Journal of Experimental and Theoretical Physics, Vol. 93, No. 1, 2001, pp. 1–23.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 1, 2001, pp. 5–30.

NUCLEI, PARTICLES, AND THEIR INTERACTION

Summing Divergent Perturbative Series in a Strong Coupling Limit. The Gell-Mann–Low Function of the φ^4 Theory

I. M. Suslov

Kapitza Institute of Physical Problems, Russian Academy of Sciences, Moscow, 117334 Russia
e-mail: suslov@kapitza.ras.ru

Received December 27, 2000

Abstract—An algorithm is proposed for determining asymptotics of the sum of a perturbative series in the strong coupling limit using given values of the expansion coefficients. Application of the algorithm is illustrated, methods for estimating errors are developed, and an optimization procedure is described. Applied to the φ^4 theory, the algorithm yields the Gell-Mann–Low function asymptotics of the type $\beta(g) \approx 7.4g^{0.96}$ for large g. The fact that the exponent is close to unity can be interpreted as a manifestation of the logarithmic branching of the type $\beta(g) \sim g(\ln g)^{-\gamma}$ (with $\gamma \approx 0.14$), which is confirmed by independent evidence. In any case, the φ^4 theory is self-consistent. The procedure of summing perturbative series with arbitrary values of the expansion parameter is discussed. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

This paper presents a systematic description of the algorithm proposed previously in a brief communication [1]. Operation of the algorithm is illustrated by test examples, methods for estimating errors are developed, and an optimization procedure is described. Using this algorithm, the Gell-Mann–Low function of the φ^4 theory—the main physical result of this study—can be reconstructed with a tenfold greater precision.

The abstract formulation of the problem is as follows. Let some function $W(g)$ be expanded into a series of the perturbation theory in powers of a coupling constant g:

$$W(g) = \sum_{N=0}^{\infty} W_N(-g)^N. \quad (1)$$

The first several expansion coefficients W_N can be obtained by straightforward diagram calculations. The high-order terms can be determined using the Lipatov method [2], which is applicable to most of the important problems and yields for W_N an asymptotic behavior of the type (see reviews [3–5]):

$$W_N^{as} = ca^N\Gamma(N+b) \approx ca^N N^{b-1}N!. \quad (2)$$

Matching asymptotics (2) to the first coefficients provides information about all terms of the series and allows the $W(g)$ function to be approximately restored, but this procedure requires using special methods for summing divergent series. Implementation of this approach allowed the critical indices of the phase transition theory to be determined to within the third decimal position [6–8], thus rendering the intermediate coupling region ($g \sim 1$) principally accessible. However, this direction was not developed further because the problem of renormalon contributions arose that cast doubt [9] on the applicability of the Lipatov method. The interest in this field had dropped sharply and no breakthrough into the strong coupling region took place.

Expanding the theory into the strong coupling region is required in many fields of theoretical physics. The most known cases, related to the dependence of the effective coupling constant g on the spatial scale L, include the problem of electrodynamics at very small distances and the confinement problem. The dependence of g on L in renormalizable theories is determined by the equation

$$-\frac{dg}{d\ln L^2} = \beta(g) = \beta_2 g^2 - \beta_3 g^3 + \beta_4 g^4 - \ldots. \quad (3)$$

In the general case, this description requires information on the Gell-Mann–Low function $\beta(g)$ for arbitrary g. The possible variants were classified by Bogolyubov and Shirkov [10]. In the case of $\beta_2 > 0$, the situation reduces to the following. If the function $\beta(g)$ possesses a root at g_0, then $g(L) \longrightarrow g_0$ as $L \longrightarrow 0$. If $\beta(g)$ at large g behaves as g^α with $\alpha \le 1$, then $g(L) \longrightarrow \infty$ at small L; should $\beta(g)$ grow as g^α with $\alpha > 1$, the theory is no longer self-consistent and cannot describe the behavior of $g(L)$ in the entire range of L.

1063-7761/01/9301-0001$21.00

The first attempt at restoring the β function in the φ^4 theory with the Euclidean action

$$S\{\varphi\} = \int d^4x\left\{\frac{1}{2}(\partial\varphi)^2 + \frac{16\pi^2}{4!}g\varphi^4\right\} \quad (4)$$

was undertaken by Popov *et al.* [11]. The Shirkov group attempted to move into the strong coupling region [12] and obtained for large g asymptotics of the type $0.9g^2$, which differs only by a coefficient from a one-loop law $1.5g^2$ valid for $g \longrightarrow 0$. A close asymptotic behavior ($1.06g^{1.9}$) was obtained by Kubyshin [13], while the more recently developed variational perturbation theory of Sissakian *et al.* [14] yields $2.99g^{1.5}$. All these results give evidence that the φ^4 theory is not self-consistent.[1] This is, however, rather strange from the standpoint of condensed-matter applications, where a quite reasonable disordered system model [16, 17] well defined in the continuum limit is mathematically strictly reduced to the φ^4 theory. Another argument follows from the author's recent study [9] showing the φ^4 theory to contain no renormalon singularities, which can be considered as evidence of self-consistency. This situation makes revision of the above results an urgent task.

In this paper, an algorithm is proposed for restoring asymptotics of the sum of a perturbative series in the strong coupling limit using given values of the expansion coefficients (Section 2). Application of the algorithm is illustrated by test examples with both known expansion coefficients (Section 4) and the coefficients obtained by interpolation (Sections 5 and 6). Methods for estimating errors and an optimization procedure are developed (Sections 3 and 6). The problem of summing the perturbative series with finite g is considered, and it is demonstrated that knowledge of the $W(g)$ asymptotics significantly increases precision of the results (Section 7). The main physical result of this study consists in reconstructing the Gell-Mann–Low function of the φ^4 theory (Section 8). The task is solved proceeding from the same information as that used in [13], namely, the first four coefficients of expansion of the $\beta(g)$ function in the subtraction scheme [15, 18]

$$\beta(g) = \frac{3}{2}g^2 - \frac{17}{6}g^3 + \frac{154.14}{8}g^4 - \frac{2338}{16}g^5 + \ldots, \quad (5)$$

and their asymptotics according to Lipatov, taking into account the first-order correction [19]:

$$\beta_N = \frac{1.096}{16\pi^2}N^{7/2}N!\left\{1 - \frac{4.7}{N} + \ldots\right\}. \quad (6)$$

[1] It should be noted that Kazakov *et al.* [12] do not insist on this conclusion, emphasizing the preliminary character of their results (see also [15]).

Note that the interaction term in expression (4) corresponds to the "natural" charge normalization, for which the parameter a in asymptotics (2) is unity. It will be demonstrated that the results obtained in [12, 13] are not artifacts: they objectively reflect the behavior $\beta(g)$ in the interval $1 \lesssim g \lesssim 10$. However, the true asymptotics is manifested at still greater g and gives evidence of self-consistency of the φ^4 theory.

2. RELATIONSHIP BETWEEN $W(g)$ ASYMPTOTICS AND EXPANSION COEFFICIENTS

Let us formulate the problem of restoring the asymptotics

$$W(g) = W_\infty g^\alpha, \quad g \longrightarrow \infty, \quad (7)$$

using the coefficients W_N of the series (1). The coefficients with large N, increasing according to the factorial law (2), are assumed to be set numerically. By analogy with the case of critical indices introduced in the phase transition theory, the slow (logarithmic) corrections to (7) are considered as overstating the accuracy. For exponentially growing $W(g)$, which can be revealed by abnormally large values of α, the series (1) is considered upon preliminarily taking the logarithm.

2.1. Standard (Conform-Borel) Summing Procedure

Considering the sum of series (2) in the Borel sense [20], we use a modified definition of the Borel image $B(g)$,

$$W(g) = \int_0^\infty dx e^{-x}x^{b_0-1}B(gx),$$
$$B(g) = \sum_{N=0}^\infty B_N(-g)^N, \quad B_N = \frac{W_N}{\Gamma(N+b_0)}, \quad (8)$$

where b_0 is an arbitrary parameter (convenient for optimization of the summation procedure [6]). It was suggested by Le Guillou and Zinn-Justin [6] and recently proved for the φ^4 by the author [9] that the Borel image is analytical in the complex plane g cut from $-1/a$ to $-\infty$ (Fig. 1a). The analytical continuation of $B(g)$ from the convergence circle $|g| < 1/a$ to an arbitrary complex g value is provided by a conformal mapping $g = f(u)$ of the plane with a cut into a unity circle $|u| < 1$ (Fig. 1b). The re-expansion of $B(g)$ into a series in u,

$$B(g) = \sum_{N=0}^\infty B_N(-g)^N\Big|_{g=f(u)} \longrightarrow B(u) = \sum_{N=0}^\infty U_N u^N, \quad (9)$$

gives a series converging for any g. Indeed, all the possible singular points (P, Q, R, …) of the $B(g)$ function occur on the cut and their images (P, Q, Q', R, R', …) fall on the boundary $|u| = 1$ of the circle. Therefore, the

second series in (9) converges at any $u < 1$, but the interior of this circle is in a single-valued correspondence with the region of analyticity in the g plane.

The conformal mapping is defined by the formulas

$$g = \frac{4}{a}\frac{u}{(1-u)^2} \quad \text{or} \quad u = \frac{(1+ag)^{1/2}-1}{(1+ag)^{1/2}+1}, \tag{10}$$

from which we readily find a relationship between U_N and B_N:

$$U_0 = B_0, \quad U_N = \sum_{K=1}^{N} B_K\left(-\frac{4}{a}\right)^K C_{N+K-1}^{N-K} \quad (N \geq 1). \tag{11}$$

In order to establish a relationship between asymptotics (7) and the expansion coefficients, we will use the fact that the behavior of U_N at large N is determined by a sum of the contributions from singular points occurring on the boundary $|u| = 1$. This can be readily checked by expressing U_N in terms of $B(u)$,

$$U_N = \oint_C \frac{du}{2\pi i}\frac{B(u)}{u^{N+1}}, \tag{12}$$

and deforming the integration contour (enclosing the point $u = 0$) so as to make it passing around the cuts from all singular points to infinity. A singularity of the type $A(1-u/u_0)^\beta$ at the point $u_0 = e^{i\varphi}$ makes a contribution to U_N of the type

$$\frac{A}{\Gamma(-\beta)}\frac{e^{-i\varphi N}}{N^{1+\beta}}. \tag{13}$$

Now we can readily find the contributions to U_N from the singular points of the initial Borel image $B(g)$. For power singularities at the points $g = \infty$, $g = -1/a$, and $g = g_0$ with $g_0 \in (-\infty, -1/a)$, the corresponding expressions are as follows:

$$B(g) = Ag^\alpha \longrightarrow U_N = \frac{A}{\Gamma(2\alpha)}\left(\frac{4}{a}\right)^\alpha \frac{1}{N^{1-2\alpha}},$$

$$B(g) = A(g+1/a)^\beta \longrightarrow U_N = \frac{A}{(4a)^\beta\Gamma(-2\beta)}\frac{(-1)^N}{N^{1+2\beta}}, \tag{14}$$

$$B(g) = A(g-g_0)^\beta \longrightarrow U_N = \frac{2A}{\Gamma(-\beta)}\left(\frac{\cos(\varphi/2)}{a\sin^3(\varphi/2)}\right)^\beta \frac{\cos(\varphi N - \pi\beta/2)}{N^{1+\beta}},$$

where $\varphi = \arccos(1 + 2/ag_0)$.

