) as should be the case.

Since the identity $\hat{c}_{k\alpha}\hat{c}^+_{k\alpha}|\mathrm{vac}\rangle \equiv |\mathrm{vac}\rangle$ is obtained for any k, the electron pair creation operator can be expressed as

$$\hat{A}^+_{Kk} = \left\{\tilde{\varphi}_K(k, k) + \sum_{q\neq k}\tilde{\varphi}_K(k, q)\hat{c}^+_{K/2+q\uparrow}\hat{c}^+_{K/2-q\downarrow}\hat{c}_{K/2-k\downarrow}\hat{c}_{K/2+k\uparrow}\right\} \times \hat{c}^+_{K/2+k\uparrow}\hat{c}^+_{K/2-k\downarrow}, \tag{2.9}$$

having isolated the term corresponding to the forward scattering amplitude. The ground state of the system has the form

$$|0\rangle = \alpha\prod_{Kk}\left\{1 + \sum_{q\neq k}\alpha_K(k, q) \times \hat{c}^+_{K/2+q\uparrow}\hat{c}^+_{K/2-q\downarrow}\hat{c}_{K/2-k\downarrow}\hat{c}_{K/2+k\uparrow}\right\}|F\rangle, \tag{2.10}$$

where $\alpha_K(k, q) = \tilde{\varphi}_K(k, q)/\tilde{\varphi}_K(k, k)$, $\alpha \equiv \prod_{Kk}\tilde{\varphi}_K(k, k)$. In fact, a systematic permutation of pairs of electron creation and annihilation operators having the same quasi-momenta and spins does not lead to any change in the sign of the wave function; in addition, in the sum in (2.10) we obtain $q \neq k$ since no nonzero contractions of the operators exist. Since by definition $K/2 \pm k \in \{F\}$, then conversely $K/2 \pm q \notin \{F\}$ since otherwise the product of the creation operators in (2.9) would contain a term identically equal to zero. Thus, the operators $\hat{c}_{K/2+k\uparrow}$ and $\hat{c}_{K/2-k\downarrow}$ create holes inside $\{F\}$ while the operators $\hat{c}^+_{K/2+q\uparrow}$ and $\hat{c}^+_{K/2-q\downarrow}$ create electrons outside $\{F\}$. We can confirm that the wave function of the ground state (2.10) is normalized to unity: $\langle 0|0\rangle = 1$.

Systematically multiplying the (commuting) operators, (2.10) can be expressed in the form $|0\rangle = \alpha\{1 + \hat{G}\}|F\rangle$ where the operator $\hat{G}$ may be written as the sum

$\hat{G} = \hat{G}^{(1)} + \hat{G}^{(2)} + \ldots$, where the first term contains all possible quadruplets of the operators $\hat{c}^{+}_{K/2+q\uparrow}$ $\hat{c}^{+}_{K/2-q\downarrow}$ $\hat{c}_{K/2-k\downarrow}$ $\hat{c}_{K/2+k\uparrow}$, the second term contains quadratic combinations of these quadruplets, and so on. For example,

$$\hat{G}^{(1)} = \sum_{Kkq} \alpha_K(k, q)\hat{c}^{+}_{K/2+q\uparrow}\hat{c}^{+}_{K/2-q\downarrow}\hat{c}_{K/2-k\downarrow}\hat{c}_{K/2+k\uparrow} \quad (2.11)$$

describes the states of a system with a hole pair inside $\{F\}$ and an electron pair outside $\{F\}$, the operator $\hat{G}^{(2)}$ corresponds to states with two hole pairs and two electron pairs, and so on.

From the energy of the ground state (2.10) of a system of electrons having the Hamiltonian $\hat{H}$, $E_0 = \langle 0|\hat{H}|0\rangle$ we can isolate the electron energy in the Hartree–Fock approximation, $E_{HF} = \langle F|\hat{H}|F\rangle$; the correlation energy of the electron system $E_{\text{corr}} = E_0 - E_{HF}$ is expressed in the form

$$E_{\text{corr}} = \frac{\langle F|\hat{G}^{+}(\hat{H} - E_{HF})\hat{G}|F\rangle}{1 + \langle F|\hat{G}^{+}\hat{H}\hat{G}|F\rangle}. \quad (2.12)$$

In order to calculate the average in (2.12) we shall confine ourselves to the simplest approximation $\hat{G} \approx \hat{G}^{(1)}$ which is linear with respect to the parameters $\alpha_K(k, q)$. Since in the Hartree–Fock approximation $\hat{G} = 0$, the expression $\hat{G} = \hat{G}^{(1)} + \hat{G}^{(2)} + \ldots$ can be considered as an expansion of the operator $\hat{G}$ in powers of the parameters $\alpha_K(k, q)$ in a certain sense similar to the gas parameter in the kinetic theory of gases. In this case $\hat{G} \approx \hat{G}^{(1)}$ corresponds to the self-consistent field approximation for electron and hole pairs whereas the term $\hat{G}^{(2)}$ allows for pair interaction between isolated pairs.

The correlation correction to the kinetic energy has the form

$$T_{\text{corr}} = |\alpha|^2 \sum_{Kkq} \alpha^*_K(k, q)\left\{\left[\varepsilon\left(\frac{K}{2} + q\right) + \varepsilon\left(\frac{K}{2} - q\right)\right] - \left[\varepsilon\left(\frac{K}{2} + k\right) + \varepsilon\left(\frac{K}{2} - k\right)\right]\right\}\alpha_K(k, q). \quad (2.13)$$

Here the first term in braces is the kinetic energy of an electron pair excited outside $\{F\}$, and the second term is the kinetic energy of a pair inside $\{F\}$. If, as is usually the case, the energy of a pair as an elementary excitation is measured from twice the chemical potential, it is clear that the second term in (2.13) corresponds to the kinetic energy of a hole pair inside $\{F\}$.

Calculations of the correlation energy produced by the electron–electron interaction give a bilinear expression in terms of the scattering amplitudes $U_{\text{corr}} = U^{(h-h)} + U^{(e-e)} + U^{(h-e)}$, where

$$U^{(h-h)} = |\alpha|^2 \sum_{Kkq\kappa} \alpha_K(k - \kappa, q)\tilde{U}(\kappa)\alpha^*_K(k, q) \quad (2.14)$$

is the interaction energy of holes forming a pair inside $\{F\}$,

$$U^{(e-e)} = |\alpha|^2 \sum_{Kkq\kappa} \alpha^*_K(k, q - \kappa)\tilde{U}(\kappa)\alpha_K(k, q) \quad (2.15)$$

is the interaction energy of electrons forming a pair outside $\{F\}$. Here

$$\tilde{U}(q) = \frac{1}{S}\int d^2 r U(r)\exp(-iqr) \quad (2.16)$$

is the Fourier transform of the 2D potential of the electron–electron interaction. The potential $U(r)$ should be understood as the result of averaging the three-dimensional screened Coulomb potential along the c-axis of the structure (averaging with corresponding envelope functions). Since the thickness of the conducting layer has atomic dimensions, the result of the averaging may be taken to be the potential at $z = 0$ (z-axis parallel to c). Thus, for estimates we can use an expression of the type

$$U(r) = \frac{e^2}{r}\exp\left(-\frac{r}{r_0}\right), \quad (2.17)$$

where r_0 is the screening radius and r is the 2D position vector. Bearing in mind that the thickness of the barrier layers separating the conducting layers also has atomic dimensions, we can neglect the screening characteristics in two-dimensional systems. Hence, for hole concentrations typical of HTSC compounds the screening radius can be estimated in the Thomas–Fermi approximation:

$$r_0 = [4\pi e^2 g(\mu)]^{-1/2}. \quad (2.18)$$

Here $g(\mu)$ is the density of states at the Fermi level per unit volume.

We shall not give an explicit expression for the energy $U^{(h-e)}$ of the electron–hole interaction in pairs inside and outside $\{F\}$ which in principle leads to electron and hole pairing; also, we shall not give the terms in U_{corr} which are linear with respect to the pair scattering amplitudes and the electron occupation numbers in the $\{F\}$ state since, in the very simple approximation used here, these contributions to the correlation energy cannot influence the envelope functions of the relative motion of the electron and hole pairs.

Summation over κ (unlike the sums over q and k which correspond to certain specific ranges of values of

these vectors for each quasi-momentum of the pair K) is performed over the entire Brillouin zone. In (2.14) the vector k plays the role of an independent variable whereas the vector q determines the state of the internal motion of a hole pair. Conversely, in (2.15) the independent variable is q and k should be considered to be the quantum number relating to the internal state of an electron pair. For the hole-type conductivity of a HTSC compound it is usually convenient to convert from an electron to a hole representation where the Fermi contour defines the region $\{\bar{F}\}$ which adds $\{F\}$ to the complete Brillouin zone.

3. EQUIVALENT HAMILTONIAN OF A HOLE PAIR

In order to write the equivalent Hamiltonian of a hole pair having the quasi-momentum K, we shall use expression (2.13) in which the second term in braces, taken with the opposite sign, has the meaning of the kinetic energy of a hole pair and (2.14) which describes the direct screening Coulomb interaction of holes forming the pair. We set $k \longrightarrow \hat{k} = -i\nabla$ and express the equivalent Hamiltonian in the form $\hat{H} = \hat{T} + \hat{U}$ where the kinetic energy operator of the pair is

$$\hat{T} = \varepsilon\left(\frac{K}{2} - i\nabla\right) + \varepsilon\left(\frac{K}{2} + i\nabla\right). \tag{3.1}$$

Converting to the r-representation in (2.14), we have

$$\sum_{K,k} \tilde{\varphi}_{Kq}(k-\kappa, q)\tilde{U}(\kappa)\tilde{\varphi}^*_{Kq}(k, q) = \frac{1}{S}\sum_k \iint d^2r d^2r' \varphi_{Kq}(r')U(r')\varphi^*_{Kq}\exp[ik(r-r')]. \tag{3.2}$$

Here summation over κ is performed, as we have noted, assuming that this variable covers the entire Brillouin zone. Conversely, summation over the variable k is bounded by a certain region $\{K\}$ typical of each quasi-momentum of the pair K which should correspond to those hole quasi-momenta k_1 and k_2 which in accordance with (2.6) are situated in the region $\{F\}$, i.e., outside the Fermi contour for the holes. The compact region $\{K\}$ is thus a certain set of points belonging to that part of the 2D Brillouin zone not filled with holes. The dimensions and shape of the regions $\{K\}$ depend on the position of the vector K relative to the Fermi contour.

Let Ξ_K be the area of the region $\{K\}$ of permissible values of the quasi-wave vector of the relative motion for given K. Then, if this region is fairly large, we can approximately assume

$$\frac{1}{S}\sum_{k\in\{K\}} \exp[ik(r-r')] \approx \Xi_K a^2 \delta(r-r'), \tag{3.3}$$

where the coefficient of the delta function is of the order of magnitude of the ratio of Ξ_K to the area of the 2D Brillouin zone (a is the interatomic distance in the plane of the conducting layer). Thus, the potential energy operator of the holes forming a pair can be given as

$$\hat{U}_K \approx \Xi_K a^2 U(r). \tag{3.4}$$

The factor $\Xi_K a^2 < 1$ can be considered to be a certain correlation attenuation factor [28] attributable to the Pauli principle.

If the region $\{K\}$ is small, the sum in (3.3) contains a few oscillating terms and the operator $\hat{U}$ becomes essentially nonlocal. The Fourier components appearing in the kernel of this operator may have different signs at similar points in r-space and for this reason, when analyzing the interaction between holes which leads to the formation of quasi-steady states of hole pairs, we can confine ourselves to pairs having quasi-momenta K for which the region of permissible values of the vector k is fairly large (note that these regions are situated near almost rectilinear sections of the Fermi contour). In this case, the equivalent Hamiltonian of a hole pair having the quasi-momenta K has the form

$$\hat{H}_K = \varepsilon\left(\frac{K}{2} - i\nabla\right) + \varepsilon\left(\frac{K}{2} + i\nabla\right) + \Xi_K a^2 U(r). \tag{3.5}$$

The probabilities of the filling of hole states in the range $\{\bar{F}\}$ are determined by the scattering amplitudes $\tilde{\varphi}_K(k, q)$. Here, as we noted above $k \in \{F\}$ and $q \in \{\bar{F}\}$. In general, the scattering amplitudes are largest when k and q are situated in the immediate vicinity of the Fermi contour. Assuming that the characteristic width Δk of the most probable hole distribution band adjacent to the Fermi contour in the region $\{F\}$ is determined by the potential (2.17) where the value of the argument corresponds to the average distance between the hole $\bar{r} = p^{-1/2}$, where p is the concentration of 2D holes, we can easily confirm that

$$\Delta k \approx \frac{e^2}{\hbar|v_t(\mu)|\bar{r}}\exp\left(-\frac{\bar{r}}{r_0}\right), \tag{3.6}$$

where $v_t(\mu)$ is the component of the hole velocity perpendicular to the Fermi contour. Using an anisotropic approximation of the dispersion law for order-of-magnitude estimates we can obtain $v_t(\mu) \sim \hbar k_F/m$ where $k_F = (2\pi p)^{1/2}$ is the value of the Fermi wave vector, and m is a certain characteristic effective mass. As an estimate for the density of states we can use the expression $g(\mu) \approx m/\pi\hbar^2 d$, where d is the distance between neighboring conducting layers. Also using the notation $a^* = \hbar^2/me^2$, we have

$$\Delta k \sim \frac{1}{\sqrt{2\pi}a^*}\exp\left(-\frac{2}{\sqrt{a^* dp}}\right). \tag{3.7}$$

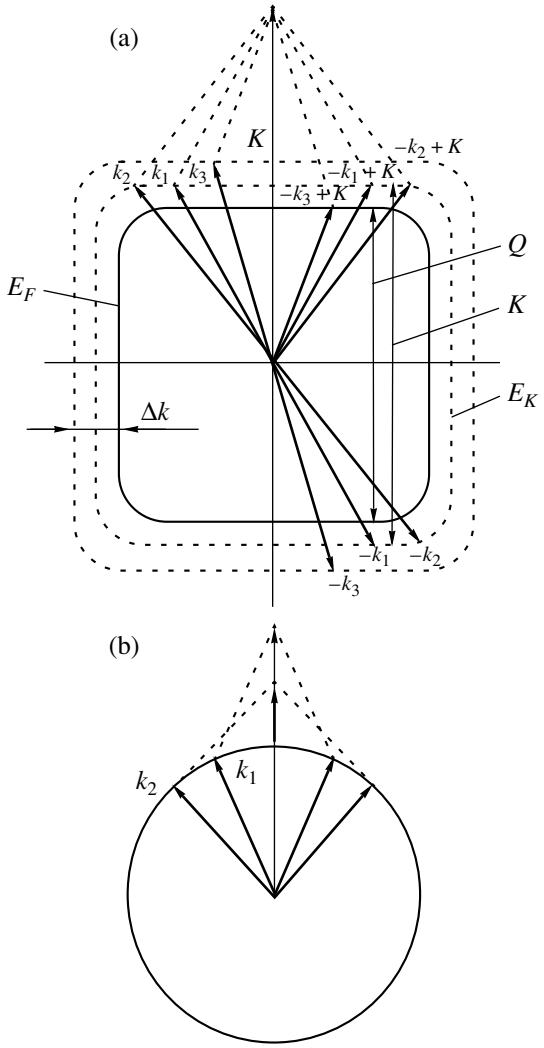

Fig. 1. Determination of regions of permissible values of the quasi-momenta k_i of holes forming a pair having the quasi-momentum K for constant-energy lines having the form of squares with rounded corners (a) and for circles (b).

Assuming that all important scattering processes leading to the formation of quasi-steady states take place in a band of width Δk adjacent to the Fermi contour, we can take the intersection of this band and the region $\{K\}$ determined above as the region of determination of the quasi-momentum of the relative motion. Figure 1 illustrates the principle of determining this region of permissible values of k_1 and k_2. The region thus determined will then be denoted by $\{K\}$ as before and its area by Ξ_K since for most vectors $K/2$ we can assume that the width of the region $\{K\}$ is the same as Δk.

The region $\{K\}$ is a narrow strip (Fig. 1a) if $K/2$ is located near an almost rectilinear section of the Fermi contour and the length of this strip l_K depends on the orientation of the vector K, i.e., $l_K = l_K(\vartheta)$, where ϑ is the polar angle of the vector K. It is easy to establish that in *p*-type HTSC cuprates where the Fermi contour is a square with rounded corners [6–10], l_K reaches a maximum when the direction of K coincides with the directions of the k_x and k_y coordinate axes parallel to the long, almost rectilinear parts of the Fermi contour. As ϑ increases from 0 to $\pi/4$, the value of l_K decreases, reaching a minimum at $\vartheta = \pi/4$. Thus, $l_K(\vartheta)$ is a periodic function having the period $\pi/2$. Near the rounded corners of the square, the area of the region $\{K\}$ is small and this region may have a fairly complex profile. In this case, the effective interaction between the holes forming the pair is small so that the formation of quasi-steady states is exclusively attributed to the fact that the isolines near the Fermi contour have long almost rectilinear sections. The area of the region $\{K\}$ can then be estimated as $\Xi_K \approx l_K \Delta k$; together with $l_K(\vartheta)$ it depends on the polar angle of the vector K.

The explicit form of the function $l_K(\vartheta)$ is determined by the structure of the constant-energy lines near the Fermi contour. When the constant-energy lines are squares with rounded corners, neglecting the rounded edges, i.e., assuming that the constant-energy lines are simply squares, we can write $l_K(\vartheta)$ explicitly as $l_K(\vartheta) = 2k_F(1 - \tan\vartheta)$. Here k_F is half the side of the square forming the Fermi contour $0 \le \vartheta \le \pi/4$. Bearing in mind that $\Delta k \ll k_F$, for $0 \le \vartheta \le \pi/4$ we can approximately write $l_K(\vartheta) \approx K_x - K_y$; for $\pi/4 \le \vartheta \le \pi/2$ however, we have $l_K(\vartheta) \approx K_y - K_x$. Thus, $l_K(\vartheta)$ is determined for $0 \le \vartheta \le \pi/2$ and the values of this function for an arbitrary polar angle can be obtained from the periodicity condition: $l_K(\vartheta + \pi/2) = l_K(\vartheta)$.

It can be seen from Fig. 1a that the value of K should be larger than Q. Then, first all pairs of particles having the wave vectors k_i and $-k_i + K$ corresponding to the rectilinear section with energy ε_K have the same total momentum K. In addition, each state in the layer Δk having energy higher than ε_K (for example, having the wave vector k_3) has a partner with $\varepsilon_F < \varepsilon < \varepsilon_K$ (with $k_4 = -k_3 + K$) to ensure the same total wave vector K. Thus, a given K corresponds to a macroscopic number of states with the area Ξ. For a given doping level there is an optimum value of the four vectors K perpendicular to each side of the Fermi contour. For the vector K equal to Q the corresponding area is zero.

Consequently, when the Fermi contour has rectilinear sections, hole pairs apparently focused in momentum space (focused pairs) are formed. This focusing corresponds to the macroscopic filling of one state for given K at the same energy. This focusing cannot be achieved for example for a Fermi contour in the form of a circle (see Fig. 1b).

The formation of an electron pair (inside the Fermi contour for the holes) corresponds to a different value of K and different values for the binding energy because of the lack of symmetry of the constant-energy profiles and the effective masses relative to the Fermi contour. This evidently results in a lack of symmetry of

the states below and above the Fermi level in the superconducting phase which should be observed as asymmetry, $V \longrightarrow -V$, of the tunnel characteristics, and Andreev reflection characteristics.

Figure 2 shows the 2D Brillouin zone of a crystal possessing tetragonal symmetry and shows the distribution of the isolines calculated in the strong coupling approximation allowing for interaction in the first, second, and third coordination spheres. The dimensionless parameters of the dispersion law

$$E(k_x, k_y) = 2 - 2t_1(\cos k_x a + \cos k_y a) - 4t_2 \cos k_x a \cos k_y a - 2t_3(\cos 2k_x a + \cos 2k_y a) \quad (3.8)$$

were selected so that [23] the profile of the isolines reflected the observed profile of the Fermi contour of HTSC cuprates [6–10]: $t_1 = 0.50$, $t_2 = -0.15$, $t_3 = 0.10$. The Brillouin zone is centered at the point $(\pi/a, \pi/a)$ in reciprocal space. The point M at the center of the square is the absolute maximum of the electron band (minimum of the hole band) and m and m' are the absolute and side minima (maxima) of the electron (hole) bands. The saddle points S are connected by the isoline (separatrix) separating the families of closed constant-energy lines.

Near the half-filling isoline there is a fairly extensive region $\{S\}$ embracing the saddle points in which the principal values of the tensor of reciprocal effective masses have different signs (light region in Fig. 2). This region should have vectors $K/2$ which in principle can correspond to relatively long-lived quasi-steady states.

The width of the most probable hole distribution band in accordance with (3.7) increases with increasing hole concentration. Thus, as p increases and, under conditions of isoline nesting, the number of hole pairs capable of forming quasi-steady states can increase. An increase in the hole concentration generally leads to a nonmonotonic dependence $l_K(p)$: if the isolines have relatively long, almost rectilinear sections in a fairly large part of the Brillouin zone, initially as p increases, the function $l_K(p)$ can increase slowly or remain almost constant. A further increase in the hole concentration brings the Fermi level to severely distorted isolines, causing an abrupt decrease in the regions $\{K\}$ or even transferring them outside the region $\{S\}$ so that $l_K(p)$ becomes a fairly rapidly decreasing function, as is illustrated in Fig. 3 for the dispersion law (3.8). This figure gives the average length of the $\{K\}$ band $\langle l_K \rangle$ as a function of the hole energy in the region $\{S\}$ where the effective mass components have different signs. Also plotted for comparison are data for the dispersion law obtained in the strong coupling approximation allowing only for the interaction of nearest neighbors (dashed curve). It can be seen that the dispersion law (3.8) corresponds to isoline nesting in the region $\{S\}$ over a wide range of energy. Thus, as the doping level is varied when the Fermi level systematically passes from one isoline to another, the nesting conditions for the Fermi contour are conserved over a fairly wide range of concentrations. Figure 3 clearly shows the asymmetry of the dispersion law relative to the type of doping: nesting conditions are only observed for p-type cuprates.

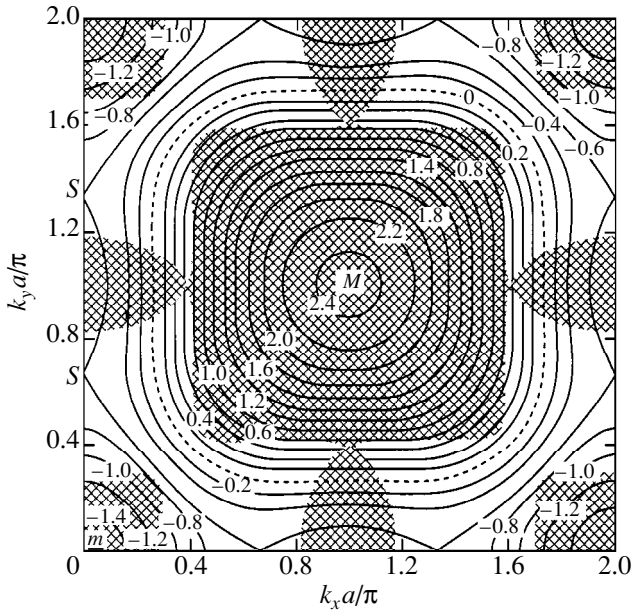

Fig. 2. Distribution of constant-energy lines and singular points in the 2D Brillouin zone of a crystal with tetragonal symmetry. The light regions correspond to different signs of the principal values of the reciprocal effective mass tensor of the dispersion law (3.8).

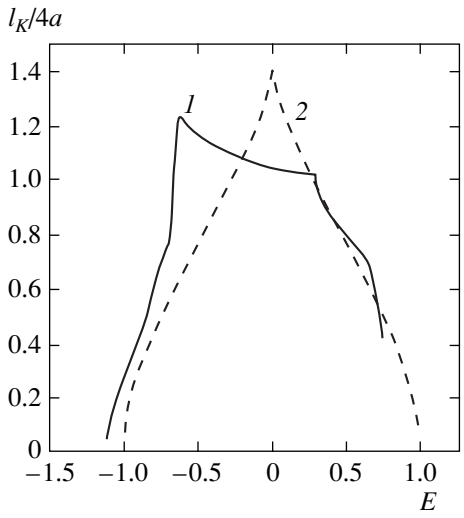

Fig. 3. Dependence of the characteristic length of the band of permissible quasi-momentum values of holes forming a pair on the hole energy in the Brillouin zone corresponding to different signs of the principal values of the reciprocal effective mass tensor. The solid curve gives the 2D dispersion law (3.8). Also plotted for comparison is the similar dependence corresponding to the simplest dispersion law with $t_n = 0$ for $n \geq 2$ (nearest-neighbor approximation, dashed curve).

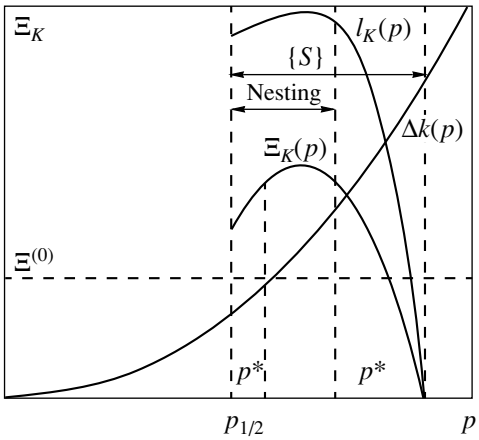

Fig. 4. Area of the region of permissible values of the quasi-momenta $\Xi_K(p)$, and the values l_K and Δk for holes forming a pair as a function of the hole concentration (schematic).

However, the simplest dispersion law which only allows for nearest neighbors is symmetric with respect to the type of doping although, exhibiting ideal nesting at half filling, it leads to an abrupt decrease in the value of $\langle l_K \rangle$ with increasing distance from the center of the zone.

The area $\Xi_K(p)$ as a function of the hole concentration thus has the form of a curve with a maximum (Fig. 4).

In the effective mass approximation we need to confine ourselves to the second derivatives in the expansion of the equivalent Hamiltonian (3.5) in powers of ∇; the coefficients of the expansion of the kinetic energy operator form the tensor of reciprocal reduced effective masses for the point $K/2$. Transforming this tensor to a diagonal form (i.e., suitably selecting the directions of the k_1 and k_2 coordinate axes), we have

$$\varepsilon\left(\frac{K}{2} - i\nabla\right) + \varepsilon\left(\frac{K}{2} + i\nabla\right) \approx 2\varepsilon\left(\frac{K}{2}\right) - \frac{\hbar^2}{2m_1^{(K)}}\frac{\partial^2}{\partial x_1^2} - \frac{\hbar^2}{2m_2^{(K)}}\frac{\partial^2}{\partial x_2^2}, \tag{3.9}$$

where $m_\alpha^{(K)}$ are the principal values for the effective masses relating to the point $K/2$.

Provided that $\Delta k \ll K/2 \sim k_F$ we can assume that $\varepsilon(K/2)$ approximately coincides with the Fermi level: $\varepsilon(K/2) \approx \mu$. Then, if we stipulate that the energy of the hole pair is measured from the value 2μ, the equivalent Hamiltonian (3.5) may be expressed as

$$\hat{H}_K = -\frac{\hbar^2}{2m_1^{(K)}}\frac{\partial^2}{\partial x_1^2} - \frac{\hbar^2}{2m_2^{(K)}}\frac{\partial^2}{\partial x_2^2} + \Xi_K a^2 U(x_1, x_2). \tag{3.10}$$

It has been noted that for quasi-momenta K for which relatively long-lived quasi-steady states of hole pairs can appear, the reduced effective masses $m_1^{(K)}$ and $m_2^{(K)}$ have different signs.