The singularities of $B(g)$ change depending on the parameter b_0 in formulas (8). For the Borel images $B(g)$ and $\tilde{B}(g)$ corresponding to b_0 and b_1, we readily obtain a recalculation formula

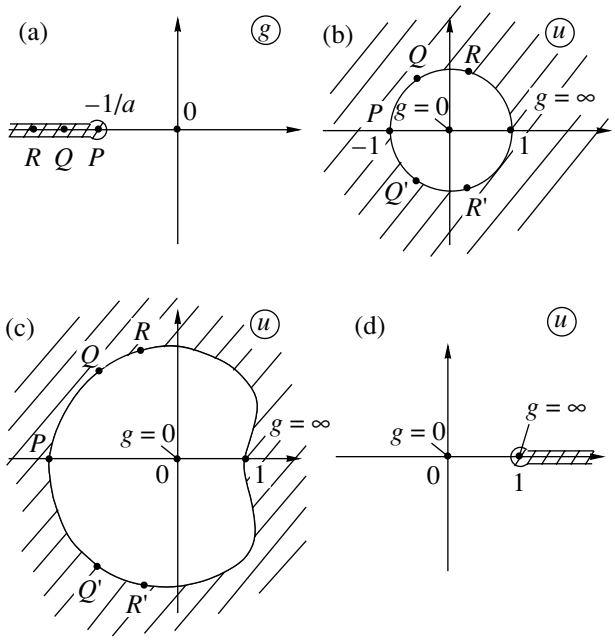

Fig. 1. (a) The Borel image is analytical in the complex plane with $(-\infty, -1/a)$ cut; (b) this analyticity region can be conformally mapped onto the unity circle; (c) restricting the consideration to analytical continuation to the positive semiaxis, the conformal mapping is admitted onto any region in which the point $u = 1$ is the closest boundary point to the origin; (d) in the extremal form (18) of this mapping, the analyticity region can be conformally mapped onto the plane with $(1, \infty)$ cut.

$$\tilde{B}(g) = \frac{1}{\Gamma(b_1-b_0)}\int_0^\infty dx \frac{x^{b_1-b_0-1}}{(1+x)^{b_1}} B\left(\frac{g}{1+x}\right), \tag{15}$$

and a rule of singularity transformation at a finite (g_0) or infinite points on the passage from b_0 to b_1:

$$B(g) = A\Gamma(-\beta)\left(\frac{g_0-g}{g_0}\right)^\beta \longrightarrow \tilde{B}(g)$$

$$= A\Gamma(-\beta-b_1+b_0)\left(\frac{g_0-g}{g_0}\right)^{\beta+b_1-b_0}, \tag{16}$$

$$B(g) = \frac{A}{\Gamma(\alpha+b_0)}g^\alpha \longrightarrow \tilde{B}(g) = \frac{A}{\Gamma(\alpha+b_1)}g^\alpha.$$

As is seen, an increase in b_0 weakens the singularities at a finite point, while the character of singularity at infinity remains unchanged. For sufficiently large b_0, the contributions from finite points to U_N are suppressed and the corresponding asymptotic behavior is determined by the singularity of $B(g)$ (and, hence, of $W(g)$) at $g \longrightarrow \infty$:

$$U_N = \frac{W_\infty}{\Gamma(2\alpha)\Gamma(b_0+\alpha)}\left(\frac{4}{a}\right)^\alpha N^{2\alpha-1}, \quad N \longrightarrow \infty. \tag{17}$$

This formula solves the problem: the coefficients U_N are related by a linear transformation (11) to the initial coefficients W_N (see Eq. (8)), while their asymptotic behavior (17) determine the parameters W_∞ and α of asymptotics (7).

Formulas (14) indicate that a contribution to U_N from the singular point $g = \infty$ is monotonic, while the contributions from other points are oscillating. Therefore, increasing b_0 leads to a change in the U_N behavior from oscillating to monotonic. This phenomenon was observed in [6] and, albeit not given any satisfactory explanation, regularly employed for improving the divergence of perturbative series.

2.2. Modified Conformal Mapping

A more effective algorithm is provided by using a modified conformal mapping.

According to the Riemann theorem [21], the conformal mapping of a simply connected region into a unity circle is single-valued to within the so-called normalization, which can be defined by setting the images of two (internal and boundary) points. Under the convention that the point $g = 0$ is imaged by $u = 0$ and $g = \infty$ by $u = 1$, conformal mapping (10) is the only one that allows the Borel image to be analytically continued to arbitrary complex g values. However, this is not necessary: to perform the integration in (8), the analytical continuation to positive semiaxis is sufficient. Then, any conformal mapping into a region of the type depicted in Fig. 1c is admissible, in which the point $u = 1$ is the boundary point closest to the origin. The second series in expansion (9) is convergent at $u < 1$ and, in particular, in the interval $0 < u < 1$ imaging the positive semiaxis. An advantage of this conformal mapping is that the contributions from singular points P, Q, Q', R, R' … to U_N are exponentially suppressed and the U_N asymptotics for all b_0 is determined by a contribution of the singular point at $u = 1$ related to the singularity of $W(g)$ at $g \longrightarrow \infty$.

Let us use an extremal form of such mapping, imaging the plane with cut $(-\infty, -1/a)$ into the plane with cut $(1, \infty)$ (Fig. 1d). This mapping is given by the formula

$$g = \frac{u}{a(1-u)}, \tag{18}$$

which leads to the following relationship between U_N and B_N:

$$\begin{aligned} U_0 &= B_0, \\ U_N &= \sum_{k=1}^{N} \frac{B_K}{a^K}(-1)^K C_{N-1}^{K-1} \quad (N \geq 1). \end{aligned} \tag{19}$$

The asymptotic behavior of U_N for large N is

$$U_N = U_\infty N^{\alpha-1}, \quad N \longrightarrow \infty, \tag{20}$$

$$U_\infty = \frac{W_\infty}{a^\alpha \Gamma(\alpha)\Gamma(b_0+\alpha)}. \tag{21}$$

As a result, we arrive at a simple algorithm: calculate coefficients B_N by formula (8) using preset W_N, recalculate B_N to U_N using relationship (19), and take the power limit (20) for large N to determine parameters W_∞ and α for asymptotics (7).

2.3. Random Error Growth

The above algorithms possess an implicit drawback that significantly restricts the accuracy of description. Let us introduce a reduced coefficient function:

$$\begin{aligned} F_N &= \frac{W_N}{W_N^{as}} = \frac{W_N}{ca^N\Gamma(N+b)} \\ &= 1 + \frac{A_1}{N} + \frac{A_2}{N^2} + \ldots + \frac{A_K}{N^K} + \ldots, \end{aligned} \tag{22}$$

which varies within finite limits and admits a regular expansion in the powers of $1/N$. The latter can be checked by calculating sequential corrections to the Lipatov asymptotics [19]. In practice, F_N is set with a certain accuracy δ_N (calculation or round-off error), which leads to a random error in U_N. The error dispersion for the algorithm considered in Section 2.2 is as follows:

$$(\delta U_N)^2 = \sum_{K=1}^{N} \left[c\delta_K \frac{\Gamma(K+b)}{\Gamma(K+b_0)} C_{N-1}^{K-1} \right]^2. \tag{23}$$

For the round-off errors, the value of $\delta_K = \delta$ is independent of K. A sum calculated by the steepest descent method for large N,

$$\delta U_N \sim 2^N \delta, \tag{24}$$

demonstrates a catastrophic growth of the error. Calculation with a double computer accuracy yields $\delta \sim 10^{-14}$, so that δU_N is on the order of unity for $N \approx 45$.[2] This restricts the accuracy of determining the parameters of asymptotics (7) to approximately 1%. According to expression (23), an increase in b_0 decreases the error so that the permissible N level grows. However, large b_0 values delay the process of attaining the asymptote (20), so that no advantages are eventually gained.

For the algorithm considered in Section 2.1, the error grows at a still higher rate,

$$\delta U_N \sim (\sqrt{2}+1)^{2N}\delta \sim 5.8^N \delta, \tag{25}$$

and the requirement of using sufficiently large b_0 significantly restricts the possibility of optimization (see Section 3). Nevertheless, this algorithm may still be useful

[2] This error growth is observed in fact in the form of rapidly increasing irregular oscillations.

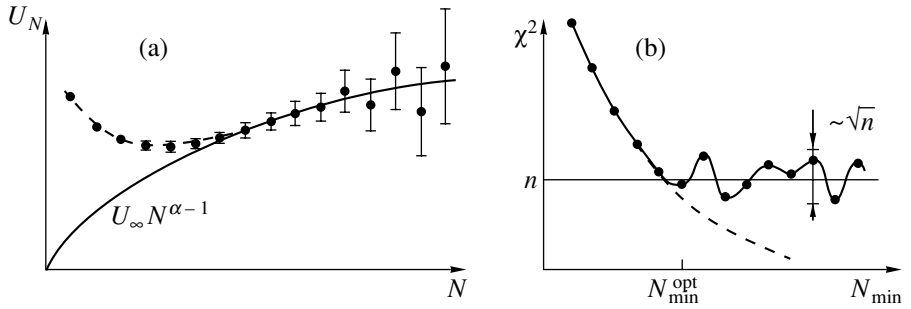

Fig. 2. The U_N treatment according to the power law: (a) a typical situation whereby large N correspond to a large statistical error and small N, to a large systematic error; (b) the plot of χ^2 versus $N_{\min}$ at a constant number of points n.

to increase the accuracy of calculations in the region of small g (Section 7). Below we dwell on the algorithm of Section 2.2 based on a modified conformal mapping, which offers indisputable advantages in the region of strong coupling.

The above considerations indicate that the computer round-off errors restrict the accuracy of the algorithm to ~1% even for test examples where the W_N values are precisely known. In real cases, the accuracy of W_N calculations is much worse and the situation might appear as hopeless. However, this is not so in fact because we mostly deal with interpolation errors, the influence of which has a quite different character. The linear relationship (19) known in mathematics as the Hausdorff transformation [20] possesses a remarkable property

$$\sum_{K=1}^{N} K^m(-1)^K C_{N-1}^{K-1} = 0, \quad m = 0, 1, \ldots, N-2, \quad (26)$$

that makes smooth errors (well approximated by polynomials) insignificant even despite their large magnitude.[3] Of course, limitations related to the computer round-off error are still valid, but a 1% accuracy is quite sufficient for real problems and this level can hardly be improved for the level of information accessible at present.

Strictly speaking, the problem of round-off errors is purely technical and can be solved by means of special precise arithmetic programs which allow the calculations to be performed with arbitrary number of significant digits [22], however, the accuracy of α and W_∞ restoration logarithmically depends on the computation accuracy. Algorithms that are more perfect in this respect do exist, but their consideration falls outside the scope of this paper; such methods, albeit providing for a high accuracy in the test examples, are insufficiently robust and work unsatisfactorily under conditions of restricted information. The algorithm under consideration is quite stable and, in the author's opinion, ideally suited to obtaining a reliable zero-order approximation.[4]

Treating U_N by the power law can involve a standard procedure of minimization of χ^2 [22]:

$$\chi^2 = \sum_i \left(\frac{y_i - y(x_i)}{\sigma_i}\right)^2, \quad (27)$$

where y_i are the values set at the points x_i with a statistical error σ_i and fitted to the theoretical function $y(x)$. In this process, it is important to select properly the interval $N_{\min} \le N \le N_{\max}$ for the U_N treatment. Indeed, large N values lead to large statistical errors determined by formula (23), while small N values increase the role of a systematic error related to the fact that U_N still did not attain asymptote (20) (Fig. 2a). A sufficiently high upper limit $N_{\max}$ can be chosen using the condition $\delta U_N \sim U_N$, since the points with greater N provide no additional information; this choice is not very critical since the procedure of χ^2 minimization automatically discriminates the points with large statistical errors, which are used in averaging with a weight of $1/\sigma_i^2$. The lower limit $N_{\min}$ has to be selected taking into account the χ^2 value, which reaches an extremely high level for small $N_{\min}$ but attains a "normal" level of $n \pm \text{const}\sqrt{n}$ (n is the number of points) with increasing $N_{\min}$ (Fig. 2b). The optimum value of $N_{\min}$ corresponds to the left end of the "plateau," where a systematic error becomes smaller than the statistical error and the available information is most completely employed.