4. QUASI-STEADY STATES OF HOLE PAIRS

In the effective mass approximation, the envelope functions of the free motion of a hole pair can be selected in the form of plane waves where the energy of this pair is expressed in terms of the components of the quasi-wave 2D vector q_1 as

$$E(K, q) = \frac{\hbar^2}{2m}\sum_{l=1}^{2} \nu_K^{(l)} q_l^2; \tag{4.1}$$

here $\nu_K^{(l)}/m = 1/m_l^{(K)}$ are the principal values of the tensor of the reciprocal effective masses (the dimensionless parameters $\nu_K^{(l)}$ in the region $\{S\}$ have different signs), m is a parameter having dimensions of mass, $l = 1, 2$. The energy measured from twice the chemical potential 2μ can therefore have either sign. Thus, the spectrum of the operator (3.10) in the region $\{S\}$ is not bounded.

Since the real interaction between holes is screened Coulomb interaction (2.17), we can approximately assume that the holes interact by means of the short-range singular potential

$$U(r) = 2\pi e^2 r_0 \delta(r). \tag{4.2}$$

The relative motion of a hole pair is infinite, as should be the case for a continuous spectrum. This does not eliminate the possible formation of relatively long-lived states assigned to the continuous spectrum and manifest as fairly well-defined density-of-states maxima. The resulting quasi-steady state is characterized by the energy E_0 corresponding to the density-of-states maximum and the finite lifetime τ. We can therefore assume [29] that this state corresponds to the complex energy $\tilde{E}_0 = E_0 - i\Gamma$ with the damping $\Gamma = \hbar/\tau$ and is observed as the pole of the scattering amplitude at energy $E = \tilde{E}_0$.

The wave function corresponding to the scattering has the form

$$\varphi_{Kq}(r) = \exp(iqr) + \chi_{Kq}(r), \tag{4.3}$$

where the Fourier transformation of the scattered wave for scattering at the δ-potential is expressed as

$$\tilde{\chi}_{Kq}(k) = -\frac{w_K}{1 + w_K B_K(\omega)}\frac{1}{k^2 - \omega - i0\,\mathrm{sgn}\,\omega}. \tag{4.4}$$

Here we have

$$w_K = \frac{4\pi r_0}{a^*}\Xi_K a^2, \quad k^2 = \sum_{l=1}^{2} \nu_K^{(l)} k_l^{(2)},$$

$$\omega = \sum_{l=1}^{2} \nu_K^{(l)} q_l^2,$$

the sign function $\operatorname{sgn}\omega$ provides the necessary condition for $\tilde{\chi}_{Kq}(k)$ to correspond to a diverging (or damped) wave,

$$B_K(\omega) = \int \frac{1}{k^2 - \omega - i0\operatorname{sgn}\omega} \frac{d^2k}{(2\pi)^2} \equiv B_{K1}(\omega) + iB_{K2}(\omega). \tag{4.5}$$

The real and imaginary parts of the function (4.5) are expressed as

$$B_{K1}(\omega) = P\int \frac{1}{k^2 - \omega} \frac{d^2k}{(2\pi)^2} \tag{4.6}$$

(P is the symbol denoting the principal value of the integral),

$$B_{K2}(\omega) = \pi\int \delta(\omega - k^2) \frac{d^2k}{(2\pi)^2} \operatorname{sgn}\omega. \tag{4.7}$$

The denominator of the scattering amplitude generally does not go to zero since it is a sum of two independent functions. An exception may be the case when one of the functions and specifically $B_{K2}(\omega)$ goes identically to zero. The poles of the scattering amplitude determined from the equation

$$1 + w_K B_{K1}(\omega) = 0, \tag{4.8}$$

then correspond to bound states. For $w_K > 0$ this situation for a hole pair can occur near the maximum M of the hole band.

If equation (4.8) has a solution for some (complex) value $\omega = \omega_K^0$ and $B_{K2}(\omega_K^0) \neq 0$, then provided that $B_{K2}(\omega_K^0)$ is fairly small and positive, a quasi-steady hole-pair state appears. Near the point ω_K^0 the function $B_{K1}(\omega)$can be expressed as (the prime denotes differentiation with respect to ω)

$$B_{K1}(\omega) \approx B_{K1}(\omega_K^{(0)}) + B'_{K1}(\omega_K^{(0)})(\omega - \omega_K^{(0)}). \tag{4.9}$$

The scattering amplitude is expressed in the form

$$f_K(E) = -\frac{w_K}{1 + w_K B_K(\omega)} \approx -\frac{\hbar^2}{2m} \frac{1}{B'_{K1}(\omega_K^{(0)})} \frac{1}{E - E_K^{(0)} + i\Gamma_K/2}. \tag{4.10}$$

Here $E_K^{(0)} = \hbar^2\omega_K^{(0)}/2m$ is the energy of the quasi-steady state, $E = \hbar^2\omega/2m$, and the damping has the form

$$\Gamma_K = \frac{\hbar^2}{m} \frac{B_{K2}(\omega_K^{(0)})}{B'_{K1}(\omega_K^{(0)})}. \tag{4.11}$$

For fairly large ω, (4.6) is a monotonically increasing function; in accordance with the definition (4.7) we have $B_{K2}(\omega) > 0$ which gives a solution of equation (4.8) corresponding to positive damping of the quasi-steady state.

In order to identify the nature of the possible solution of equation (4.8) we consider the region $\{K\}$ near the almost rectilinear section of the Fermi contour and assume for simplicity that the point $K/2$ is close to one of the symmetric points of the Brillouin zone, for example, adjacent to the axis of symmetry. Then the direction of one of the principal axes of the reciprocal effective mass tensor (the k_1 axis) is almost parallel to the rectilinear section of the Fermi contour; the direction of the second principal axis (k_2) is then almost perpendicular to the Fermi contour. Since the directions of the rectilinear section of the constant-energy line and the k_1 axis almost coincide, the corresponding effective mass $m_1^{(K)} = m/\nu_K^{(1)}$ has a high absolute value; it can be seen from Fig. 2 that this is positive. Conversely the effective mass $m_2^{(K)} = m/\nu_K^{(2)}$ along the k_2 axis is negative and $\nu_K^{(2)} < 0$. For estimates we can assume that $\nu_K^{(2)} = -1$, $\nu_K^{(1)} \equiv \nu \ll 1$.

It follows from the definition (4.7) that for any $\omega > 0$ and $\nu = 0$ we have the equality $B_{K2}(\omega) = 0$ and in order to obtain a finite damping value we need to go outside the effective mass approximation. For $\nu \ll 1$ there is consequently a specific range of ω in which the quasi-steady state damping is weak. It follows directly from the definition (4.7) that this region corresponds to $\omega \geq \nu l_K^2/4$, where l_K is the previously determined characteristic size of the region $\{K\}$ along the Fermi contour. It can also be seen that for $\omega < \nu l_K^2/4$ we have to within logarithmic accuracy $B_{K2} \sim \nu^{-1/2}$ so that under the assumption $\nu \ll 1$ the damping (4.11) for $\omega < \nu l_K^2/4$ is definitely large.

The function $B_{K1}(\omega)$ cannot be represented as a combination of elementary functions although for fairly high energies and specifically for $\omega - \nu l_K^2/4 \gg (\Delta k)^2/4$ this function may be approximately expressed as

$$B_{K1}(\omega) \approx -\frac{\Xi_K}{(2\pi)^2} \frac{1}{\omega}. \tag{4.12}$$

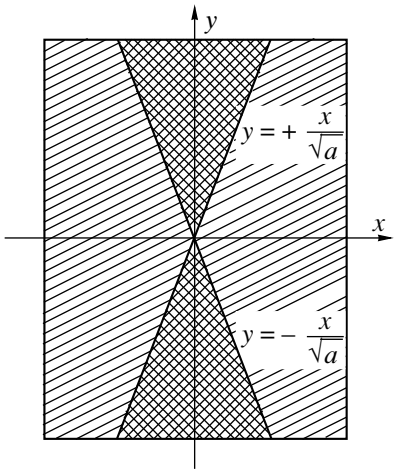

Fig. 5. Character of the wave function of the relative motion of the pair (5.1) corresponding to a diverging (light region) and damped (dark region) wave. The lines separating these regions are caustics.

It can be confirmed that this expression qualitatively reflects the behavior of the function (4.6) for all $\omega \geq \nu l_K^2/4$, i.e., in that range of energies which corresponds to fairly weak damping of the quasi-steady state. We obtain the energy of this state which depends on the quasi-momentum of the pair K in the form

$$E_K^{(0)} = \frac{e^2 r_0}{2\pi}(\Xi_K a)^2, \tag{4.13}$$

or, taking into account (3.7),

$$E_K^{(0)} = \frac{e^2}{2a^*}\frac{(l_K a)^2}{4\pi}\left(\frac{d}{a^*}\right)^2 \exp\left(-\frac{4}{\sqrt{pa^* d}}\right). \tag{4.14}$$

Thus, the energy of the quasi-steady state is expressed in terms of the parameters of the electron spectrum of the HTSC compound and its doping level. Assuming that in *p*-type HTSC cuprates an extensive vicinity of the saddle point corresponds to relatively weak dispersion of the hole band [11], i.e., a fairly high effective mass m, we can assume for a rough estimate that the effective Bohr radius a^* is of the same order of magnitude as the interatomic distance: $a^* \sim a \sim d$. In this case, for a hole concentration corresponding to a half-filled band $pa^2 = 1$, the exponential factor in (4.14) is around 2×10^{-2} and the effective Rydberg $e^2/2a^*$ is around 2.5–3.0 eV. Consequently, the quasi-steady energy is mainly determined by the parameter l_K which for long almost rectilinear sections of the Fermi contour may reach values of the order of π/a. In these cases, the quasi-steady state energy may be between 10 and 100 meV. It follows from (4.14) that longer regions $\{K\}$ produced by nesting near the Fermi contour play a dominant role in its formation: a decrease in l_K leads to an abrupt reduction in the energy of this state.

5. WAVE FUNCTION OF HOLE PAIR

The function of the relative motion of the pair (4.4) corresponding to a diverging wave has the following form in the r-representation:

$$\chi_{Kq}^{(+)}(r) = \frac{1}{4\sqrt{\nu}} \times \begin{cases} H_0^{(2)}\left(q\sqrt{\frac{x^2}{\nu} - y^2}\right), & 0 < |y| < \frac{|x|}{\sqrt{\nu}}, \\ \frac{2i}{\pi}K_0\left(q\sqrt{y^2 - \frac{x^2}{\nu}}\right), & \frac{|x|}{\sqrt{\nu}} < |y| < \infty. \end{cases} \tag{5.1}$$

Here $H_0^{(2)}(z)$ is a Hankel function and $K_0(z)$ is a Macdonald function. Thus, the plane x, y is divided into four sections, two of which $0 < |y| < |x|/\sqrt{\nu}$ correspond to waves diverging from the origin; in the other two sections $|x|/\sqrt{\nu} < |y| < \infty$ the wave function (5.1) is damped (Fig. 5). The lines $y = \pm x/\sqrt{\nu}$ dividing these sectors are caustics near which $\chi_{Kq}^{(+)}(r)$ has a logarithmic singularity.

The function $\chi_{Kq}^{(+)}(r)$ and the linearly independent function $\chi_{Kq}^{(-)}(r)$ corresponding to a wave converging toward the center are eigenfunctions of the equivalent Hamiltonian (3.10) with the singular potential (4.2) relating to complex eigenvalues. Thus, in order to study the spatial distribution of the hole density at an arbitrary time we need to solve a time-dependent equation with the Hamiltonian (3.10). However, such an exhaustive solution cannot be obtained by separating the time and space variables [29]. Consequently, in order to make a qualitative analysis of the nature of the wave function over the quasi-steady state lifetime of a hole pair, we can use the adiabatic approximation [30] which separates the "fast" motion along y and the "slow" motion along x. In fact, for $t \leq \tau$ the quasi-steady state can be considered to an almost steady state having the energy $E_K^{(0)}$; in addition, as a result of the extremely strong anisotropy of the effective masses when $m_1 = m/\nu$, $m_2 = -m$, and $\nu \ll 1$ a similar approximate separation of the motion along x and y becomes possible.

We express the envelope function which is a solution of the Schrödinger equation with the Hamiltonian (3.10) in the form $\psi(x, y) = \varphi(x, y)\Phi(x)$ where the function $\varphi(x, y)$ which depends on x and on the parameter is one of the solutions of the equation

$$\left[\frac{\hbar^2}{2m}\frac{\partial^2}{\partial y^2} + \Xi_K a^2 U(x, y)\right]\varphi_n(x, y) = V_n(x)\varphi_n(x, y), \tag{5.2}$$

where n numbers the eigenvalues $V_n(x)$ of the operator on the left-hand side of (5.2). Quite clearly, at least one of these eigenvalues which we shall denote by $V_0(x)$ is nonnegative. We shall confine ourselves to this case and write the equation for $\Phi(x) = \Phi_0(x)$:

$$\left[-\nu\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V_0^*(x)\right]\Phi_0(x) = E\Phi_0(x), \tag{5.3}$$

where the effective potential for the slow motion is

$$V_0^*(x) = V_0(x) + \nu\frac{\hbar^2}{2m}\int_{-\infty}^{\infty}\left(\frac{\partial\varphi_0(x,y)}{\partial x}\right)^2 dy. \tag{5.4}$$

Since $V_0^*(x) > 0$, the function $\Phi_0(x)$ corresponds to the continuous spectrum, i.e., may be expressed as a linear combination of converging and diverging waves. Conversely for any $U(x, y) > 0$ and any x the function of the one-dimensional motion $\varphi_0(x, y)$ corresponds to the steady state of the discrete spectrum and may be normalized by the condition

$$\int_{-\infty}^{\infty}[\varphi_0(x,y)]^2 dy = 1. \tag{5.5}$$

This implies that in a certain region of the x axis and for a specific energy the envelope function $\Phi_0(x)$ may increase sharply which corresponds to the quasi-steady state [29].

Characteristic features of the function $\varphi_0(x, y)$ can be established if we express the interaction in (5.2) in the form of the separable potential

$$U(x, y) = U_0 r_0^2 f(x) f(y), \tag{5.6}$$

where the function $f(x)$ is assumed to be bounded and normalized to unity. The characteristic spatial scale for this function is the screening radius r_0 and the parameter $U_0 = (2\pi e^2/r_0)\Xi_K a^2$. Assuming for simplicity that $f(y) = (2r_0)^{-1}$ for $|y| < r_0$ and $f(y) = 0$ for $|y| > r_0$, we can easily find $\varphi_0(x, y)$ for $|y| > r_0$, i.e., outside the radius of action of the potential (5.6):

$$\varphi_0(x, y) = [\gamma(x)]^{1/2}\exp(-\gamma(x)|y|). \tag{5.7}$$

Here we have $\gamma(x) = \sqrt{2mV_0(x)}/\hbar$. We assume that the potential $U(x, y)$ leads to a single discrete level as in (4.2). The position of this level can then be approximately estimated as [31]

$$\gamma(x) \approx \frac{2mr_0^2}{\hbar^2}U_0 f(x). \tag{5.8}$$

The estimate (5.8) is obtained if $2mr_0^2 U_0/\hbar^2 < 1$. This inequality can be rewritten in a form such that it is obvious if it is satisfied:

$$\frac{2\pi e^2}{r_0}\Xi_K a^2 < \frac{\hbar^2}{2mr_0^2}. \tag{5.9}$$

In fact, $2\pi e^2/r_0$ and $\hbar^2/2mr_0^2$ have the same order of magnitude and the inequality $\Xi_K a^2 \ll 1$ is satisfied by definition.

The function (5.7) is normalized so that its exact form for $|y| < r_0$ is unimportant for the following estimates. Assuming that $f(x)$ varies rapidly [between $(1/2)r_0$ and zero] on the scale r_0 (this characteristic scale of the preexponential factor in (5.7) characterizes the localization of the quasi-steady state along the x-axis), it follows from (5.7) that for $|x| \le r_0$ the characteristic scale along the y axis for the function $\varphi_0(x, y)$ is $\gamma^{-1}(0) = \hbar^2/2mr_0U_0 \sim r_0/\Xi_K a^2 \gg r_0$. We can assume that during the lifetime of the quasi-steady state the hole pair is localized in the region $|x| \le r_0$, $|y| \le r_0/\Xi_K a^2$. These inequalities determine the size of the region in r-space in which the motion of the two holes can be considered to be essentially correlated. Thus, the correlation length for a hole pair having the quasi-momentum K is anisotropic.

A consequence of the crystal symmetry is that all the wave functions $\tilde{\chi}_{\hat{g}Kq}(k)$ corresponding to the set of vectors $\hat{g}K$ ($\hat{g}$ is the transformation from the crystal symmetry group which converts the vector K to its equivalent) which form the star of the vector K are equivalent. Hence the wave function of the pair reflecting the crystal symmetry should be expressed as the linear combination

$$\tilde{\Phi}_{Kq}^{(\Gamma)}(k) = \sum_{\{\hat{g}K\}} c_{\hat{g}K}^{(\Gamma)}\tilde{\chi}_{\hat{g}Kq}(k), \tag{5.10}$$

where the choice of coefficients $c_{\hat{g}K}^{(\Gamma)}$ is determined by the irreducible representation of the Γ crystal symmetry group whose basis function is (5.10). This function corresponds to zero quasi-momentum of the pair since $\sum_{\{\hat{g}K\}}\hat{g}K = 0$.

For crystals possessing tetragonal symmetry (C_{4m} 2D symmetry group) all the elements of the star of the arbitrary vector K can be divided into two sets. One of these, which includes the vector K, contains the vectors $\hat{g}K$ which are interrelated by reflections relative to the k_x and k_y coordinate axes; the second is generated similarly by a vector obtained from K as a result of reflection from the diagonal of the square representing the 2D Brillouin zone. The trivial representation A_{1g} corresponds to the coefficients $c_{\hat{g}K}^{(A_{1g})}$ which are the same for all the star vectors. For the B_{1g} representation the coefficients $c_{\hat{g}K}^{(B_{1g})}$ have the same absolute values and differ

in sign for the vectors $\hat{g}K$ belonging to different groups. Taking into account the explicit form (4.4) of the functions $\tilde{\chi}_{\hat{g}Kq}(k)$ we can easily confirm that with a suitable choice of coordinates all the vectors $\hat{g}K$ from the first set correspond to $k^2 = \nu k_x^2 - k_y^2$ whereas the vectors of the second set correspond to $k^2 = \nu k_y^2 - k_x^2$. Thus, the function (5.10) corresponding to the irreducible representation A_{1g} for $\nu \ll 1$ has the form

$$\tilde{\Phi}_{Kq}^{(\Gamma)}(k) \sim -\frac{2q^2}{(q^2 + k_x^2)(q^2 + k_y^2)} \tag{5.11}$$

and exhibits complete symmetry with respect to all transformations of the crystal symmetry group, i.e., can be arbitrarily assigned to the *s*-type of orbital symmetry. For the function corresponding to the irreducible representation B_{1g} for $\nu \ll 1$ we have

$$\tilde{\Phi}_{Kq}^{(\Gamma)}(k) \sim -\frac{k_x^2 - k_y^2}{(q^2 + k_x^2)(q^2 + k_y^2)}, \tag{5.12}$$

and this function can be arbitrarily assigned to *d*-type orbital symmetry.

We shall consider a pair with given K. Its characteristic dimensions along the x and y axes in r-space are r_0 and $r_0/\Xi_K a^2$, respectively. Thus, in k-space the dimensions of the pair along the k_x and k_y axes are of the order of magnitude r_0^{-1} and $\Xi_K a^2/r_0$, respectively. The area in k-space assigned to a single pair with given K can be estimated as $\Xi_K(a/r_0)^2 \sim \Xi_K$ since at hole concentrations corresponding to less than half filling we can assume $r_0 \sim a$. Consequently, over the entire range of determination of the quasi-momentum of the relative motion of the pair $\{K\}$ there is in fact only one pair. This conclusion naturally holds for the normalized correct wave function of the pair (5.10) which corresponds to zero quasi-momentum. The correlation length can then be estimated as $\xi \sim \Xi_K^{-1/2}$ and has upper and lower constraints: $\xi_m \le \xi \le \xi_M$; here $\xi_M \sim (\Xi^{(0)})^{-1/2}$ and the minimum value of ξ_m is determined by the largest length l_K of the almost rectilinear section of the Fermi contour for a given doping level. Quite clearly the correlation length may be of the order of magnitude of only a few interatomic distances.

The formation of a quasi-steady hole pair state results in the removal of the quasi-momenta of the holes $k_1, k_2 \in \{K\}$ so that the density of single-particle states near the Fermi level decreases, which may be interpreted as the formation of a pseudogap in the spectrum of single-particle excitations.

The energy of the pair (4.13) forming a quasi-steady state for given K includes the kinetic energy of the relative motion of the holes and their Coulomb interaction energy. Let us assume that N_{pair} holes pairs forming such a state has appeared in the system. The correlation energy of these quasiparticles can be estimated as the sum of the energy of the noninteracting pair gas $E_0 N_{\text{pair}}$ where $E_0 = (e^2 r_0/2\pi)(\Xi a)^2$ is a certain average value of (4.13), and the pair interaction energy $E^{(B)}$. The latter can be estimated in the first order with respect to the hole concentration if we consider the N_{pair} hole pairs as a weakly nonideal Bose gas [32]:

$$E^{(B)} = \frac{1}{2} N_{\text{pair}} n_{\text{pair}} U^{(B)} (\Xi a^2)^2. \tag{5.13}$$

Here n_{pair} is the pair concentration, the energy parameter of the interaction can be estimated as $U^{(B)} \approx 4\pi(2e)^2 r_0^2$, and the inclusion in (5.13) of the second power of the correlation attenuation factor allows for the fact that processes of simultaneous scattering of two holes in states already filled with other holes should be excluded. In the absence of pairing, $2N_{\text{pair}}$ holes may be considered as a weakly ideal Fermi gas since in this case (4.13) gives

$$E^{(F)} = \frac{1}{4} N_{\text{pair}} n_{\text{pair}} U^{(F)} (\Xi a^2) \tag{5.14}$$

(for estimates we use the expression for the energy of a weakly nonideal Fermi gas with short-range repulsion between particles [32] in the first order in terms of the particle concentration). Here the energy parameter of the interaction between two holes is estimated as $U^{(F)} \approx 4\pi e^2 r_0^2$ and the correlation attenuation factor as in (3.10) appears in the first order since we only allow for the relative scattering processes of two holes. Since $\Xi a^2 \ll 1$, it is clear that $E^{(B)} \ll E^{(F)}$ and the formation of a quasi-steady state of hole pairs is energetically favorable when $E_0 N_{\text{pair}} < E^{(F)}$, which can be rewritten as

$$n_{\text{pair}} > \frac{1}{2\pi^2} \frac{\Xi}{r_0}. \tag{5.15}$$

Bearing in mind that $\Xi^{-1/2}$ is the characteristic size of the quasi-steady state in the plane of the conducting layer, we can interpret inequality (5.15) as follows: the formation of this hole pair state leads to a reduction in the ground-state energy when their two-dimensional concentration begins to exceed a value of the order of Ξ, i.e., when the pairs begin to overlap in r space.

The gain in energy E_0 corresponding to the formation of a hole pair with repulsive interaction and negative effective mass is compensated by a reduction in the positive contribution to the energy for a finite pair concentration as a result of the attenuation of the interpair correlations compared with the positive contribution to the energy made by the correlations of holes not coupled to form pairs (fermions).

The basis for using the conclusion that the correlation energy is positive in the model of a weakly nonideal gas with short-range repulsion [32] is as follows.

All the properties of cuprate HTSCs are determined by a low concentration of holes in the doping, since dielectric correlations are conserved up to optimum doping. The screening radius r_0 is determined by the high concentration of metallic forephase carriers if the correlation dielectric length is substantially greater than r_0.

In this approach the superconducting state must be described as the Bose condensation of these focused pairs. The superconducting order parameter will be determined by the anomalous mean $\langle c_k^+ c_{-k+K}^+ \rangle$, i.e., will correspond to a state of the type [33, 34] with a nonzero total pair momentum. However, under the same conditions for the profile of the Fermi surface for which focused pairs may appear, it is also possible for electron–hole pairs to appear, having the order parameter $\langle c_k^+ c_{k+K_1} \rangle$ and the total momentum K_1 close to K. Making the necessary allowance for both the parameters $\langle c_k^+ c_{-k+K}^+ \rangle$ and $\langle c_k^+ c_{k+K_1} \rangle$ should lead to the appearance of a parameter $\langle c_k^+ c_{-k+K-K_1}^+ \rangle$, which for $K \approx K_1$ corresponds to Cooper pairing. The role of the order parameter in doping corresponding to the superconducting region may be played by the near antiferromagnetic order. At present there is some evidence that a new type of ordered state ("hidden" order parameter) may exist in this region. Such a state may be electron–hole pairing with an imaginary order parameter for which, as we know [35], closed spontaneous currents appear [36].

REFERENCES

1. G. V. M. Williams and J. L. Tallon, Phys. Rev. B **57**, 10984 (1998).
2. Y. D. Leu, W. N. Huang, C. M. Wang, and H.-C. I. Kao, Physica C (Amsterdam) **261**, 284 (1996).
3. I. Bosovic and J. N. Eckstein, in *Physical Properties of High Temperature Superconductors*, Ed. by D. M. Ginsberg (World Scientific, Singapore, 1996), Vol. V.
4. J. Kane, K.-W. Ng, and D. Moecher, Physica C (Amsterdam) **294**, 176 (1998).
5. S. Kashiwaya, T. Ito, K. Oka, *et al.*, Phys. Rev. B **57**, 8680 (1998).
6. J. C. Campuzano, G. Jennings, M. Faiz, *et al.*, Phys. Rev. Lett. **64**, 2308 (1990).
7. C. G. Olson, R. Liu, D. W. Lynch, *et al.*, Phys. Rev. B **42**, 381 (1990).
8. D. M. King, Z.-X. Shen, D. S. Dessau, *et al.*, Phys. Rev. Lett. **70**, 3159 (1993).
9. R. O. Anderson, R. Claessen, J. W. Allen, *et al.*, Phys. Rev. Lett. **70**, 3163 (1993).
10. H. Ding, A. F. Bellman, J. C. Campuzano, *et al.*, Phys. Rev. Lett. **76**, 1533 (1996).
11. Z.-X. Shen, W. E. Spicer, D. M. King, *et al.*, Science **267**, 343 (1995).
12. R. A. Klemm, Int. J. Mod. Phys. B **12**, 2920 (1998).
13. Y. Koike, T. Takabayashi, T. Noji, *et al.*, Phys. Rev. B **54**, R776 (1996).
14. S. M. Anlage, D.-H. Wu, J. Mao, *et al.*, Phys. Rev. B **50**, 523 (1994).
15. F. Hayashi, E. Ueda, M. Sato, *et al.*, J. Phys. Soc. Jpn. **67**, 3234 (1998).
16. C. Kendziora, R. J. Kelley, and M. Onellion, Phys. Rev. Lett. **77**, 727 (1996).
17. G. V. M. Williams, J. L. Tallon, E. M. Haines, *et al.*, Phys. Rev. Lett. **78**, 721 (1997).
18. M. R. Norman, H. Ding, M. Randeria, *et al.*, Nature **392**, 157 (1998).
19. B. Batlogg and V. Emery, Nature **382**, 20 (1996).
20. C. Renner, B. Revaz, K. Kadowaki, and I. Maggio-Aprile, Phys. Rev. Lett. **80**, 3606 (1998).
21. H. Ding, M. R. Norman, T. Yokoya, *et al.*, Phys. Rev. Lett. **78**, 2628 (1997).
22. S. Massidda, N. Hamada, J. Yu, and A. J. Freeman, Physica C (Amsterdam) **157**, 571 (1989).
23. A. J. Freeman and J. Yu, Helv. Phys. Acta **61**, 401 (1988).
24. E. F. Gross, V. I. Perel', and R. I. Shekhmamet'ev, Pis'ma Zh. Éksp. Teor. Fiz. **13**, 320 (1971) [JETP Lett. **13**, 229 (1971)].
25. K. A. Brueckner, Phys. Rev. **100**, 36 (1955).
26. H. A. Bethe, Phys. Rev. **103**, 1353 (1956).
27. H. J. Lipkin, *Quantum Mechanics* (North-Holland, Amsterdam, 1973; Mir, Moscow, 1977).
28. M. Schafroth, S. Butler, and J. Blatt, Helv. Phys. Acta **30**, 93 (1957).
29. A. I. Baz', Ya. B. Zel'dovich, and A. M. Perelomov, *Scattering, Reactions and Decays in Nonrelativistic Quantum Mechanics* (Nauka, Moscow, 1971, 2nd ed.; Israel Program for Scientific Translations, Jerusalem, 1966).
30. B. Velicky and I. Sak, Phys. Status Solidi **16**, 147 (1966).
31. L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-Relativistic Theory* (Nauka, Moscow, 1989, 4th ed.; Pergamon, Oxford, 1977, 3rd ed.).
32. E. M. Lifshitz and L. P. Pitaevskiĭ, in *Statistical Physics* (Nauka, Moscow, 1978; Pergamon, New York, 1980), Part 2.
33. A. I. Larkin and Yu. N. Ovchinnikov, Zh. Éksp. Teor. Fiz. **47**, 1136 (1964) [Sov. Phys. JETP **20**, 762 (1964)].
34. R. Fulde and R. Ferrel, Phys. Rev. **135**, 550 (1964).
35. B. A. Volkov, A. A. Gorbatsevich, Yu. V. Kopaev, and V. V. Tugushev, Zh. Éksp. Teor. Fiz. **81**, 729 (1981) [Sov. Phys. JETP **54**, 391 (1981)].
36. C. M. Varma, Phys. Rev. B **61**, R3804 (2000).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 91, No. 4, 2000, pp. 832–843.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 118, No. 4, 2000, pp. 959–972.