In fact, the conditions for a strict statistical treatment of χ^2 were not fulfilled because the errors δU_N for various N were not independent (see Eq. (23)). This was

[3] This implies that, in the case when many W_N values are known with low precision, the data should be used upon approximation by a smooth function rather than directly.

[4] This situation is well known in computational mathematics [22]. All algorithms can be roughly divided into two groups: those in the first group possess moderate accuracy and convergence rate but are highly reliable (an example is offered by seeking for a root of equation through segment halving); algorithms of the second group show high accuracy and ensure rapid convergence but pose stringent requirements with respect to the function smoothness (e.g., in seeking a root with the forecast for several derivatives).

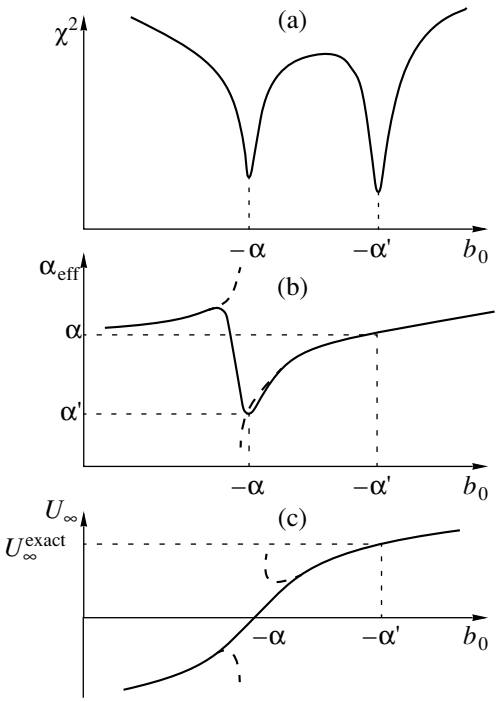

Fig. 3. Theoretical plots of (a) χ^2, (b) α_{eff}, and (c) U_∞ versus b_0 constructed with neglect of the correction terms indicated by dots in expression (29).

manifested by the fact that χ^2 values decreased below the "normal" level (dashed curve in Fig. 2b), while the statistical uncertainty of α and W_∞ became very small and did not reflect real errors even in the order of magnitude. For this reason, we considered the choice of $N_{\min}$ as satisfactory when the χ^2 values were on the correct order of magnitude ($\sim n$); small changes in $N_{\min}$ did not significantly influence the results.

3. DEPENDENCE ON THE PARAMETER b_0 AND ACCURACY ESTIMATION

Direct application of the algorithm described in Section 2.2 is insufficiently effective since the results depend on the arbitrary parameter b_0, which implies that an additional investigation is necessary to select the optimum value.

It is naturally expected that corrections to asymptote (7) have the form of a regular expansion with respect to $1/g$. However, even the simplest examples show that, in the general case, this assumption is not valid: in the zero-dimensional case, the corrections follow the powers of $g^{-1/2}$; for an anharmonic oscillator, the corrections follow the powers of $g^{-2/3}$ (see Section 4). For this reason, we admit the power corrections in the general form:

$$W(g) = W_\infty g^\alpha + W'_\infty g^{\alpha'} + \dots. \tag{28}$$

Accordingly, the asymptotic behavior of U_N written by analogy with (20) and (21) is described as

$$U_N = \frac{W_\infty}{a^\alpha \Gamma(\alpha)\Gamma(b_0+\alpha)} N^{\alpha-1} + \frac{W'_\infty}{a^{\alpha'}\Gamma(\alpha')\Gamma(b_0+\alpha')} N^{\alpha'-1} + \dots. \tag{29}$$

First, let us neglect the correction terms indicated by dots in expansion (29). A formal treatment of this expression according to the power law (20) yields quite satisfactory results because the truncated function (29) in the double logarithmic scale varies smoothly and is well approximated by a straight line. However, this approximation only leads to certain effective values of α and U_∞.

Note that, because of the poles of the gamma function, the first and second terms in (29) become zero for $b_0 = -\alpha$ and $b_0 = -\alpha'$, respectively. These b_0 values correspond to the purely power laws, $U_N \propto N^{\alpha'-1}$ and $U_N \propto N^{\alpha-1}$, which results in increasing quality of the approximation and a sharply decreasing χ^2 value. Within a fixed working interval $N_{\min} \le N \le N_{\max}$, the pattern is as follows (Fig. 3): the χ^2 versus b_0 curve exhibits sharp minima at $b_0 = -\alpha'$ and $b_0 = -\alpha$; the effective index α_{eff} drops down to α' in the vicinity of $b_0 = -\alpha$ and is close to α outside this region (being exactly equal to α at $b_0 = -\alpha'$); the effective parameter U_∞ corresponds to exact W_∞ at $b_0 = -\alpha'$ and crosses the zero level in the vicinity of the point $b_0 = -\alpha$. The slope of a linear portion of the curve near this root is

$$U_\infty \approx \frac{W_\infty}{a^\alpha \Gamma(\alpha)}(b_0+\alpha), \tag{30}$$

which provides for an W_∞ estimate not too sensitive with respect to α errors. The rejected terms in (29) may only slightly perturb this pattern.

The pattern outlined above was actually observed, but the behavior of α_{eff} and U_∞ in the vicinity of $b_0 = -\alpha$ is usually discontinuous (as indicated by dashed branches in the curves of Fig. 3). However, this circumstance is not physically significant and only reflects features of the mathematical procedure involving taking logarithm of the U_N modulus,

$$\ln|U_N| = \ln|U_\infty| + (\alpha-1)\ln N, \tag{31}$$

followed by using a linear fitting algorithm [22]. The sign of U_∞ is determined by calculating χ^2 for $U_\infty = |U_\infty|$ and $-|U_\infty|$ and selecting a variant with the minimum value. This procedure leads to rather senseless results in the case of U_N changing sign, but this is only possible in a small vicinity of the point $b_0 = -\alpha$, while the sign of U_N outside this narrow interval is determined by the sign of the first term in the right-hand part of Eq. (29).

Smoothness of the $U_\infty(b_0)$ function is restored when the treatment according to power law (20) is performed by varying only U_∞ at a fixed (approximate) α value. Small variations of α virtually do not affect the position of the root of $U_\infty(b_0)$, while significantly influencing the W_∞ value determined from the slope of the linear relationship (30). The above considerations suggest four different methods for estimating the α index, based on (i) the α_{eff} value at the first minimum of χ^2 (counting from large b_0), (ii) the position of the second χ^2 minimum, (iii) the change in the sign of U_∞ upon the logarithmic treatment, and (iv) the change in the sign of U_∞ upon treatment at a fixed α value (taken equal to a preliminary estimate).

The first two estimates ensure, in the general case, a higher precision, since their uncertainty is determined by the ratio of rejected terms in the right-hand part of expansion (29) to the characteristic value of the first term outside the narrow vicinity of $b_0 \approx -\alpha$. The accuracy of the last two estimates is determined by the ratio of the second term to the first term. When the rejected terms in (29) are comparable with the second term (this condition can be monitored by reproducibility of the α' value), all four methods are on the same footing. In practice, it is always important to monitor the change in the sign of U_∞ because this point reliably indicates the minimum in χ^2 corresponding to $b_0 = -\alpha$ (the numbering of minima may change because of their disappearance, appearance of spurious minima, etc. (see below).

There are three possible estimates of W_∞, which use either (i) the U_∞ value at the first minimum of χ^2 or (ii, iii) the slope of a linear portion of the $U_\infty(b_0)$ curve in the vicinity of the root for the treatment at a fixed α (variation of the latter parameter within the interval of α uncertainty obtained by the four methods indicated above provides the upper and lower estimates for W_∞, respectively).

As can be readily shown, a difference between various estimates of α and W_∞ is on the same order of magnitude as the deviation of each estimate from the exact value. This correlation can be used for estimating errors. The availability of several estimates is of great significance: while any two estimated values can accidentally be close to each other (leading to understated value of the predicted error), the accidental proximity of three or four estimates is hardly probable.

4. TEST EXAMPLES

The operation of the proposed algorithm can be illustrated by application to several test systems.

4.1. Zero-Dimensional Case

The first example is offered by the integral

$$W(g) = \int_0^\infty d\varphi \varphi^{n-1} \exp(-\varphi^2 - g\varphi^4), \tag{32}$$

which can be considered as a zero-dimensional limit of the functional integral in the n-component φ^4 theory. Here, it is easy to calculate the expansion coefficients

$$W_N = ca^N \frac{\Gamma\left(N + \frac{n+2}{4}\right)\Gamma\left(N + \frac{n}{4}\right)}{\Gamma(N+1)} \tag{33}$$

and their behavior for large N:

$$W_N = ca^N \Gamma(N+b)\left\{1 + \frac{A_1}{N} + \ldots\right\}, \tag{34}$$

where

$$a = 4, \quad b = \frac{n-1}{2}, \qquad c = \frac{2^{n/2}}{4\sqrt{\pi}}, \quad A_1 = \frac{(n-2)(4-n)}{16}. \tag{35}$$

Asymptotic behavior of the integral at $g \longrightarrow \infty$ is described by the following relationships:

$$W(g) = W_\infty g^\alpha, \quad \alpha = -n/4, \qquad W_\infty = \Gamma(n/4)/4, \tag{36}$$

with the corrections having the form of a series in powers of $g^{-1/2}$. In the test, the required number of coefficients W_N was set with a double computer accuracy ($\delta \sim 10^{-14}$), after which the α and W_∞ values were restored assuming their Lipatov asymptotics to be known.

(i) $n = 1$. Figure 4 shows the $\tilde{U}_N$ against N curves calculated for various values of the parameter b_0 (points) and the results of treatment according to the power law (solid curves). For better illustration, the data are presented in the form of coefficients,

$$\tilde{U}_N = U_N \Gamma(b_0 + N_0), \tag{37}$$

normalized so as to tend to a finite limit for $b_0 \longrightarrow \infty$; N_0 is the lower limit of summation in relationship (19), which can differ from unity when several first terms of the series (1) are zero. As is seen, all curves in fact exhibit a power asymptotic behavior for large N. Attaining the asymptote is delayed for $b_0 \gg 1$ and $b_0 \longrightarrow -N_0$, because of the existence of the corresponding large parameters in relationship (19). In contrast, the power law holds even for small N for $b_0 = 0.82$ corresponding to the first minimum of χ^2.