MISCELLANEOUS

Interaction of Nucleic Acid Segments as a Result of Modification of the Network of Hydrogen Bonds of the Solvent

V. L. Golo[a, *], Yu. M. Yevdokimov[b], E. I. Kats[c, d], and V. I. Salyanov[b]

[a]*Moscow State University, Moscow, 119899 Russia*
**e-mail: golo@mech.math.msu.su*
[b]*Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, 117984 Russia*
[c]*Landau Institute for Theoretical Physics, Russian Academy of Sciences, Moscow, 117940 Russia*
[d]*Laue-Langevin Institute, 38042, Grenoble, France*

Received May 26, 2000

Abstract—It is shown that experimental results on the influence of various factors in the formation efficiency and structure of cholesteric liquid-crystal dispersions of nucleic acids cannot be consistently described using conventional theories of liquid crystal formation. A new model is proposed for the interaction of nucleic acid segments which allows for a change in the particular structure of the solvent hydrogen bonds in the presence of nucleic acid molecules. The conclusions of the model agree with existing spectroscopic and structural investigations of DNA dispersions. According to our model, interaction between nucleic acid molecules and solvent modifies proton tunneling processes in the latter, leading to effective interaction between the nucleic acids. A theoretical analysis of the model is made using a pseudospin formalism in which the effective interaction potential of the nucleic acid segments is calculated. It is shown that this potential may lead to nematic ordering for small distances between the nucleic acid molecules ($R \le 30$ Å) and cholesteric ordering for large distances. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

A well-developed method exists for preparing liquid-crystal dispersions of nucleic acids [1–3] (relevant experimental data are also presented in our previous study [4]). Usually, rigid segments of double strand DNA or RNA molecules having lengths of the order of one or two persistent lengths (i.e., up to 1000 Å) are used to prepare dispersions. Aggregation of these segments in an aqueous solution is induced by adding various chemical agents which create osmotic pressure (for example, ethylene or polyethylene glycol) and positive counterions which partially compensate for the negative charge of the phosphate groups of the nucleic acids (NA).

As a result of the action of these agents at a dispersion density corresponding to the average spacing between the NA molecules $R \approx 50$ Å a macroscopic cholesteric structure forms having a helix pitch of the order of a few micron (which corresponds to a molecular angle of relative rotation of the long axes of the molecules of the order of a degree). The cholesteric structure remains stable up to distances R around 30 Å (in the range 50–30 Å the pitch of the macroscopic helix varies negligibly, increasing from 2.4 to 2.5 μm) and at shorter distances ($R \le 30$ Å) the molecules exhibit parallel ordering.

This factor (the small angle φ of relative rotation of the molecules in neighboring quasi-nematic cholesteric layers) cannot be explained in terms of purely steric forces of interaction between macromolecular segments. In fact, an elementary geometric analysis shows that as a result of the mutual impermeability of the rigid macromolecular segments (steric constraints) these should be packed in layers in each of which the axes of the NA molecular helixes should be parallel. With this packing of the rigid helixes, the neighboring layers must rotate relative to each other through the angle

$$\varphi = \arctan(d/h), \qquad (1)$$

where d is the diameter of a macromolecular (20 Å for DNA) and h is the pitch of the double helix (36 Å for DNA). Even this primitive estimate (which for example, neglects the fact that the thickness of the quasi-nematic layer should be greater than d) gives for the angle $\varphi \approx 10°$ whereas a more accurate analysis (see, for example [5] and also the recent review [6]) gives an even larger difference from the experimental data. By allowing for dispersion or electrostatic forces of attraction in addition to steric repulsion (see [7–12]) we can obtain the angle $\varphi \approx 1°$ which agrees in principle with the experimental data for liquid-crystal NA dispersions although all the system parameters must be specially selected for this purpose.

In Section 3 we give a more comprehensive analysis of existing theories of liquid-crystal ordering and show that these do not give a unified consistent description of the complete set of experimental data for NA dispersions. The next section is devoted to a description of the experimental data. In our view, the main disadvantage

of the existing theories (or more accurately, their applications to aqueous NA dispersions) is that they neglect the specific properties of the solvent, water, in which the liquid-crystal NA dispersions are formed. The literature devoted to the study of the influence of NA on the water surrounding them is quite extensive (this research is briefly summarized in Section 4).

Interaction between NA molecules and solvent (water) leads to substantial modification of the structure of hydrogen bonds in the nearest aqueous environment of the NA segments. Then, as a result of modification of the structure of the hydrogen bonds effective interaction takes place between the NA segments. The radius of this interaction does not exceed the characteristic depth of "penetration" of the hydrogen bond modification $\xi \le 50$ Å although these scales determine all the characteristic features of the behavior of the liquid-crystal NA dispersions. In Section 5 we describe our model of the formation of liquid-crystal dispersions in aqueous NA solutions. According to our model, interaction between NA and solvent changes the structure of the hydrogen bonds and the probability of proton tunneling along them, leading to interaction between the NA molecules. In a certain range of system parameters and external conditions these interactions of an entropic nature may compete with van der Waals, steric, and electrostatic forces. For a theoretical description of the model we use a pseudo-spin formalism in which the pair interaction potential is calculated. Finally, the concluding section of the study is devoted to a discussion of the main consequences of our model and the possibilities for refining and checking it.

2. PROPERTIES OF LIQUID-CRYSTAL DISPERSIONS OF NUCLEIC ACIDS

A considerable amount of experimental data has been accumulated so far on the conditions of formation, types, and properties of liquid-crystal phases and dispersions of nucleic acids [13, 14]. It is well established that linear, double strand, rigid DNA or RNA molecules of low molecular mass ($\le 10^6$) form liquid-crystal phases in concentrated aqueous–polymer or aqueous–salt solutions. On transition from the isotropic to the liquid-crystal phases the parameters of the secondary NA structure change negligibly. In order to obtain liquid-crystal dispersions two conditions need to be satisfied: the negative charges of the phosphate groups must be compensated (using cations of metal salts dissolved in water) and a sufficient density of macromolecular segments must be created (using a neutral polymer with respect to the nucleic acids, such as polyethylene glycol).

According to theoretical estimates and experimental data [13, 14] positively charged ions and polycations can screen up to 80% of the positively charge phosphate groups of macromolecules, providing the degree of neutralization needed to compensate for the NA segments in an aqueous salt solution. Importantly all the parameters of the linear secondary structure are conserved even in a complex (NA macromolecule + cation). As the polyethylene glycol concentration (or the molecular mass) varies, the osmotic pressure created by the polyethylene glycol and consequently the density of the macromolecular segments also varies.

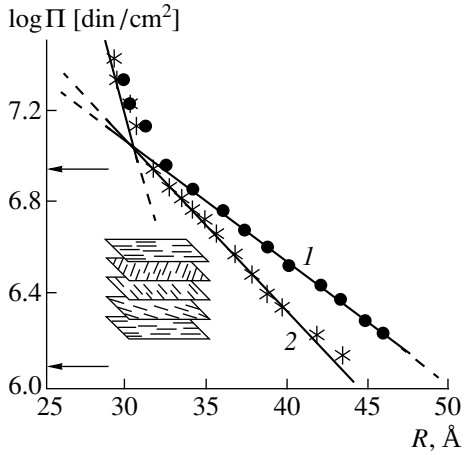

Fig. 1. Dependence of the osmotic pressure on the average distance between NA molecules for DNA (*1*) and RNA (*2*) molecules.

Instead of the density, we can use the average distance between the centers of mass of the macromolecules R. A liquid-crystal structure forms for $R \le 50$ Å (for larger R an isotropic solution exists). This type of structure is identified in particular from the presence of a high-intensity band in the circular dichroism spectrum in the absorption region of nitrous bases ($\lambda_{max} \approx$ 2600 Å). From the circular dichroism and various other optical data [3, 13, 14] we can conclude that in the range 30 Å $< R <$ 50 Å the NA dispersions have a cholesteric structure with an almost constant helix pitch p = 2.4–2.5 μm whereas for large R a transition takes place to the isotropic phase. Finally, fairly dense dispersions ($R <$ 20 Å) form a hexagonal close-packed crystal phase (in this case, the circular dichroism band vanishes which indicates that the long axes of the molecules exhibit parallel, i.e., nematic, orientation).

Figure 1 gives the logarithm of the osmotic pressure Π created by polyethylene glycol as a function of the average distance R between the nucleic acids for DNA and RNA. The region of existence of the cholesteric (denoted by the arrows in the figure) correlates with the kink on the dependence of logΠ on R.

It is conventionally assumed (for further details see the following section) that direct interactions between macromolecules, which are usually considered to be steric, dispersion, and electrostatic forces, are responsible for the formation of liquid-crystal dispersions. However, a detailed analysis of the experimental data on NA dispersions shows that they are poorly described in the conventional scheme which thus requires new components at the least. Below we shall briefly men-

tion such experimental facts, referring to our special study on this topic for details [4].

2.1. Cholesteric Phases of DNA and RNA [15]

In the range 30 Å < R < 50 Å rigid segments of the double strand, right-handed-helix form B-DNA form a liquid-crystal dispersion. This dispersion is characterized by anomalous optical activity which is manifest in the circular dichroism spectrum as a high-intensity negative band. The negative sign of the anomalous band indicates that right-handed helix B-DNA molecules form liquid crystals with a left twist. However, the relationship between the sign of the circular dichroism and the direction of the cholesteric helix is not completely unique. This is because the dichroism of these systems is mainly determined by interaction between the light wave and the π-electrons of base pairs. Thus, the sign of the circular dichroism reflects the order of the dipoles of the corresponding electron transitions rather than the order of the rigid molecular cores. This observation on the relationship between the sign of the circular dichroism and the direction of the cholesteric helix implies strong coupling between the electron dipole moments and the orientations of planes of base pairs. This statement also illustrates a common property of all chiral systems. There are infinitely many parameters characterizing the chiral breaking of symmetry (in the same way that an infinite set of spherical harmonics defines the breaking of spherical symmetry). Various macroscopic properties of the system can generally depend on various chiral order parameters. In particular, the sign of the chirality cannot be determined uniquely. For example, for the case of purely steric forces a formula of the type (1) which gives right-handed cholesteric packing of right-handed helix molecules is only valid under the condition $h \ll l$ (l is the length of the macromolecular segment). In the case $h \approx l$ steric forces produce a left-handed cholesteric helix.[1]

If a dispersion of rigid segments (having the same ratio l/d, where l is the length of the rigid macromolecular segment and d is the diameter) of the right-handed, double strand A-form of RNA forms under the same conditions (polyethylene glycol concentration, temperature, type of cations, and so on), a right-handed cholesteric liquid-crystal dispersion is obtained. These facts are difficult to explain merely in terms of modification of the direct interaction between rigid macromolecular segments. At the molecular masses used to form liquid-crystal dispersions the entire difference between the B form of DNA and the right-handed double strand structure of A-RNA can be reduced to the replacement of a single base, i.e., the replacement of methyluracil (thymine) by uracil. Under these conditions we can only expect a small modification of the direct interparticle interaction rather than the fundamental rearrangement of the entire structure observed experimentally [15].

[1] This purely geometric fact is a particular case of the general property of chiral objects noted in the review [6] whereby a chiral object can be continuously converted into its mirror image without being converted into an achiral object along the path (in chiral order parameter space).

2.2. Damage to the Cholesteric Structure by Heating and Modification of DNA Molecules [16, 17]

The process of formation of liquid-crystal dispersions of DNA as a result of phase exclusion in polymer-containing solutions is reversible and reproducible. However, both the sign of the cholesteric helix and the properties of the cholesteric are very sensitive even to a small change in the external conditions or modification of the DNA molecules. For example, it was established in [16] that the circular dichroism band characteristic of a cholesteric dispersion formed by DNA segments 150 base pairs long disappears when the system is heated to temperatures at which [18] no substantial changes take place in the secondary structure of the DNA B-form (or other double strand polynucleotides), the optical and electronic spectra, and other properties of individual molecules. It is therefore natural to assume that the direct intermolecular interaction cannot change substantially. Nevertheless, the structure of the dispersion changes fundamentally [16].

Similarly, it was shown in [17] that in cases of quite negligible (from the viewpoint of molecular properties) modification of the DNA molecule (a single change per 10^2–10^3 base pairs) the cholesteric helix is destroyed and nematic ordering occurs.

2.3. Immiscibility of Right-Handed Cholesteric RNA Dispersions and Left-Handed DNA Dispersions [4, 13]

For classical low-molecular liquid crystals the miscibility in any proportions of mesophases with their textures conserved is conventionally used (see, e.g., [9]) in the classification of liquid crystal types to assign them to a single group. If this method is applied to a mixture of a left-handed cholesteric DNA dispersion and a right-handed cholesteric RNA dispersion, the formation of a mixed liquid crystal would be predicted.

For a mixed liquid crystal any property which depends on the order parameter, such as the circular dichroism, is determined by minimizing the free energy of the mixture. We denote the circular dichroism of the mixture by A and that of the components by A_1 and A_2, respectively. As a result of a difference between A and A_1 and A_2, the mixed liquid crystal has excess free energy which can be expressed as follows in the principal (harmonic) approximation in the sense of the Landau theory:

$$\Delta F = \frac{1}{2}\alpha_1(1-x)(A-A_1)^2 + \frac{1}{2}\alpha_2 x(A-A_2)^2, \quad (2)$$

where x is the mixture concentration, and α_1 and α_2 are the phenomenological coefficients of the Landau

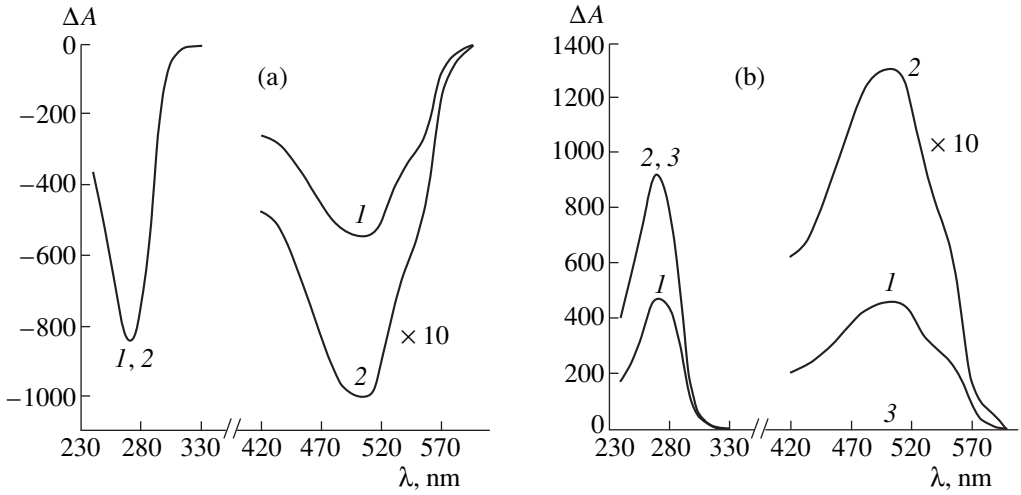

Fig. 2. Circular dichroism spectra of NA molecules and dye: (a) dye added to cholesteric phase of dispersion; (b) NA molecules treated with dye and then the cholesteric phase is obtained. Curve *1*, $c_d = 6.8 \times 10^{-6}$ M, curve *2*, $c_d = 27.2 \times 10^{-6}$ M, and curve *3* was obtained after washing out the dye ($c_d = 0$).

expansion of the free energy which generally differ for both mixture components.

Minimizing ΔF with respect to A gives the dependence of the circular dichroism of a mixed liquid crystal on the concentration[2]

$$A(x) = \frac{\alpha_2 A_2 x + \alpha_1 A_1 (1-x)}{(1-x)\alpha_1 + x\alpha_2}. \qquad (3)$$

In order to check equation (2) for a mixed DNA + RNA dispersion the authors of [4] determined the compensation point $A(x) = 0$ (we note that the signs of A_2 and A_1 differ) and then (at this point in terms of concentration) formed polymer chelate crosslinks (bridges) between the molecules. The corresponding bridges have a molecular mass considerably lower than the molecular mass of the DNA or RNA segments. In addition, the concentration of these bridges is low. Thus, for a mixed liquid crystal it should be predicted that the compensation point does not change as a result of this chemical cross-linking, i.e., $A(x) = 0$. However, the experiment [4] contradicts this, showing that as a result of cross-linking we have $A(x) \neq 0$. The only possible consistent description of this result (see [4]) dictates the conclusion that left- and right-handed cholesteric dispersions of DNA and RNA do not mix, i.e. the DNA and RNA molecules form independent liquid crystals.

2.4. Circular Dichroism Band of Dyes in Nucleic Acid Dispersions [4, 19]

Extremely important information on the structure of cholesteric dispersions can also be obtained by studying the circular dichroism of dye molecules incorporated into the nuclei acid molecules by some method. Such investigations were reported in [4, 19] using two types of dye: daunomycin (D) and mitoxanthrone (M).

If the dye molecules are added directly to a cholesteric dispersion formed from double strand DNA molecules, i.e. a complex (DNA–dye) is formed, two bands appear in the circular dichroism spectrum of the cholesteric dispersion. One of these lies in the DNA absorption band, the other lies in the absorption band of the dye. Both bands have the same negative signs which corresponds to the incorporation (intercalation) of dye molecules between nucleic acid base pairs.

If these dyes are added to linear double strand DNA molecules and then a cholesteric phase is formed from the DNA–dye complexes as a result of phase exclusion, the results are different. When molecules of dye D are added, there is a critical concentration c_d^* at which the signs of the bands in the circular dichroism spectrum lying in the absorption regions of the DNA and the dye undergo an abrupt inversion, i.e., two positive bands appear in the circular dichroism spectra, not negative bands as in the first case. However, if the dye is "washed out" of the DNA–dye complex in the cholesteric dispersion thus formed [by adding a surface-active agent such as sodium dodecyl sulfate (SDS)], the band corresponding to the dye disappears in the circular dichroism spectrum although the band in the DNA absorption region remains unchanged, i.e., keeps its positive sign. This result shows that both "classical" cholesteric dispersions having a negative band in the circular dichroism spectrum and dispersions having a positive band can be formed from the same DNA molecules. Figure 2 shows circular dichroism spectra of NA dispersions treated with a dye: adding the dye to the cholesteric phase (Fig. 2a, two negative circular dichroism bands are observed) and adding the dye to

[2] We note that the trivial linear dependence $A(x) = A_1(1-x) + A_2 x$ is only obtained from (2) in the particular case $\alpha_1 = \alpha_2$.

the isotropic phase (Fig. 2b). Curve *3* was obtained after washing out the dye. The circular dichroism band is then absent and the intrinsic NA band remains positive.

It should also be stated that this effect is only observed for dye D and its analogs, not for dye M. Dye M induces no inversion of the band in the circular dichroism spectrum of cholesteric DNA dispersions under any conditions and, other conditions being equal, the absorption band of this dye only retains the negative band. However, if dye M is added to a cholesteric DNA dispersion "inverted" by the addition of dye D, the appearance of a positive band may also be observed in the absorption band of dye M. This result shows once again that two different types of cholesteric DNA dispersions may be obtained depending on the method of formation. These results indicate that the behavior exhibits unique hysteresis which is consistent with the possibility of a first-order cholesteric–nematic conformational phase transition in DNA dispersions discussed in Section 5.

Summarizing the experimental data presented in this section we can state that the properties of liquid-crystal NA dispersions differ substantially from the properties of classical low-molecular liquid crystals. This difference suggests that the formation of structure of NA dispersions depend on some additional component of the system which is not significant (or plays a passive, auxiliary role) in low-molecular liquid crystals.

3. CONVENTIONAL DESCRIPTION OF CHOLESTERIC ORDERING

Any systematic theory of condensed structures formed by rigid segments of NA molecules (as in the theory of any liquid or liquid-crystal phases) can only be constructed numerically using direct *ab initio* quantum-mechanical calculations of atom–atom potentials. Quite clearly, these calculations are laborious and, even with the latest supercomputers, can only be made for a relatively small (for macroscopic condensed systems) number of particles (several thousand). In addition, it is difficult to use the results of calculations made for specific values of the parameters to analyze the behavior of the system when these parameters vary and in any case, these calculations are, so to speak, too detailed if we are interested in qualitative behavioral characteristics (which are important for numerous physical and biophysical applications of condensed NA structures).

However, there is an approach which has proved successful in the theory of low-molecular liquid crystals in which the liquid crystal structure is determined in the mean field approximation by competition between dispersive (van der Waals) attraction and steric repulsion. These contributions dominate in cholesteric liquid crystals, although in this case in order to determine the pitch of the cholesteric helix we need to allow for the first nonvanishing contribution to the expansion of the pair interaction energy in terms of multipole moment in chiral systems, i.e., dipole–quadrupole interaction [20].

These types of interactions clearly play a dominant role in the structures of condensed phases formed by rigid segments of NA molecules. However, in the case of DNA molecules the existence of nonuniformly distributed charges of phosphate groups and condensed counterions leads to additional anisotropic interactions of an electrostatic nature (see, for example [7]). In [11] the authors also considered the contribution of dipole–dipole forces which in cases of helix ordering of the dipole moments can also lead to cholesteric ordering of the molecules.

In any case, the complete pair interaction of two rigid segments of NA molecules is usually described in a fairly local approximation as follows:

$$V_{12}(r_{12}, \varphi_{12}) = J_{12}(r_{12})\cos(2\varphi_{12}) + I_{12}(r_{12})\sin(2\varphi_{12}), \tag{4}$$

where r_{12} is the distance between the centers of mass of molecules 1 and 2, φ_{12} is the angle between the long axes of the molecules (as is usually the case in mean field theory for cholesterics [9], we assume that molecules 1 and 2 lie in parallel quasi-nematic planes).

The functions J_{12} and I_{12} depend on the main types of forces acting between molecules 1 and 2. For example, when van der Waals interactions predominate we have $J_{12} \propto 1/r_{12}^6$; $I_{12} \propto 1/r_{12}^7$, and for electrostatic interactions $J_{12} \propto 1/r_{12}^4$; $I_{12} \propto 1/r_{12}^5$. In the more general case, interaction between multipoles of the same polarity always leads to an even function of the angle φ_{12} whereas multipoles of different parity (i.e., chiral interaction) lead to an odd function of φ_{12}. Thus, the formal multipole expansion of the pair interaction has the form

$$V_{12} = V_0(r_{12}) + \sum_{m=1}^{\infty} [J_{2m}(r_{12})\cos(2m\varphi_{12}) + I_{2m}(r_{12})\sin(2m\varphi_{12})]. \tag{5}$$

For a macroscopic description of the structure we need to find the molecular distribution function (when only the orientational order of the condensed phases is being analyzed this is the orientational distribution function $f(\varphi)$). In the mean-field approximation this distribution function can be obtained by minimizing the free energy, which has the standard form

$$F = F_0 + \rho T \int f(\varphi)\ln f(\varphi)\, d\varphi + \frac{1}{2}\rho^2 \int f(\varphi_1) f(\varphi_2) V_{12}\, dr_1 dr_2 d\varphi_1 d\varphi_2, \tag{6}$$

where T is the temperature, ρ is the density, F_0 is the component of the free energy of the isotropic phase which does not depend on the orientational order, the

second term in (6) allows for a reduction in the entropy of the system as a result of the orientational order; the last term is associated with direct pair interaction between the macromolecular segments, and this pair interaction V_{12} is defined by expressions (1) or (2).

This procedure for $I_{12} \neq 0$ (or $I_{2m} \neq 0$) leads to cholesteric twisting of the structure for any system densities [i.e., any (r_{12})] and was discussed in detail in the literature (see, for example, [7, 9, 11, 20]). Some very serious practical and fundamental theoretical claims can be put forward for this procedure. For low-molecular systems using a certain number of unknown (phenomenological) parameters the mean-field theory put forward above can fairly accurately describe the experimental data, for example, the dependence of the helix pitch on the temperature or concentration of chiral impurities. However, the number of fitting parameters of the theory is by no means small (four or five) and, worse still, for various types of experimental data it is sometimes necessary to use a completely different set of fitting parameters. For condensed NA phases even this fitting of the experimental data (see previous section) is not feasible with a natural choice of fitting parameters. The theoretical basis of the multipole expansion (1) or (2) also has serious shortcomings since it assume that the intermolecular distances r_{12} are much larger than the characteristic size of the molecules $(d^2l)^{1/3}$ (d is the diameter and l the length) which is not satisfied in most real condensed phases.

We also note that in the approach described above the solvent participates fairly passively and specifically as a carrier of electrostatic or dispersive interactions although even in this simplified model, anomalous compensation of the chirality by the solvent may occur subject to the condition [12]

$$\varepsilon_m = \varepsilon_l \varepsilon_t, \tag{7}$$

where ε_m is the permittivity of the solvent, and ε_l and ε_t are the longitudinal and transverse permittivities of the interacting chiral and anisotropic molecules.

The condition (7) clearly cannot be satisfied at arbitrary frequencies and thus the sign of the chirality can only change (and go to zero) if equation (7) is satisfied at frequencies which make the main contribution to the dispersion forces (for example, near the fundamental absorption band of the NA and solvent molecules in the infrared) and if, in addition, all three permittivities depend weakly on frequency in this region. If this is not the case, in this theory the solvent plays no significant role in the intermolecular reactions responsible for the formation of liquid-crystal order in the NA condensed phase.

Interaction between NA molecules associated with fluctuations of their profile is also discussed in some studies (see, for example [21]). However, this interaction can only be significant for fairly long NA molecules when l is considerably greater than the persistent length.

4. INFLUENCE OF NUCLEIC ACID MOLECULES ON PROPERTIES OF SURROUNDING WATER

The literature contains numerous data which indicate that on small scales (≤50 Å) the solvent around NA molecules (usually an aqueous salt) is in many respects a structured system [22–24]. We know [25] for example, that when ice melts no more than 15% of the hydrogen bonds are broken. This structure occurs as a result of the presence of a network of hydrogen bonds in the water whose molecules interact with base pairs of NA segments and phosphate groups. Below we present experimental data which suggest that the packing of NA molecules during the formation of liquid-crystal dispersions is related to the properties of the water molecules situated between NA molecules which converge during the phase separation process.

The role of water in the stabilization and "reflection" of the structure and conformation of nucleic acids has been the subject of study and discussion for some time [22, 26]. We know that water molecules forming hydrogen bonds with NA molecules form a hydrate shell around the latter, which has a complex structure. It was shown in [27, 28] that the hydration schemes of RNA and DNA molecules differ substantially. From this it follows that when the interaction of nucleic acids with the solvent is taken into account, the difference between the liquid-crystal dispersions of DNA and RNA noted in Section 2 does not appear absolutely inexplicable. The difference between the structure of the hydrate shells of the A form of RNA and the B form of DNA is particularly noticeable near the narrow groove of the NA molecule where the 2'-hydroxyl group is positioned in A-RNA [27, 28].