Figure 5 shows the plots of χ^2, α_{eff}, and $\tilde{U}_\infty = U_\infty \Gamma(b_0 + N_0)$ versus b_0 calculated in the interval $24 \le N \le 50$. For the first minimum of χ^2 corresponding to $b_0 = 0.82$, estimates obtained according to Section 3 are as follows:

$$\alpha = -0.247, \quad W_\infty = 0.892, \quad \alpha' = -0.82. \tag{38}$$

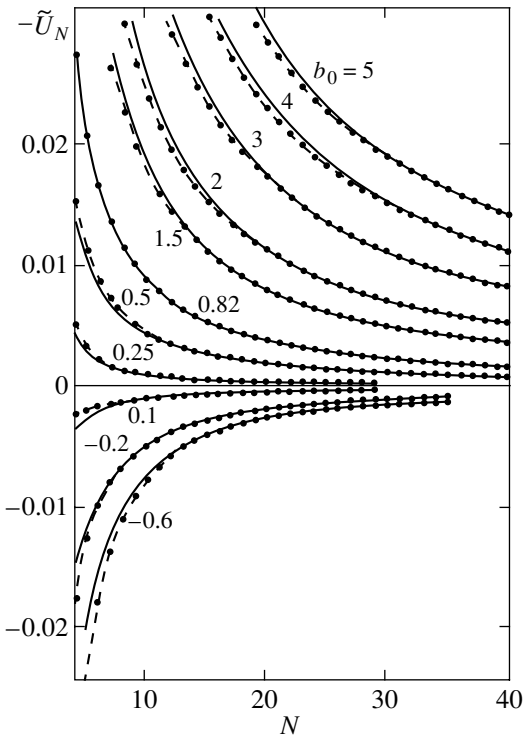

Fig. 4. The plots of $\tilde{U}_N = U_N\Gamma(b_0 + 1)$ versus N calculated for various b_0 (points and dashed curves). Solid curves show the results of treatment according to the power law using integral (32) with $n = 1$.

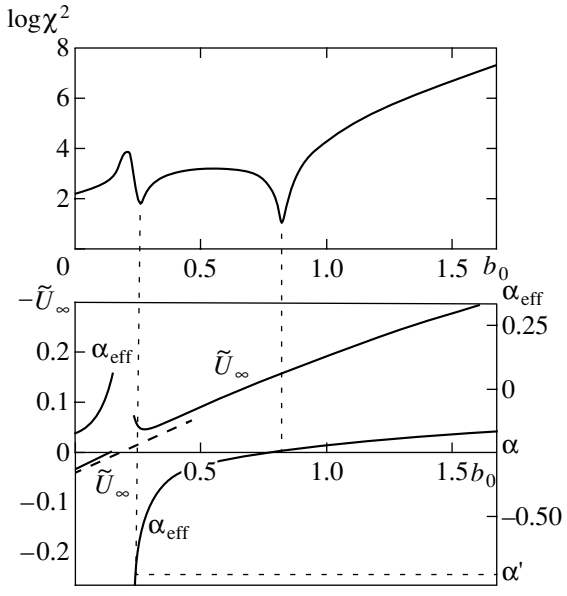

Fig. 5. The plots of χ^2, α_{eff}, and $\tilde{U}_\infty = U_\infty\Gamma(b_0 + 1)$ versus b_0 for integral (32) with $n = 1$ in the averaging interval of $24 \le N \le 50$. Dashed line shows a portion of the $U_\infty(b_0)$ curve in the vicinity of the root, obtained by the treatment at a constant index $\alpha = -0.25$.

The second minimum of χ^2 taking place at $b_0 = 0.26$ yields

$$\alpha = -0.26, \quad \alpha' = -0.67. \tag{39}$$

The U_∞ value changes sign at $b_0 = 0.210$ and 0.215 for the treatment with taking a logarithm and at a fixed index, which yields the estimates $\alpha = -0.210$ and -0.215, respectively. The slope of a linear portion in the $U_\infty(b_0)$ curve in the vicinity of the root (dashed line in Fig. 5 constructed upon treatment at a fixed index) yields the W_∞ values depending on the preselected α value: for $\alpha = -(0.21–0.26)$, the estimates range within $W_\infty = 0.883–0.933$. Summarizing all these estimates, we obtain the set of estimates

$$\alpha = -0.235 \pm 0.025, \quad W_\infty = 0.908 \pm 0.025, \quad \alpha' = 0.75 \pm 0.08, \tag{40}$$

which are consistent with the exact values

$$\alpha = -0.25, \quad W_\infty = 0.9064, \quad \alpha' = -0.75. \tag{41}$$

Since the α' values in (38) and (39) agree satisfactorily, we may conclude that the rejected terms in expansion (29) are small as compared to the second term. Therefore, the best estimates for α are provided (see Section 3) by relationships (38) and (39). Restricting to these estimates, we obtain

$$\alpha = -0.253 \pm 0.007, \quad W_\infty = 0.887 \pm 0.005 \tag{42}$$

instead of set (40). Here, the accuracy of determining α really increased, but the error of W_∞ is somewhat underestimated.

The shape of the χ^2 curves is highly sensitive to selection of the lower boundary of the working interval $N_{\min} \le N \le N_{\max}$. As the $N_{\min}$ value decreases, the χ^2 minima tend to smear, while an increase in $N_{\min}$ leads to flattening of the curves and the appearance of small-scale fluctuations hindering identification of the minima. In attempts at obtaining the clearest minima corresponding to χ^2 values of the correct order in magnitude, the choice was usually made between two–three $N_{\min}$ values.[5] A change in the working interval most significantly affects the estimates (39), with the α and α' variations approximately corresponding to a difference between (38) and (39).

(ii) $n = 2$. The χ^2 plots in Fig. 6 exhibit sharp minima at $b_0 = 1.26$ and 0.50. The first χ^2 minimum yields

$$\alpha = -0.4996, \quad W_\infty = 0.442, \quad \alpha' = -1.26, \tag{43}$$

while the other three methods give $\alpha = -0.5000$ accurate to within the last digit. An estimate for α' obtained using the second χ^2 minimum amounts to about 20,

[5] It should be noted that, in displaying the results of calculations with fixed decimal point, the χ^2 minima are well distinguished by the configuration of digits even in the course of a rapid on-screen computer survey.

which is inconsistent with (43). Therefore, the rejected terms in (29) are comparable with the second, so that all four possible estimates are on the same footing. Treatment of a linear portion of the $U_\infty(b_0)$ curve near the root yields $W_\infty = 0.460$. As a result, we obtain

$$\alpha = -0.5000 \pm 0.0004, \quad W_\infty = 0.451 \pm 0.009, \quad (44)$$

in good agreement with the exact values

$$\alpha = -0.50, \quad W_\infty = 0.4431. \quad (45)$$

(iii) $n = 3$. Here, the $\chi^2(b_0)$ plots exhibit minima at $b_0 = 1.07$ and 0.77, which yield

$$\alpha = -0.704, \quad W_\infty = 0.192, \quad \alpha' = 1.07 \quad (46)$$

and

$$\alpha = -0.77, \quad \alpha' = -1.42, \quad (47)$$

respectively. Estimates obtained using U_∞ changing sign are $\alpha = -0.86$ for the treatment with taking a logarithm and $\alpha = -0.84$ for the treatment at a fixed index. Determining W_∞ from the slope of a linear portion in the $U_\infty(b_0)$ curve in the vicinity of the root yields 0.311, 0.420, and 0.751 for $\alpha = -0.704$, –0.77, and –0.86, respectively. Since the two values of α' reasonably agree with each other, the estimates (46) and (47) for α must be more precise. Taking only these estimates into account, we obtain

$$\alpha = -0.737 \pm 0.033, \quad W_\infty = 0.306 \pm 0.114, \quad \alpha' = -1.25 \pm 0.18, \quad (48)$$

in good agreement with the exact values

$$\alpha = -0.75, \quad W_\infty = 0.3063, \quad \alpha' = -1.25. \quad (49)$$

An allowance for all four estimates of α yields

$$\alpha = -0.78 \pm 0.08, \quad W_\infty = 0.47 \pm 0.28 \quad (50)$$

with markedly greater errors.

In this case, we may also point out difficulties arising due to an additional "spurious" minimum appearing at $b_0 = 1.90$. However, this minimum can be excluded from consideration upon identifying the minimum at $b_0 = 0.77$ as corresponding to $b_0 = -\alpha$ (by U_∞ changing sign) and the minimum at $b_0 = 1.07$ as corresponding to $b_0 = -\alpha'$ (by the consistent α' values). In the general case, the process of identifying useful minima resembles the situation in spectroscopy under high noise conditions: selecting informative signals requires certain skill.

(iv) $n = 4$. In this case, application of the algorithm encounters the "hidden rock" of this method. Based on the usual estimates, we obtain a quite precise result:

$$\alpha = -1.500 \pm 0.004, \quad W_\infty = -0.222 \pm 0.005. \quad (51)$$

However, these values do not agree with (36). The discrepancy is caused by the fact that the main contribution to the U_N asymptotics vanish because the gamma function exhibits a pole at the exact value of the index $\alpha = -1$ (see Eq. (29)), so that the next term of the expansion becomes significant with the parameters

$$\alpha' = -1.50, \quad W'_\infty = -\sqrt{\pi}/8 = -0.2216. \quad (52)$$

Therefore, the proposed algorithm is incapable of restoring correct asymptotics described by Eq. (7) in the case of nonpositive integer α values. In order to avoid these problems, the algorithm has to be supplemented by the following rule: if the treatment yields a negative α value, the result must be checked by taking a negative or fractional power of series (1) and summing the reexpanded series.

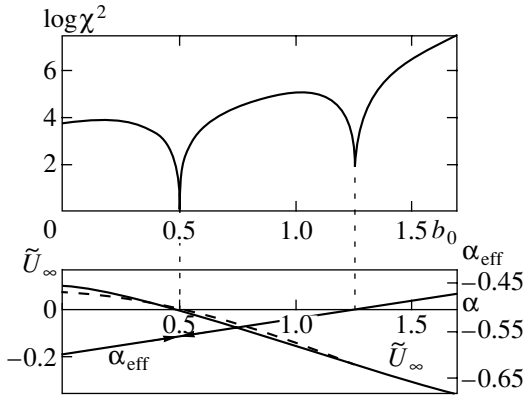

Fig. 6. The plots of χ^2, α_{eff}, and $\tilde{U}_\infty = U_\infty\Gamma(b_0 + 1)$ versus b_0 for integral (32) calculated with $n = 2$ in the averaging interval of $20 \le N \le 50$. Dashed line shows a portion of the $U_\infty(b_0)$ curve in the vicinity of the root, obtained by the treatment at a constant index $\alpha = -0.5$. The α_{eff} for $b_0 = 0.5$ falls far outside the diagram boundaries.