Quantum-chemical calculations [29] show that the first hydration shell of water around NA molecules consists of 10–11 water molecules per NA monomer. The next 8–9 water molecules have different infrared spectroscopic characteristics from water molecules in the first hydration shell and from normal bulk liquid water. Direct quantum-chemical calculations [30] qualitatively confirm the spectroscopic data. It follows from these calculations that around 20 water molecules form a layer filling the so-called main groove of the double strand DNA molecule. The water molecules in this layer as it were screen the polar groups of base pairs entering the main groove in the ratio of approximately one "localized" water molecule per base pair. The water molecules in the next hydration shells form hydrogen bridges with the first shell although the orientation of these hydrogen bonds is more or less random.

Consequently, the relative influence of the solvation water and the NA molecules extends to scales which depend very much on the property being considered. If the orientation order of the hydrogen bonds does not penetrate more than a few angstroms into the bulk, the correlation length characterizing the NA-induced change in the frequency of proton tunneling along the

system of hydrogen bonds may exceed 10^2 Å. Note that a similar trend has been thoroughly investigated in numerous studies of the spectra of the so-called disjoining pressure in thin water films (see, e.g., the review [31] and the literature cited there). Numerical modeling of the behavior of water in porous matrices [32] also indicates that the properties of water are modified substantially in small volumes. For example, under certain conditions (degree of hydration, temperature, and pore size) the dynamic structure factor of water in small pores is similar to that for supercooled liquids near the glass phase transition [32], for example, a so-called boson peak is observed. Since the characteristic pore sizes discussed in [32] are of the order of the typical intermolecular distances for NA dispersions similar "glass" behavior of water may be envisaged in dispersions.

An important characteristic of our case (liquid-crystal NA dispersions) is that unlike the generally uniform and smooth surfaces used to study the properties of water in thin planar capillaries, the surface of a rigid NA segment of the order of 150 base pairs long is fundamentally heterogeneous because of its very structure. The heterogeneity of the macromolecules leads to nonuniform binding of water molecules with base pairs (and thus leads to a nonuniform change in the proton tunneling energy, which is in fact a source of anisotropic interaction forces which strongly influence the structure of the liquid-crystal dispersion). Quantum-chemical calculations and spectroscopic experimental data can be used to estimate the binding energy of a water molecule with various NA groups. For example, it is stated in [33] that a water molecule may be bound with two neighboring base pairs simultaneously (e.g., with an N_7 atom and NH_2 group of adenine, with an O_2 atom and NH group of thymine, and with N_7 and O_6 or N_3 atoms and an NH_2 group of guanine). The highest binding energy (−12.6 kcal/mol) is obtained for N_7 and O_6 atoms of guanine, and the weakest binding is between water and thymine (–9.5 kcal/mol). These energies nevertheless exceed the characteristic scales of intermolecular interaction in bulk water (6.7 kcal/mol).

Thus, to summarize the experimental observations listed in this section we can conclude that on scales between 10 Å and 50 Å the properties of water around NA molecules differ substantially from the bulk properties. On the smallest scale (up to 10 Å) a relatively stable (static) network of hydrogen bonds appears in the water. On larger scales (up to 50 Å) the network of hydrogen bonds is more mobile although the nature of the proton tunneling still differs from the bulk process. The properties of the NA hydrate shells are influenced by various factors (which from the point of view of the secondary structure of these nucleic acids are relatively weak) listed in Section 2.

A specific feature of the dispersion of nucleic acids is the presence of an aqueous solvent. It is well known (see, e.g., [34]) that for NA molecules at short distances so-called hydration forces play an important role. These forces decay exponentially with distance. The corresponding characteristic length λ_h for DNA under natural conditions is of the order of a few angstroms.[3] However, the detailed nature of these forces is not very clear. In any case, it is assumed that pure entropy effects will make a substantial contribution to the hydration forces. Water molecules ordered in some way around the surface of NA molecules have a lower entropy than the water surrounding (at large distances from the surface) interacting molecules. This separation of water molecules into internal and external relative to the interacting surfaces is only unique for the geometry of a planar capillary or film. For a liquid-crystal NA dispersion the separation of water molecules into external and internal has a quite different meaning. We also note that water molecules have fairly large dipole moments. Hence the short-range hydration interactions under conditions where the structure of the hydrogen bonds is frozen-in to a certain extent can be modified by long-range dipole–dipole forces.

It is important to note that the total entropy of a system consisting of NA molecules, water, and salt ions increases as the average distance between the NA molecules decreases (the density increases) [21, 34]. At the same time, the degree of order of rigid segments of NA molecules can only increase under compression. Similarly the degree of adsorption of counterions on NA molecules increases as the system is compressed. Consequently both macromolecules and counterions make a negative contribution to the total entropy and thus the only component of the system that can increase the entropy of the system is water.

In order to describe this contribution theoretically we recall that in accordance with the experimental data noted in this section, the first few shells of water surrounding the NA molecules differ substantially from the more distant shells. On the basis of these data we can assume that the nearer (internal) water first possesses a certain order (or more accurately frozen-in property) in the orientations of the hydrogen bonds which lowers its entropy and second proton tunneling along these hydrogen bonds in the nearer water is possibly made easier and in this case, the proton distribution becomes more uniform, thus increasing the entropy. The total effect can, in principle, have any sign.

The estimates made in [34] give an excess entropy contribution to the free energy of the order of 0.3 kcal/mol under normal conditions and at room temperature,

[3] In fact there are two different ranges of parameters. At short distances (between 3 and 7 Å) the repulsion forces increase exponentially with decreasing water concentration (dehydration) or, which is equivalent, the average intermolecular distance with the characteristic length $\lambda_h \approx 3$ Å. At larger distances between interacting molecules $\lambda_h \approx 6$ Å such an exponential dependence on distance occurs, for example, as a result of screening of Coulomb and van der Waals interactions by counterions.

which is substantially lower than the characteristic energy of the hydrogen bond in bulk water although summation over all molecules in the nearest neighborhood to the NA molecules may give an appreciable effect. In the following section we describe a simple model which allows for this change in the probability of proton tunneling and we propose a formalism which can be used to calculate the entropy interaction of rigid NA segments which occurs as a result of modification of the network of hydrogen bonds and as a result of proton tunneling in this structure.

5. MODEL OF PSEUDOSPIN INTERACTION OF NUCLEIC ACID MOLECULES

Using a method first developed by Blinc [35], each hydrogen bond can be set in correspondence with a two-level system, i.e., the pseudospin $\hat{S}$ such that

$$S^z|\mu\rangle = \pm\frac{\varepsilon}{2}|\mu\rangle, \tag{8}$$

where the z-axis is directed along the bond, and $\pm\varepsilon/2$ is the proton energy in two of its possible states relative to the oxygen ions of water and the vector of state $|\mu\rangle$ determines the wave function of the two-level system. If in the immediate vicinity of the NA segments we have a specific structure of hydrogen bonds of water, this implies that we have a certain graph (network) having the pseudospin operators S_i^α defined at the vertices, where i is the number of the vertex and α is the projection of the pseudospin on the local reference axis in the vertex i (i.e., $\alpha = x, y, z$). Consequently in this case the solvent is not a passive carrier of interaction between molecules but has an intrinsic structure and internal degrees of freedom. In order to describe this structure in space we can introduce a local Frenet reference point at each point, i.e., three orthonormalized vectors $\mathbf{v}_j$ (or the matrix $X_{ij} = (v_i)_j$). The topology of this network, which resembles the structure of the hydrogen bonds of ice, is consistent with the Bernal–Fowler rules [25], and specifically: (1) there are two protons in the vicinity of each oxygen atom; (2) each proton may be associated with a single hydrogen bond.

Segments of nucleic acids positioned in this network X_{ij} cause unavoidable deformation which then influences the surrounding molecules leading to their effective anisotropic interaction.

Thus, taking into account the experimental data presented in the previous section we can confirm that in water shells "nearer" to the NA molecules there is a relatively stable network of hydrogen bonds, in which the orientational correlations decrease exponentially ($\exp(-r/\lambda_0)$) with distance, and pseudospin ordering which also decreases exponentially with distance $\propto\exp(-r/\lambda_s)$ where the corresponding pseudospin length λ_s may exceed λ_0 several times.[4]

The assumptions put forward above on the network of hydrogen bonds surrounding the NA molecules can be reformulated using a simple phenomenological model. We shall first consider the region of scales of the order λ_0 where orientational ordering of the hydrogen bonds occurs. A DNA or RNA molecule is chiral so that the orientation of the network of hydrogen bonds induced by this molecule may have a specific degree of chirality, i.e., twist. However, twisting of the network of hydrogen bonds leads to deformation energy. Thus, at distances of the order of λ_0, competition occurs between the NA-induced chirality and the deformation of the bond network. The induced chirality gives the following contribution to the system energy:

$$E_{\rm ch} = \frac{1}{2}\alpha_{\rm ch}(\omega_{ij} - \omega_{ij}^0)^2, \tag{9}$$

where ω_{ij} is the skew-symmetric twist matrix of the local reference frame $\mathbf{v}_i$ of the network of hydrogen bonds (see, e.g., [36]), ω_{ij}^0 is the equilibrium twist matched with the chiral NA molecules, and α_{ch} is a phenomenological coefficient which determines the chirality of the network of hydrogen bonds induced by the hydration forces (in accordance with the experimental data discussed above this coefficient depends on the average intermolecular distance R in the dispersion according to the law $\exp(-R/\lambda_h)$).

At the same time, the nonuniform distribution of the local reference frame orientations in the network of hydrogen bonds may lead to elastic deformation of the latter:

$$E_{\rm el} = \frac{1}{2}\alpha_{\rm el}(\nabla\omega_{ij})^2, \tag{10}$$

where (having in mind the qualitative nature of the model) we assumed elastic isotropy of the network of hydrogen bonds and the elastic modulus $\alpha_{\rm el}$ depends on R as does $\alpha_{\rm ch}$.

The two expressions (9) and (10) may be simplified still further if we assume that the twist only changes substantially in one direction linking the centers of mass of the molecules. In this case, the twist matrix is reduced to the angle of rotation θ about this direction. Then the energies (9) and (10) reduce to

$$E = \frac{1}{2}\alpha_{\rm ch}(\theta - \theta_0)^2 + \frac{1}{2}\alpha_{\rm el}\left(\frac{d\theta}{dz}\right)^2 \tag{11}$$

(the z-axis is selected in the direction indicated above). The equilibrium angle of twist θ_0 induced by the NA

[4] Note that the correlation lengths λ_0 and λ_s of the pseudospin order generally differ from the characteristic scales of decrease of the hydration forces λ_h discussed in the previous section.

molecules should depend on z according to the "cholesteric" law [9]:

$$\theta_0 = q_0 z \tag{12}$$

(q_0 is the NA-induced wave vector of the twist modulation; it is natural to assume that q_0 is related to the helix pitch h of the secondary structure of the NA molecules, i.e., $q_0 \sim 1/h$).

A simple procedure for minimizing the energy (11) under the condition (12) gives the criterion for the absence of twist:[5]

$$\frac{1}{2}\alpha_{ch} < \alpha_{el} q_0^2. \tag{13}$$

In the opposite case ($\alpha_{ch} > 2\alpha_{el} q_0^2$) the network of hydrogen bonds becomes chiral and may then induce cholesteric ordering of the NA molecules. This induced ordering may enhance the "bare" cholesteric order associated with direct interactions between NA molecules (steric, van der Waals, and electrostatic) and may attenuate it.

We considered the region of scales where $R \le \lambda_0$. On large scales $\lambda_s \ge R > \lambda_0$ the elasticity of the network of hydrogen bonds α_{el} and the induced chirality α_{ch} may be neglected. Thus, at these distances there is no mean-field effective interaction between NA segments occurring as a result of deformation of the network of hydrogen bonds. Nevertheless (as we noted in Section 4) on these scales the properties of the water surrounding the NA molecule are modified substantially and specifically, the frequency of proton tunneling along the hydrogen bonds and in consequence, the tunnel splitting energy ε which appears in Eq. (8) in the pseudospin representation of the hydrogen bonds, are modified.

As we know [37, 38] in the pseudospin formalism the system of hydrogen bonds of water may be represented by a Hamiltonian having the following form:

$$H = -\Omega \sum_l S_l^x + \frac{1}{2}\sum_{l,l'} [J_3 S_l^z S_{l'}^z + J_2 S_l^y S_{l'}^y + J_1 S_l^x S_{l'}^x]. \tag{14}$$

In terms of its physical meaning the x-component of the pseudospin defines the dipole moment operator, the y-component determines the local current operator, and the x-component determines the tunneling operator. The coefficients of the pseudospin Hamiltonian (14) are expressed in terms of the proton tunneling energy along a particular hydrogen bond (numbered l) and the matrix elements of the proton interaction at neighboring hydrogen bonds, the operators S^x, S^y, S^z possess the commutation relationships of Pauli matrices and for solid ferroelectrics for which the pseudospin formalism is also used [37] we usually have $\Omega > J_1 > J_2 \sim J_3$. However, for water (and in particular for the "nearer" water surrounding the NA molecules) this relationship between the parameters of the Hamiltonian is not proven and thus we retained all the terms in (14).

The next important step is to allow for interaction between the pseudospin system and the NA molecules. We shall consider two molecules having centers of mass at points $\mathbf{r}_1$ and $\mathbf{r}_2$. Then as a result of interaction between these molecules and the nearer water the following terms are added to the Hamiltonian (14)

$$\delta H = \int d^3 r \sum_{j=1}^{3} \sum_m [\delta J_j \delta(\mathbf{r} - \mathbf{r}_1) S_m^j S_{m+1}^j + \delta J_j \delta(\mathbf{r} - \mathbf{r}_2) S_m^j S_{m+1}^j]. \tag{15}$$

Here δJ_j is the change in the corresponding parameters of the pseudospin Hamiltonian (14) describing modification of the tunneling in hydrogen bonds as a result of interaction between molecules of the nearer water and segments of NA molecules. By calculating the partition functions of a system with the Hamiltonian (14) and (15) using some approximation, we can find the contribution to the free energy which depends on the difference between the coordinates of the NA molecules $\mathbf{r}_1 - \mathbf{r}_2$, which has the meaning of the effective interaction between macromolecules produced by a change in the proton tunneling energy in the nearer water surrounding the macromolecules. However, bearing in mind the qualitative and highly approximate nature of our model, we express the final answer to within a numerical factor:

$$V_{12} = -\text{const} \times T \ln\left[1 + \frac{4l^2}{R^2}\sin^2\frac{\varphi}{2}\right]. \tag{16}$$

Here $R = |\mathbf{r}_1 - \mathbf{r}_2|$ is the distance between the centers of mass of rigid segments of NA molecules, l is the length of a segment of NA molecule, and φ is the angle between the long axes of the segments.