4.2. Anharmonic Oscillator

The second example is offered by the problem of determining the ground sate $E_0(g)$ of an anharmonic oscillator described by the Schrödinger equation

$$\left\{-\frac{d^2}{dx^2} + \frac{x^2}{4} + \frac{gx^4}{4}\right\}\psi(x) = E\psi(x). \quad (53)$$

This problem can be reduced to a one-dimensional φ^4 theory. Consider $E_0(g)$ as the $W(g)$ function with the initial terms of the perturbative series having the following form:

$$W(g) = \frac{1}{2} + \frac{3}{4}g - \frac{21}{8}g^2 + \frac{333}{16}g^3 - \frac{30885}{128}g^4 + \ldots. \quad (54)$$

Bender and Wu [23] calculated the first 75 coefficients W_N up to the 12th decimal digit and obtained an expres-

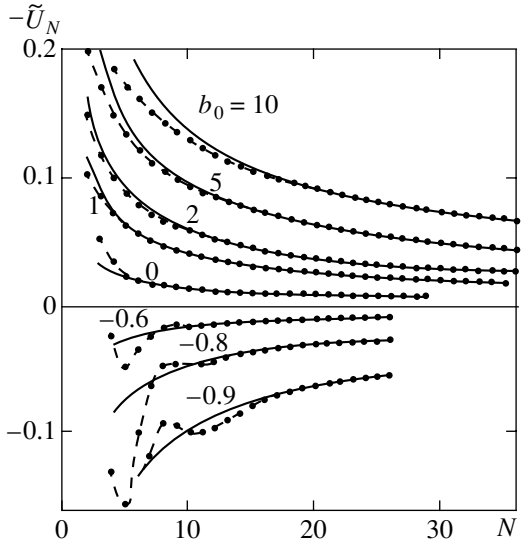

Fig. 7. The plots of $\tilde{U}_N = U_N\Gamma(b_0 + 1)$ versus N for an anharmonic oscillator. The notations are the same as in Fig. 4.

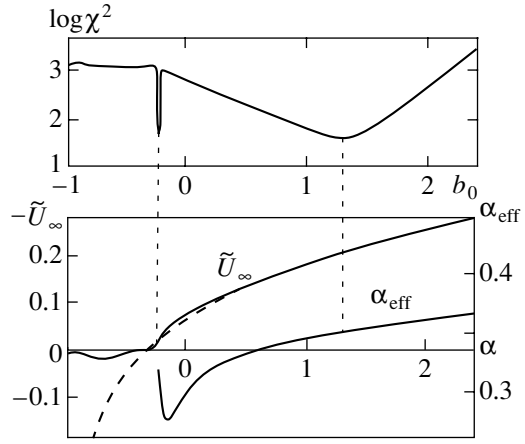

Fig. 8. The plots of χ^2, α_{eff}, and $\tilde{U}_\infty = U_\infty\Gamma(b_0 + 1)$ versus b_0 for an anharmonic oscillator in the averaging interval of $24 \le N \le 45$. Dashed line shows the result of treatment at a constant index $\alpha = 0.34$.

sion describing behavior of the expansion coefficients with large N:

$$W_N = -\frac{\sqrt{6}}{\pi^{3/2}}3^N\Gamma\left(N+\frac{1}{2}\right)\left\{1-\frac{95/72}{N}+\ldots\right\}. \quad (55)$$

The asymptotics of $E_0(g)$ for $g \longrightarrow \infty$ is revealed by substituting $E_0(g) = \lambda_0 g^{1/3}$and $x \longrightarrow xg^{-1/6}$, after which Eq. (53) transforms into

$$\left\{-\frac{d^2}{dx^2}+\frac{x^2}{4}+\frac{x^2}{4g^{2/3}}\right\}\psi(x) = \lambda_0\psi(x). \quad (56)$$

For $g \longrightarrow \infty$, the last term in braces is insignificant and λ_0 tends to a constant value of 0.6679863 that can be determined by the variational method [24]. Thus, the $W(g)$ asymptotics is described by power series (7) with the parameters

$$\alpha = 1/3, \quad W_\infty = 0.668, \quad (57)$$

and the corrections having the form of a series in powers of $g^{-2/3}$.

Figure 7 presents the plots of $\tilde{U}_N$ against N and the results of their treatment according to the power law. Figure 8 shows the plots of χ^2, α_{eff}, and $\tilde{U}_\infty$ versus b_0. As is seen, χ^2 exhibits minima $b_0 = 1.30$ and –0.34 corresponding to

$$\alpha = 0.349, \quad W_\infty = 0.602, \quad \alpha' = -1.80 \quad (58)$$

and

$$\alpha = 0.34, \quad \alpha' \approx 20, \quad (59)$$

respectively. Estimates obtained using U_∞ on the same footing changing sign are $\alpha = 0.285$ for the treatment with taking logarithm and $\alpha = 0.337$ for the treatment at a fixed index. Determining W_∞ from the slope of a linear portion in the $U_\infty(b_0)$ curve in the vicinity of the root yields values in the interval from 0.616 to 0.883. The two values of α' having nothing in common indicates that all α estimates are on the same footing. As a result, we obtain

$$\alpha = 0.317 \pm 0.032, \quad W_\infty = 0.74 \pm 0.14, \quad (60)$$

in good agreement with the exact values (57).

The above examples show that the accuracy of restoring the $W(g)$ asymptotics, while depending significantly on the particular problem, is generally correlated with the character of corrections to the U_N asymptotics described by relationship (20). An average accuracy on the order of 10^{-2} is attained in the zero-dimensional case with odd n, where the corrections to (20) have the form of power series in $N^{-1/2}$. For even n, every other correction vanishes due to the poles of the gamma function to leave a regular expansion in $1/N$, which markedly increases the resulting accuracy. A relatively low accuracy in the case of an anharmonic oscillator is related to the fact that corrections have the form of series in powers of $N^{-1/3}$.[6] It is important to note, however, that the algorithm automatically yields an estimate of the error. The estimate is rather reliable when all four possible methods for evaluating α are employed.

[6] The first term in (28) gives, in addition to the main contribution to U_N proportional to $N^{\alpha-1}$, the regular corrections $N^{\alpha-2}$, $N^{\alpha-3}$, …; the second term contributes by $N^{\alpha'-1}$, $N^{\alpha'-2}$, …, etc. As a result, the expansion in $g^{-2/3}$ converts into the expansion in $N^{-1/3}$.

5. ALGORITHM OPERATING WITH INTERPOLATED COEFFICIENT FUNCTION

The importance of interpolation was strongly underestimated, although this method can obviously provide for an increase in the accuracy of calculations. In most investigations in the field under consideration, the algorithms were formulated so as to avoid mentioning the coefficients W_N at intermediate N values. This approach is conceptually incorrect since, using a finite number of the initial coefficients and their asymptotics, it is possible to construct a function with preset behavior in infinity.[7] A reasonable problem formulation corresponds to approximately setting all W_N, after which $W(g)$ can be reconstructed with certain precision.

Thus, a necessary stage in solving the problem consists in interpolating the coefficient function, which naturally implies that this function is analytical (see Section 8.2). The interpolation stage allows the parameter c in the Lipatov asymptotics (essentially not used in the standard conform-Borel procedure [6]) to be effectively employed. In addition, it is possible to take into account smoothness of the reduced coefficient function, its regularity with respect to $1/N$, and (eventually) the information concerning asymptotics of the A_K coefficients in expansion (22) [25].

In Section 2.3, some qualitative considerations were presented suggesting that the influence of the interpolation errors is not as significant as that of the round-off errors. Unfortunately, no particular estimates illustrating this were obtained. Validity of this statement will be experimentally demonstrated for the zero-dimensional test example with $n = 1$.

With a view to modeling a situation for the φ^4 theory, let us assume that several coefficients in the expansion of series (1) are known,

$$W_{L_0}, W_{L_0+1}, \ldots, W_L, \tag{61}$$

together with the Lipatov asymptotics (2) and the corresponding first corrections in $1/N$. The interpolation is conveniently performed for the reduced coefficient function, retaining a finite number of terms in expansion (22) and selecting coefficients A_K by correspondence to set (61).

Let us consider in detail two examples of the interpolation procedure, which correspond to (i) $L_0 = 1$, $L = 5$ and (ii) $L_0 = 1$, $L = 1$. Owing to a slow character of variation of the coefficient function, the accuracy of interpolation in both cases is very high: $\sim 10^{-9}$ and $\sim 10^{-4}$, respectively. A random error of such amplitude might only lead to large fluctuations in U_N for $N \approx 30$ in the former case and $N \approx 13$ in the latter case. Real calculations indicate that no catastrophic consequences take place up to $N = 40$, when the influence of the round-off errors becomes significant. This can be seen in Table 1 presenting the values of some coefficients U_N calculated for $b_0 = 1$ using the exact and interpolated coefficients of W_N. An increase in the b_0 value improves the accuracy; when b_0 decreases, the accuracy drops somewhat, although the resulting deviations would be indistinguishable on the scale of Fig. 4.

Table 1. Comparison of U_N values calculated for $b_0 = 1$ using exact and interpolated coefficients W_N

N	U_N		
	Exact W_N values	Interpolation with $L_0 = 1$, $L = 5$	Interpolation with $L_0 = 1$, $L = 1$
30	-2.911×10^{-3}	-2.911×10^{-3}	-2.868×10^{-3}
35	-2.408×10^{-3}	-2.409×10^{-3}	-2.369×10^{-3}
40	-2.038×10^{-3}	-2.041×10^{-3}	-2.004×10^{-3}

Table 2. The parameters of asymptotics for integral (32) with $n = 1$ calculated using exact and interpolated coefficients W_N

Estimates based on	Exact W_N values	Interpolation with $L_0 = 1$, $L = 5$	Interpolation with $L_0 = 1$, $L = 1$
First χ^2 minimum	$\alpha = -0.246$	$\alpha = -0.245$	$\alpha = -0.269$
	$\alpha' = -0.827$	$\alpha' = -0.830$	$\alpha' = -0.761$
	$W_\infty = 0.893$	$W_\infty = 0.892$	$W_\infty = 0.912$
Second χ^2 minimum	$\alpha = -0.249$	$\alpha = -0.245$	$\alpha = -0.271$
	$\alpha' = -0.792$	$\alpha' = -0.849$	$\alpha' = -0.747$
U_∞ changing sign	$\alpha = -0.210$	$\alpha = -0.210$	$\alpha = -0.218$
$U_\infty(b_0)$ slope	$\alpha = -0.215$	$\alpha = -0.215$	$\alpha = -0.225$
	$W_\infty = 0.889$	$W_\infty = 0.887$	$W_\infty = 0.885$

The curve of $\chi^2(b_0)$ is analogous to (albeit not fully coinciding with) that depicted in Fig. 5. Estimates of the asymptotic parameters are listed in Table 2; for better illustration, all values refer to the same working interval of $23 \le N \le 45$ and the value $\alpha = -0.25$ used for the treatment of a linear portion of the $U_\infty(b_0)$ curve. As is seen from these data, changes in α and W_∞ caused by the interpolation fall within the scatter of various estimates and virtually do not influence the accuracy of restoration of asymptotics (7). Therefore, interpolation using a single expansion coefficient W_1 allowed the $W(g)$ asymptotics to be restored with an accuracy not worse than that achieved with the exact coefficients W_N. Of course, this is by no means a typical situation.

[7] A function of the factorial series possesses the same asymptotics of coefficients (2) but with a different parameter c [17]; the last statement in the text can be readily proved by taking an appropriate linear combination of several functions.

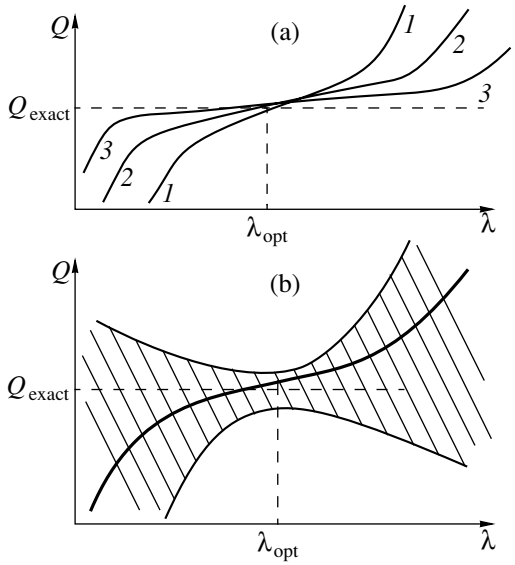

Fig. 9. Schematic diagrams illustrating the optimization procedure: (a) theoretically, any quantity Q obtained upon summation of the series must be independent of the optimization parameter λ; however, such dependence arises under the conditions of restricted information and weakens (on the passage from curve *1* to *2*, *3*, etc.) as the amount of information increases (the optimum value $\lambda = \lambda_{opt}$ occurs at the center of the plateau); (b) the choice of λ affects both the approximate Q value (thick solid curve) and the error of determination (cross-hatched area), so that a correct estimation of this error must provide for the exact value Q_{exact} being compatible with all data. In the "ideal" situation depicted, optimization with respect to λ consists in selecting the result characterized by a minimum error.

6. OPTIMIZATION OF THE INTERPOLATION PROCEDURE

Considering an example in the preceding section, we were lucky to see that the most natural method of interpolation may give good results. In the general case, the interpolation procedure requires optimization that will be demonstrated in the case of an anharmonic oscillator. Let us first discuss the general strategy of optimization, which has been significantly modified in comparison to that used in the previous works.

6.1. General Strategy of Optimization

On an abstract level, the optimization consists in introducing a certain variation of the summation procedure characterized by a parameter λ, the latter value being eventually selected in a "optimum manner." For example, the initial series (1) can be raised to the λ power and reexpanded to yield

$$W^{\lambda}(g) = \tilde{W}_0 - \tilde{W}_1 g + \tilde{W}_2 g^2 - \ldots + \tilde{c}a^N \Gamma(N+b)(-g)^N + \ldots. \quad (62)$$

The properties of this series are analogous to those of the initial one, except for a change in the Lipatov asymptotic parameter c [17]. The new series is summed upon selecting the λ value so as to provide for the best convergence of the second series in expansion (9). The optimization procedure is employed, bringing both advantages and troubles, in most investigations in the field under consideration. On the one hand, the principal possibility of improving the convergence is definitely valuable. On the other hand, the results become dependent of an arbitrary parameter λ and it is difficult to get rid of the feeling that any result can be obtained.

Theoretically, the use of series (62) is fully equivalent to the study of initial series (1) and the value of any quantity Q obtained upon summation must be independent of the parameter λ. However, under the conditions of restricted information concerning coefficients W_N, the Q value begins to depend on the choice of λ, this dependence weakening as the amount of information increases. In the general case, no uniform convergence with respect to λ takes place and an approximate Q value is close to the exact one only within a certain "plateau" region (Fig. 9a), the deviations rapidly growing outside this region. As the amount of necessary information increases, the plateau expands and flattens (see, e.g., [26]). Apparently, the best convergence takes place at the center of the plateau. However, this point is not always unambiguously selected, since the plateau may be asymmetric or poorly pronounced, the center may shift in the course of convergence, etc. Therefore, selecting the best approximation for Q and estimating the approximation uncertainty are rather subjective procedures.

In the author's opinion, the optimization problem can nevertheless be solved objectively. Indeed, since the choice of λ affects both the approximate Q value and the error of determination, a correct estimation of this error must provide for the exact value Q_{exact} being compatible with the approximate values obtained for any λ (Fig. 9b). This criterion eliminates the problem of an apparent dependence of Q on λ. Once such an "ideal" situation is attained, optimization of the procedure with respect to λ reduces to selecting the result characterized by a minimum error.

The optimization procedure is expediently performed in the interpolation stage, since all the final errors arise essentially from the uncertainties in W_N. Rewriting expansion (22) in the equivalent form

$$W_N = ca^N N^{\tilde{b}} \Gamma(N+b-\tilde{b}) \times \left\{1 + \frac{\tilde{A}_1}{N-\tilde{N}} + \frac{\tilde{A}_2}{(N-\tilde{N})^2} + \ldots + \frac{\tilde{A}_K}{(N-\tilde{N})^K} + \ldots\right\} \quad (63)$$

and using the interpolation by truncating the series and selecting coefficients $\tilde{A}_K$, we obtain a manifold of realizations of the interpolation procedure characterized by two parameters, $\tilde{b}$ and $\tilde{N}$. An analysis of the test examples shows this parametrization to be sufficiently effective: the accuracy of interpolation achieved for the opti-

mum $\tilde{b}$ and $\tilde{N}$ values can be higher by several orders of magnitude as compared to that for a random choice of these parameters. Below, the optimization with respect to $\tilde{b}$ is based on theoretical consideration, while the optimum $\tilde{N}$ value is selected based on the results of numerical calculations.[8]

6.2. Optimization with Respect to $\tilde{b}$

The process of optimization with respect to $\tilde{b}$ encounters the problem of selecting parametrization for the Lipatov asymptotics which can be written in various forms: $ca^N\Gamma(N+b)$, $ca^N N^{b-1}N!$, etc. This problem was actively discussed (see, e.g., [11, 12]), but no satisfactory solutions were proposed.

Note that the values $\tilde{b} = b$ and $\tilde{b} = b - 1$ lead to identical results:

$$N^{\tilde{b}}\Gamma(N+b-\tilde{b}) = \begin{cases} N^b\Gamma(N), & \tilde{b} = b \\ N^{b-1}\Gamma(N+1) = N^b\Gamma(N), & \tilde{b} = b-1. \end{cases} \quad (64)$$

Therefore, the approximate values of any quantity Q obtained upon summation of the series will coincide for $\tilde{b} = b$ and $b - 1$. As the amount of information concerning the coefficients W_N increases, the $Q(\tilde{b})$ function varies more and more slowly. When the characteristic scale L of this variation increases, the kth derivative of the function drops as $1/L^k$. As a result, an extremum at the point $\tilde{b} = b - 1/2$ appears in the general situation, with a plateau between the Q values corresponding to $\tilde{b} = b$ and $b - 1$and the point $\tilde{b} = b - 1/2$ being the natural center of this plateau. The error of restoring Q, like any other value, exhibits an extremum (which is naturally expected to be minimum) at $\tilde{b} = b - 1/2$ (see Section 8). Thus, the optimum choice is $\tilde{b} = b - 1/2$; this corresponds to the following parametrization of the Lipatov asymptotics:

$$W_N^{as} = ca^N N^{b-1/2}\Gamma(N+1/2). \quad (65)$$

The first correction A_1/N to this asymptotics (see expansion (22)) depends on $\tilde{b}$ as

$$A_1 = \bar{A}_1 - (b - 1/2 - \tilde{b})^2/2, \quad (66)$$

where $\bar{A}_1$ is the value of A_1 for $\tilde{b} = b - 1/2$. In all known cases, $\bar{A}_1 < 0$ (see [19, 23, 27, 28]) and a minimum correction corresponds to parametrization (65), which favors a good matching between the high-order asymptotics and the low-order behavior. Note that the asymptote according to the Lipatov method [2] is

$$\sqrt{2\pi}c(a/e)^N N^{b-1/2}N^N.$$

The above parametrization (65) corresponds to approximation

$$\sqrt{2\pi}e^{-N}N^N \approx \Gamma(N+1/2)$$

and provides for an accuracy of 4% even for $N = 1$ (thus, being a "natural" choice). For an anharmonic oscillator, the optimum parametrization coincides with (55), while in the zero-dimensional case with $n = 1$ the expression is similar to (34) and (35).

6.3. Optimization with Respect to $\tilde{N}$

The case of an anharmonic oscillator was studied in detail using the interpolation with $L_0 = 1$, $L = 9$ (i.e., using the first nine W_N coefficients), which corresponded to an accuracy of ~10^{-3}. The interpolation based on expression (22) was unsatisfactory: the χ^2 values obtained by treatment according to the power law (20) were abnormally large even for reasonable averaging intervals and gave no clear pattern with minima. The reason for this behavior is revealed by comparison of the U_N coefficients (obtained by interpolation) to the exact values. As is seen from Fig. 10a, the difference is very large, making treatment by the power law practically impossible. Deviations increase by approximately the same law as those for the random errors, but the variation is rather smooth and is analogous for different b_0 values. It appears that these deviations can be compensated in a broad range of b_0 by optimization with respect to $\tilde{N}$.

This is really so and the region of optimum $\tilde{N}$ values can be determined without knowledge of the exact result. Figure 11 shows the behavior of χ^2 in the interval of $20 \le N \le 40$ depending on $\tilde{N}$ for integer b_0 values. As is seen, small χ^2 values are immediately obtained for $b_0 = 0, 1, 2, 3$ in the interval of $\tilde{N} = -(5.0–5.5)$. This is evidence that the error of U_N can be compensated for all $b_0 \ge 0$, since greater b_0 correspond to still smaller errors (see Section 2.3). As is seen from Fig. 10b, deviations of the resulting U_N for $\tilde{N} = -5.4$ from exact values for $b_0 \ge 0$ are in fact virtually indistinguishable.

The possibility of more refined optimization is based on the fact that the interpolation errors in formula (29) play the same role as do the high-order scaling corrections indicated by dots. As $\tilde{N}$ is changed, the interpolation errors smoothly vary and (for a certain $\tilde{N}$ value)

[8] Further increase in the number of optimization parameters seems to be inexpedient: this way may lead to absurd results. In particular, a large number of parameters allows imitation of a rapid convergence of the algorithm to an erroneous result. Even using the proposed approach, it is possible to ensure coincidence of four estimates of the α value at a zero error.

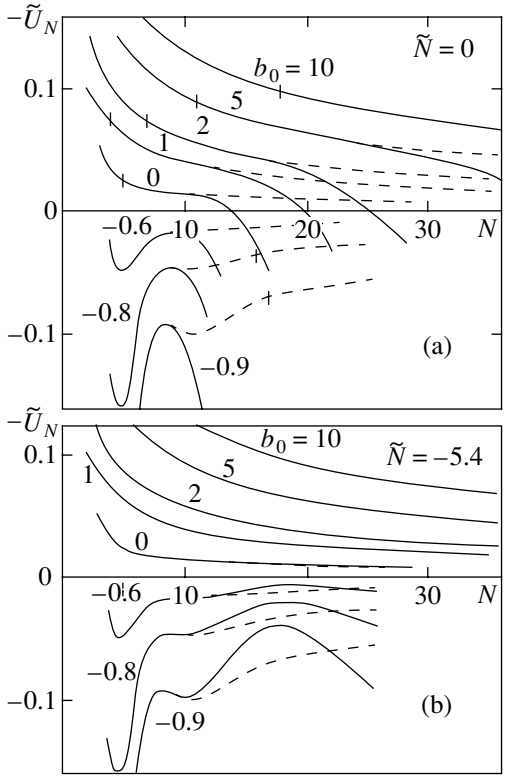

Fig. 10. Optimization of the interpolation procedure for an anharmonic oscillator: (a) a comparison of the U_N values obtained by interpolation for $\tilde{N} = 0$ using the first nine W_N coefficients (solid curves) to exact values (dashed curves); vertical bars indicate the N values above which behavior of the exact U_N values is visually indistinguishable from that according to the power law; (b) an analogous pattern after the optimization with respect to $\tilde{N}$ (for $\tilde{N} = -5.4$).