The physical meaning of Eq. (16) is fairly clear. This is in fact a reduction in the entropy of the gas of pseudospin excitations in the nearer water surrounding the NA molecules.

Summing all the interactions taking place between NA molecules in an aqueous–salt solution and bearing in mind the actual smallness of the angle of relative rotation of the long axes of the molecules, we can express the total pair potential in the following form:

$$V = v_0(R) + \frac{1}{2}v_1(R)\varphi^2 + \frac{1}{3}v_2(R)\varphi^3 - \frac{1}{2}U_1(R)\varphi^2 + \frac{1}{4}U_2(R)\varphi^4 + \frac{1}{3}\Gamma(R)\varphi^3. \tag{17}$$

[5] Note that even in the absence of any twist of the network, the chirality of the system leads to different energies of excitation of the fluctuations of different directions of twist, i.e., the fluctuation interaction remains chiral even in an achiral structure.

Here $v_0(R)$ is the isotropic part of the pair interaction, $v_1(R)$ is the anisotropic achiral part which occurs as a result of conventional (van der Waals, electrostatic, and steric) forces, $v_1(R)$ decreases according to a power law ($\propto R^{-n}$) with the intermolecular distance R, and the anisotropic anharmonic chiral contribution to the energy $v_2(R)$ has a similar nature (and dependence on R). The negative achiral contribution $\propto U_1(R)$ occurs as a result of entropy hydration forces (16) and thus decreases exponentially with R and the following term of the expansion in terms of φ from (17) proportional to $U_2(R)$ exhibits similar behavior. Finally, the chiral interaction $\propto \Gamma(R)$ is associated with the mean-field deformation of the orientation of the network of hydrogen bonds in the nearer water. The function $\Gamma(R)$ [and also $U_1(R)$ and $U_2(R)$] decrease exponentially with increasing R.

In principle, we can analyze the complete phase diagram of the system obtained by minimizing the potential (17). However, this type of detailed study is hardly justified at the present time since the theory includes too many unknown parameters [a minimum of 12 since we should define the potential amplitude and exponent or exponential function for each of the contributions appearing in (17)]. However, we note that in our model [the potential (17)] all the unusual (with respect to classical liquid crystals) properties of liquid-crystal NA dispersions described in the previous sections are given a qualitatively natural description. For this we need to add to the conventional theories of liquid-crystal order two new ingredients associated with the properties of an aqueous–salt solvent: these are the chiral mean-field interaction which is caused by deformation of the network of hydrogen bonds and also the entropy contribution from pseudospin excitations which occur in the water nearer to the macromolecules as a result of modification of the proton tunneling energy along the hydrogen bonds. We also note that the energy (17) contains a contribution proportional to φ^3 which may correspond to a first-order conformational phase transition (which is manifest in particular in the conformational hysteresis described in Section 2).

6. CONCLUSIONS

The main result of our study is the conclusion that an analysis of the entire combination of available experimental data and quantum-chemical calculations for NA dispersions shows that the aqueous–salt solvent plays an important role in the formation and properties of the solvent. A new model is proposed for the interaction of NA segments which allows for a change in the particular structure of the hydrogen bonds of the solvent in the presence of NA molecules. We used this model to derive the pair interaction potential (17) with which we can describe the complete set of experimental data known for aqueous–salt NA dispersions.

The principal new element of our theory is that (17) allows for entropy and orientational interactions between macromolecules in a gas of pseudospin excitations [terms containing the coefficients $U_1(R)$, $U_2(R)$, and $\Gamma(R)$ in (17)]. It is important to note that the quadratic contribution with respect to the angle made to the energy by the pseudospin excitations is negative, which leads to relative rotation of the long axes of the NA molecules (i.e., $\varphi \neq 0$) in neighboring quasi-nematic layers.

Consequently, allowance for modification of the solvent properties in the presence of NA molecules yields a fundamentally new picture of the formation of liquid-crystal NA dispersions. In our scheme the appearance of cholesteric ordering is associated with a phase transition between $\varphi \neq 0$ and $\varphi = 0$ states of the system which takes place at those densities for which

$$v_1(R) - U_1(R) = 0. \tag{18}$$

It is interesting to note that fairly rough estimates using Eq. (16) allowing for dipole–dipole forces for $v_1(R)$ give the value $R \sim 20$ Å from (18) which is consistent with the experimental data on the disappearance of cholesteric ordering at short distances.

Another scenario simulating the same nematic behavior of the structure has clearly not been eliminated. This is that, as the density of the dispersion increases, solidification of the dispersion must occur at fairly short distances ($R \leq 20$ Å) which is incompatible with the cholesteric twist of the molecules (since $\varphi \neq 0$ creates twist deformation which increases proportionately as the fourth power of the system size [36]).

However, if this solidification transition is merely a weak first-order transition, there should be an appreciable region of pre-transition fluctuation phenomena which lead to an increase in the pitch of the cholesteric helix (similar effects are known for a cholesteric–smectic phase transition, see, e.g., [9]).

It is easy to understand that this fluctuation contribution to the pitch of the cholesteric helix should be proportional to the correlation length for a phase transition involving partial or complete solidification and consequently in the mean field approximation we have

$$p \simeq p_0\left[1 + \left(\frac{R_c}{R - R_c}\right)^{3/2}\right]. \tag{19}$$

Here p_0 is the pitch of the cholesteric helix outside the fluctuation region, i.e., for $R - R_c \gg R_c$ where R_c is the critical intermolecular distance at which the dispersion solidifies. Bearing in mind the highly anisotropic form of the rigid segments of nucleic acid molecules ($d \approx 20$ Å, $l \approx 500$ Å), we can naturally expect partial rather than complete solidification of the cholesteric dispersion accompanied by the formation of a so-called discotic liquid crystal, i.e., a system which is liquid in the direction of preferential orientation of the long axes of the macromolecules and is characterized by a two-dimen-

sional hexagonal lattice in the perpendicular plane. The possibility of a phase transition to the discotic rather than the solid phase is also indicated by the absence of higher Bragg reflections in experiments [3] and by the conservation of appreciable fluidity in the dispersion even at $R \approx 20$ Å.[6]

If as a result of a fluctuation increase the helix pitch (19) exceeds the size of the system before a cholesteric–nematic phase transition takes place according to the estimate (18), at this point the anomalous circular dichroism band interpreted experimentally as a cholesteric–nematic transition will disappear.

In principle we can hope that both scenarios for vanishing circular dichroism [as a result of a cholesteric–nematic phase transition (18) induced by hydration forces or as a result of a fluctuation increase in the pitch of the cholesteric helix (19)] can be distinguished experimentally. For a direct cholesteric–nematic phase transition the corresponding mean-field index of the pretransition untwisting of the cholesteric is 0.5 whereas the fluctuation contribution (19) gives a faster increase in pitch (the index is 1.5). Unfortunately, the available experimental data cannot be used for a quantitative comparison with these theoretical predictions because the accessible range of variation of the NA concentration is less than one order of magnitude (see, e.g., [40]) where the authors discussed the possibility of fluctuation untwisting of the cholesteric dispersion of NA during complete solidification, i.e., a cholesteric–crystal phase transition). The dependence of the helix pitch on the length l of the rigid NA segments can also be checked. A theory based only on direct electrostatic interactions in accordance with [7] for an equilibrium angle between neighboring quasi-nematic layers gives $\varphi_{eq} \propto 1/l$ whereas our entropy fluctuation effects associated with water [see Eqs. (16), (17)] give the dependence $\varphi_{eq} \propto 1/l^2$.

ACKNOWLEDGMENTS

This work was carried out as part of the Ministry of Science Program "Statistical Physics" and was partly supported by INTAS (grant 30234) and the Russian Foundation for Basic Research (project no. 00-02-17785).

REFERENCES

1. L. S. Lerman, Proc. Natl. Acad. Sci. USA **68**, 1886 (1971).
2. Yu. M. Yevdokimov, S. G. Skuridin, and V. I. Salynov, Liq. Cryst. **3**, 1443 (1988).
3. F. Livolant and M. F. Maestre, Biochemistry **27**, 3056 (1988); F. Livolant and A. Leforestier, Prog. Polym. Sci. **21**, 1115 (1996).
4. Yu. M. Evdokimov, V. I. Salyanov, V. L. Golo, *et al.*, Sens. Sist. **14**, 245 (2000).
5. A. Ya. Grosberg and A. R. Khokhlov, Sov. Sci. Rev., Sect. A **8**, 147 (1987).
6. A. B. Harris, R. D. Kamien, and T. C. Lubensky, Rev. Mod. Phys. **71**, 1745 (1999).
7. A. A. Kornychev and S. Leikin, J. Chem. Phys. **107**, 3656 (1997); Proc. Natl. Acad. Sci. USA **95**, 13579 (1998); Phys. Rev. Lett. **84**, 2537 (2000).
8. S. Chandrasekhar, *Liquid Crystals* (Cambridge Univ. Press, Cambridge, 1977; Mir, Moscow, 1980).
9. P. G. de Gennes, *The Physics of Liquid Crystals* (Clarendon Press, Oxford, 1974; Mir, Moscow, 1977).
10. Y. R. Lin-Liu, Yu Ming Shih, Chia-Wei Woo, and H. T. Tan, Phys. Rev. A **14**, 445 (1976).
11. B. Samori, M. A. Osipov, I. Domini, and A. Bartolini, Int. J. Biol. Macromol. **15**, 353 (1993).
12. T. V. Samulski and E. T. Samulski, J. Chem. Phys. **67**, 824 (1977).
13. V. I. Salyanov and Yu. M. Evdokimov, Dokl. Akad. Nauk **368**, 700 (1999).
14. Yu. M. Evdokimov, S. G. Skuridin, and G. B. Lortkipanidze, Liq. Cryst. **12**, 1 (1992).
15. Yu. M. Evdokimov, S. G. Skuridin, S. V. Semenov, *et al.*, Biofizika **43**, 240 (1998).
16. S. Skuridin, N. Badaev, A. Dembo, *et al.*, Liq. Cryst. **23**, 51 (1998).
17. Yu. M. Evdokimov, V. I. Salyanov, A. T. Dembo, and F. Spener, Sens. Sist. **13**, 159 (1999).
18. W. Saenger, *Principles of Nucleic Acid Structure* (Springer-Verlag, New York, 1984).
19. Y. Yevdokimov, V. Salyanov, and M. Palumbo, Liq. Cryst. **131**, 285 (1985); V. I. Salyanov, M. Palumbo, and Yu. M. Evdokimov, Mol. Biol. **27**, 869 (1993).
20. B. W. van der Meer, G. Vertogen, A. J. Dekker, and J. G. J. Ypma, J. Chem. Phys. **65**, 3935 (1976); E. I. Kats, Zh. Éksp. Teor. Fiz. **74**, 2320 (1978) [Sov. Phys. JETP **47**, 1205 (1978)].
21. R. Podgornik and V. A. Parsegian, Macromolecules **23**, 2265 (1990).
22. S. Neidle, H. M. Berman, and H. S. Shieh, Nature **288**, 129 (1980).
23. E. Clementi and G. Corongiu, J. Chem. Phys. **72**, 3979 (1980).
24. W. Saenger, Nature **279**, 343 (1979).
25. J. D. Bernal and R. H. Fowler, J. Chem. Phys. **1**, 515 (1933).
26. N. A. Bul'enkov, Biofizika **36**, 181 (1991).
27. E. Clementi and G. Corongiu, Biopolymers **18**, 2431 (1979).
28. K. Kim and M. S. John, Biochim. Biophys. Acta **565**, 131 (1979).
29. D. Perahia, M. S. John, and B. Pullman, Biochim. Biophys. Acta **474**, 349 (1977); G. Minasov, V. Tereshko, B. Chernov, and L. Malinina, J. Cryst. Growth **122**, 136 (1992).

[6] In recent studies [39] based on results of small-angle X-ray scattering and theoretical analyses using Landau theory it is stated that the phase with parallel packed NA molecules of length much greater than the persistent length is a biaxial liquid crystal with nematic order of the long axes and hexatic order of the short axes.

30. M. L. Kopka, A. Fratani, H. R. Drew, and R. E. Dickerson, J. Mol. Biol. **163**, 129 (1983).
31. L. B. Boinovich and A. M. Emelyanenko, Z. Phys. Chem. (Munich) **178**, 229 (1992).
32. P. Gallo, cond-mat/0003027 (2000).
33. A. Goldblum, D. Perahia, and A. Pullman, FEBS Lett. **91**, 213 (1978).
34. S. Leikin, D. C. Rau, and A. V. Parsegian, Phys. Rev. A **44**, 5272 (1991).
35. R. Blinc, J. Phys. Chem. Solids **13**, 204 (1960).
36. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 7: *Theory of Elasticity* (Nauka, Moscow, 1982; Pergamon Press, New York, 1986).
37. R. Blinc and B. Zeks, *Soft Modes in Ferroelectrics and Antiferroelectrics* (North-Holland, Amsterdam, 1974; Mir, Moscow, 1975).
38. J. M. Ziman, *Models of Disorder: the Theoretical Physics of Homogeneously Disordered Systems* (Cambridge Univ. Press, Cambridge, 1979; Mir, Moscow, 1982).
39. H. H. Strey, J. Wang, R. Podgornik, *et al.*, Phys. Rev. Lett. **84**, 3105 (2000); R. D. Kamien and A. J. Levine, Phys. Rev. Lett. **84**, 3109 (2000).
40. D. H. van Winkle, M. W. Davidson, W. X. Chen, and R. L. Rill, Macromolecules **23**, 4140 (1990).

Translation was provided by AIP