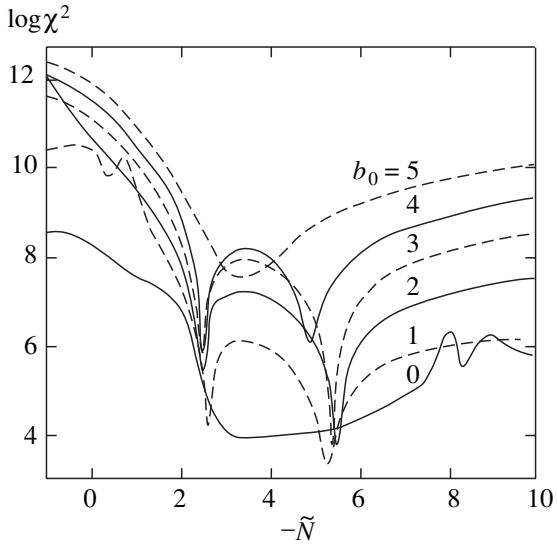

Fig. 11. The plots of χ^2 versus $\tilde{N}$ for an anharmonic oscillator in the interval of $20 \le N \le 40$ at various fixed b_0 values.

become approximately compensated by the scaling corrections. This point can be detected by the maximum proximity of various estimates obtained for the α and W_∞ values.

A systematic treatment with determination of the α and W_∞ values was carried out for $\tilde{N}$ in the interval from –5.0 to –5.6 at a step of 0.1. A "correct" pattern of χ^2 minima was observed for $\tilde{N} = -5.5$, while for $\tilde{N} = -5.6$ the first minimum disappeared and for $\tilde{N} \ge -5.4$ it was split in two. The reason for this splitting is qualitatively evident: Figs. 10 and 11 show that, at a fixed $\tilde{N}$, there is a certain b_0 value for which the effect of the interpolation error upon U_N is virtually compensated. This very b_0 corresponds to an "extra" minimum of χ^2 in comparison with the pattern of Fig. 8. Since it is difficult to decide *a priori* which of the two minima is true, the estimates were obtained for both (and proved to be very close to each other).

The results of these numerical calculations are summarized in Table 3 and depicted in Fig. 12. The scatter of α and W_∞ values allows the error to be evaluated by the order of magnitude. In order to obtain an "ideal" pattern according to Fig. 9b, the error interval should be expanded by a factor of 1.3 and 1.1 for α and W_∞, respectively (dotted curves in Fig. 12). Then the values of $\alpha = 0.38$ and $W_\infty = 0.52$ (dashed curves in Fig. 12) are compatible with the results for all $\tilde{N}$. Selecting the $\tilde{N}$ values in each particular case so as to minimize the one-side error (as indicated by arrows in Fig. 12), we obtain the following estimates:

$$\alpha = 0.38 \pm 0.05, \quad W_\infty = 0.52 \pm 0.12. \qquad (67)$$

A comparison to the set (57) shows that the error is estimated adequately, while the average values are somewhat displaced; the shift in W_∞ is induced by the shift in α.

7. SUMMING PERTURBATIVE SERIES FOR AN ARBITRARY g

When the amount of information concerning the W_N coefficients suffices for restoring the $W(g)$ asymptotics as $g \longrightarrow \infty$, summing series (1) for an arbitrary g encounters no problems: the coefficients U_N for $N \lesssim 40$ are calculated by formula (19) and the subsequent terms can be obtained according to the $U_\infty N^{\alpha-1}$ asymptotics, so that all coefficients of the converging series (9) are known. The summation error is determined by the accuracy of restoring the asymptotics,

$$\Delta_{as} = \left.\frac{\delta U_N}{U_N}\right|_{N \gg 1} = \frac{\delta U_\infty}{U_\infty} + \delta\alpha \ln N, \qquad (68)$$

which varies logarithmically with N and can be considered as constant with a restricted interval. Introducing a

Table 3. Asymptotic parameters for an anharmonic oscillator obtained by the interpolation with $L_0 = 1$, $L = 9$ (the values in parentheses for $\tilde{N} = -5.6$ were estimated at the point $b_0 = 2.20$ where the first χ^2 minimum disappears)

Estimates based on	α for $\tilde{N}$						
	–5.0	–5.1	–5.2	–5.3	–5.4	–5.5	–5.6
First χ^2 minimum	0.398	0.396	0.393	0.390	0.385	0.378	(0.373)
	0.476	0.452	0.422	0.399	0.384		
Second χ^2 minimum	0.50	0.47	0.42	0.37	0.33	0.29	0.34
U_∞ changing sign	0.585	0.535	0.485	0.445	0.405	0.365	0.335
$U_\infty(b_0)$ slope	0.495	0.445	0.40	0.36	0.32	0.29	0.26
	W_∞ for $\tilde{N}$						
	–5.0	–5.1	–5.2	–5.3	–5.4	–5.5	–5.6
First χ^2 minimum	0.490	0.495	0.500	0.505	0.513	0.529	(0.540)
	0.356	0.390	0.440	0.487	0.517		
$U_\infty(b_0)$ slope	0.226	0.290	0.373	0.463	0.572	0.675	0.712
	0.502	0.538	0.568	0.698	0.885	1.09	0.953

characteristic scale N_c on which the relative error is comparable with Δ_{as} and using the approximation

$$\frac{\delta U_N}{U_N} = \begin{cases} 0, & N < N_c \\ \Delta_{as}, & N \geq N_c, \end{cases} \tag{69}$$

we obtain for $ag \gg 1$

$$\delta B(g) = \sum_{N=N_c}^{\infty} \Delta_{as} U_N \exp\left(-\frac{N_c}{ag}\right) = \begin{cases} \Delta_{as} B(g), & ag \gg N_c \\ \Delta_{as} U_{N_c} ag \exp(-N_c/ag), & ag \ll N_c. \end{cases} \tag{70}$$

Substituting these expressions into (8) and using the steepest descent method for $ag \ll N_c$, we obtain

$$\frac{\delta W(g)}{W(g)} \sim \begin{cases} \Delta_{as}, & ag \gtrsim N_c \\ \Delta_{as} \exp\{-2(N_c/ag)^{1/2}\}, & ag \lesssim N_c \end{cases} \tag{71}$$

(where some preexponential factors are omitted for clarity). For negative α, the results for $ag \gg N_c$ are somewhat different. In particular, for $-1 < \alpha < 0$ we obtain $\delta W(g) = \Delta_{as}(W(g) - W(g_c))$, where $ag_c \sim N_c$. A natural scale for N_c is provided by the middle of the working interval $(N_{\min}, N_{\max})$, that is, $N_c \approx 30$; however, deviations from this value may be quite large because the corresponding equality holds in fact on the logarithmic

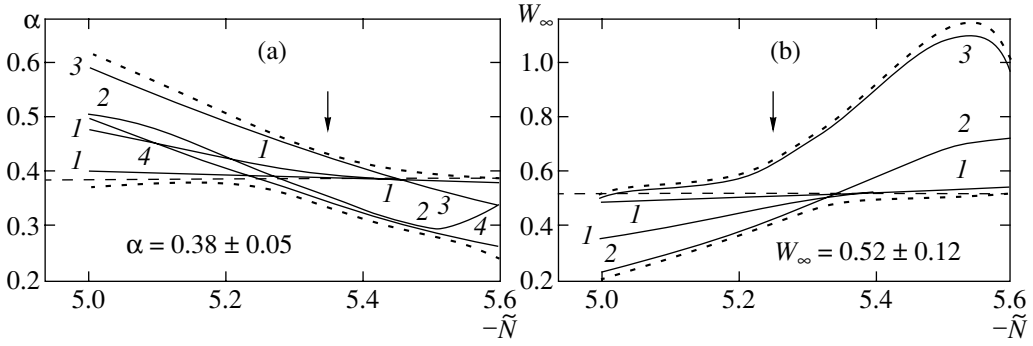

Fig. 12. The plots of α and W_∞ values estimated for an anharmonic oscillator by various methods (see Section 3): (a) α estimates based on the (*1*) first χ^2 minimum, (*2*) second χ^2 minimum, (*3*) U_∞ changing sign, and (*4*) $U_\infty(b_0)$ slope; (b) W_∞ estimates based on the (*1*) first χ^2 minimum and (*2*, *3*) $U_\infty(b_0)$ slope (upper and lower bounds, respectively). Small-dash lines indicate the error interval expanded by a factor of 1.3 and 1.1 for α and W_∞ values, respectively.

Table 4. Comparative data for the exact integral (32) with $n = 1$ and the results obtained by summing the perturbative series

g	$W(g) \times 10$			
	Exact value	Summing with exact W_N	Summing upon interpolation with $L_0 = 1$, $L = 5$	Summing upon interpolation with $L_0 = 1$, $L = 1$
1	6.842134	6.842135	6.842134	6.8436
2	6.183453	6.183454	6.183452	6.1867
4	5.497111	5.497110	5.497105	5.5034
8	4.820615	4.820608	4.820594	4.832
16	4.181699	4.181669	4.181637	4.200
32	3.597297	3.59720	3.59714	3.624
64	3.075230	3.07500	3.07490	3.113
128	2.616802	2.61633	2.61617	2.668
256	2.219222	2.2184	2.2182	2.285
512	1.877472	1.8761	1.8758	1.959
1024	1.585578	1.5835	1.5831	1.68
$g \longrightarrow \infty$	$9.064g^{-0.25}$	$8.95g^{-0.247}$	$8.95g^{-0.247}$	$9.12g^{-0.269}$

scale ($\ln N_c \approx \ln 30$). In practice, approximation (69) with a constant N_c is expedient only for large g. In the general case, estimate (71) is valid with an effective N_c value, which is determined by the number N of the maximum term $\delta U_N u^N$ in the series for $\delta B(u)$ (for small g, this value is close to $L + 1$, e.g., to the number of the first unknown coefficient W_N).

Table 5. Comparative data for the exact ground state energy $E_0(g)$ of an anharmonic oscillator and the results obtained by summing the perturbative series (the $2E_0(g)$ and $2g$ values are given in order to provide for the correspondence with the data reported in most other papers using a different normalization)

$2g$	$2E_0(g)$		
	Exact value	Summing with exact W_N ($b_0 = 1.30$)	Summing upon interpolation with $L_0 = 1$, $L = 9$ ($\tilde{N} = -5.3$, $b_0 = 3.55$)
0.5	1.241854	1.241854	1.241857
1	1.392352	1.392352	1.392396
2	1.607541	1.607545	1.60790
3	1.769589	1.769605	1.7706
4	1.903137	1.903178	1.9051
5	2.018341	2.018418	2.0214
10	2.449174	2.44961	2.4599
20	3.009945	3.0117	3.040
50	4.003993	4.0115	4.096
100	4.999418	5.018	5.19
$g \longrightarrow \infty$	$2 \times 0.668g^{1/3}$	$2 \times 0.602g^{0.349}$	$2 \times 0.511g^{0.387}$

Table 4 presents the results of calculations for the zero-dimensional case. Here, the first column gives the exact values of integral (32) with $n = 1$, while the columns from second to fourth present the results of summation obtained using exact W_N coefficients and interpolated values (with $L_0 = 1$, $L = 5$ or $L_0 = 1$, $L = 1$), respectively. In each case, the calculations were performed for b_0 corresponding to the first χ^2 minimum. A comparison to (71) indicates that $N_c \sim 200$ for the second and third columns and $N_c \sim 10$ for the fourth column.

Table 5 presents the analogous data for an anharmonic oscillator. Here, the first column gives the exact $E_0(g)$ values taken from [24], while the second and third columns present the results of summation obtained using exact W_N coefficients and interpolated values (with $L_0 = 1$, $L = 9$), respectively. In this case, the estimates give $N_c \sim 200$ for the second column and about 50 for the third column.

Information concerning the $W(g)$ asymptotics can also be taken into account within the framework of the standard conform-Borel procedure (Section 2.1) by interpolating the U_N coefficients (with the known asymptotics (17)) calculated using formula (11). For approximation (69), we obtain by analogy with (71)

$$\frac{\delta W(g)}{W(g)} \sim \begin{cases} \Delta_{as}, & ag \gtrsim N_c^2 \\ \Delta_{as} \exp\{-3(N_c^2/ag)^{1/3}\}, & ag \lesssim N_c^2. \end{cases} \tag{72}$$

This procedure is preferred in the case of sufficiently small g values (when N_c is close to $L + 1$), leading to smaller errors as compared to those obtained for (71). For greater g, the attaining of N_c values indicated above seems to be impossible.

According to the standard procedure of calculating the critical indices [6], the second series (9) is truncated on the Lth term that corresponds to the error given by (72) with $N_c = L + 1$ and $\Delta_{as} \sim 1$. In the three-dimensional case, a large number of expansion coefficients are known (for $L = 6$). These values are well matched with (2), which gives hope for restoring the asymptotics of scaling functions with an accuracy of $\Delta_{as} \sim 10^{-2}$ and for increasing N_c at the expense of interpolation. Thus, it is apparently possible to increase the accuracy of calculation of the critical indices by two–three orders of magnitude even for the currently available information. Using the modified conformal mapping may lead to a further increase in the accuracy, provided that the scale of $N_c \gtrsim 20$ would be accessible in the corresponding region of $ag \sim 0.2$.

8. THE φ^4 THEORY

8.1. Restoration of the Gell-Mann–Low Function

Now let us turn to a real physical problem of restoring the Gell-Mann–Low function in the φ^4 theory, considering $\beta(g)$ as $W(g)$ and proceeding from the information contained in relationships (5) and (6).

The interpolation was based on formula (63) with an optimum value of $\tilde{b} = 4$. Figure 13 presents the plots of $\chi^2(\tilde{N})$ versus $\tilde{N}$ calculated in the interval $20 \le N \le 40$ for several fixed b_0 values. As is seen, promising results can be expected for $\tilde{N}$ values close to zero, where the curves obtained at $b_0 = -1$, 0, –1 and 2 exhibit sharp minima. The interval $-0.5 \le \tilde{N} \le 0.5$ was studied in more detail.

Figure 14 shows the behavior of the coefficients $\tilde{U}_N = U_N\Gamma(b_0 + 2)$ in the case of a nearly optimum interpolation with $\tilde{N} = 0$. If the curves for $b_0 \gg 1$ and $b_0 \approx -2$ (attaining the asymptote with delay) are rejected, the data for large N asymptotically tend to a constant level, which correspond to a critical index α close to unity. This conclusion is consistent with the position of the second χ^2 minimum and with the change of sign in U_∞ (Fig. 15). A clear pattern with χ^2minima was observed for $\tilde{N} \le 0.2$; when the $\tilde{N}$ value increased, the first χ^2 minimum approached to and eventually merged with the second minimum. For this reason, no estimates using the first minimum could be obtained for $\tilde{N} \ge 0.3$.

The results of determining the α and W_∞ values are presented in Table 6 and Fig 16. The ideal pattern for α, corresponding to Fig. 9b, is obtained upon expanding the error interval by a factor of two (dashed lines in Fig. 16a), after which the value of $\alpha = 0.96$ is compatible with the results for all $\tilde{N}$. In the fixed interval of $20 \le N < 40$, all four estimates of α coincide for $\tilde{N} = -0.12$ on an accuracy level of 10^{-3}; the main uncertainty is related to a weak dependence on the averaging interval. With an allowance for the double error, we finally obtain

$$\alpha = 0.96 \pm 0.01. \qquad (73)$$

For W_∞ (Fig. 16b), the ideal pattern is obtained immediately and the corresponding value of $W_\infty = 7.4$ is com-

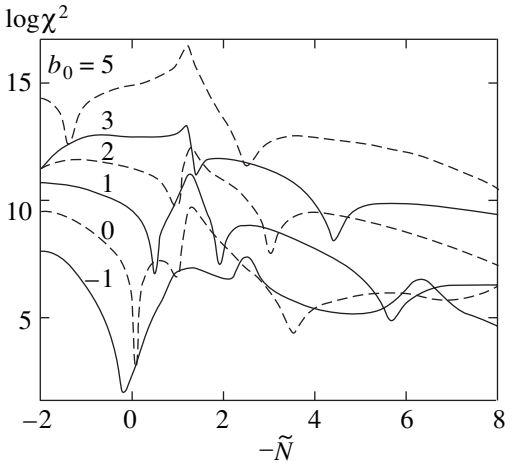

Fig. 13. The plots of χ^2 versus $\tilde{N}$ for the φ^4 theory in the interval of $20 \le N \le 40$ at various fixed b_0 values.

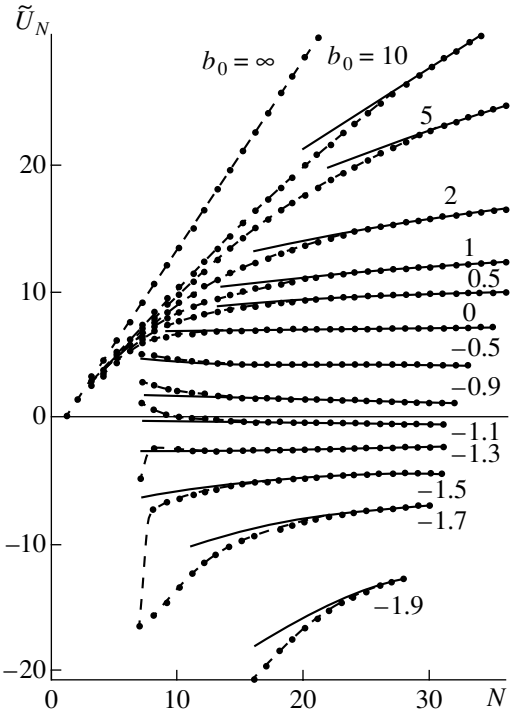

Fig. 14. The plots of $\tilde{U}_N = U_N\Gamma(b_0 + 2)$ versus N for various b_0 (points and dashed curves) and the results of treatment according to the power law (solid curves) for the φ^4 theory. The calculations were performed using a nearly optimum interpolation with $\tilde{b} = 4$, $\tilde{N} = 0$.

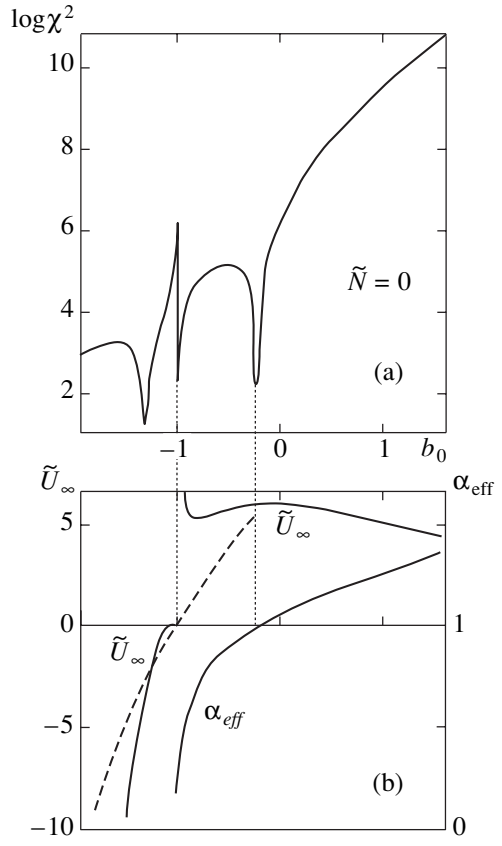

Fig. 15. (a) The pattern of minima in χ^2 for the φ^4 theory in the averaging interval of $20 \le N \le 40$. (b) The plots α_{eff} and $\tilde{U}_\infty$ versus b_0 for $\tilde{N} = 0$. The dashed curve shows the $U_\infty(b_0)$ curve for fixed $\alpha = 1$.

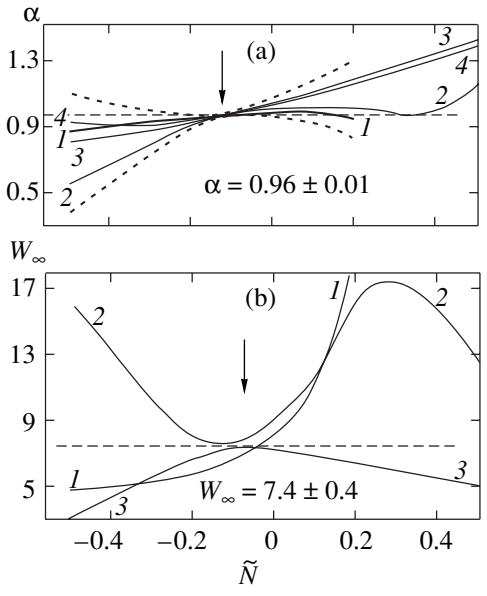

Fig. 16. The plots of various (a) α and (b) W_∞ estimates versus b_0 for the φ^4 theory. The notations are the same as in Fig. 12. Small-dash lines indicate the error interval for α expanded by a factor of two.

patible with all data. Here, the one-side error is minimum at $\tilde{N} = -0.08$, which yields

$$W_\infty = 7.4 \pm 0.4. \tag{74}$$

Correctness of the optimization with respect to $\tilde{b}$ described in Section 6.2 can be demonstrated in a somewhat heuristic manner. For an optimum value of $\tilde{N} = -0.12$ and the $\tilde{b}$ value varied in an interval from 0 to 6, a clear pattern of χ^2 minima was obtained in the middle of the interval. On approaching the boundaries, the first χ^2 minimum approached to and merged with the second minimum exactly as it was observed on increasing $\tilde{N}$. These corresponding results for α and W_∞ are presented in Fig. 17; expanding the error interval by a factor of 2 and 1.1 for α and W_∞, respectively, makes the values (73) and (74) compatible with almost all data (except for a narrow interval at $\tilde{b} = 5.5$, where the proximity of all estimates is obviously accidental. As is seen, the minimum errors also agree with (73) and (74).

Summation of the perturbative series for the Gell-Mann–Low function at finite g values was performed using a procedure analogous to that described in Section 7. The accuracy was evaluated by variation with respect to b_0 and $\tilde{N}$. The variation with respect to b_0 gave a markedly greater N_c values and allowed the $W(g)$ asymptotics to be modified without significantly affecting the results for $g \sim 1$. On varying the $\tilde{N}$ value, with b_0 adjusted so as to maintain a constant value of $\alpha = 0.96$, the most probable value of $W_\infty = 7.4$ is obtained for $\tilde{N} = -0.067$; the uncertainty range indicated in (74) corresponds to the interval $-0.09 \le \tilde{N} \le -0.05$. Table 7 lists the data for $\tilde{N} = -0.067$, with the error estimated by comparison to the results for $\tilde{N} = -0.05$ and -0.09. Note that asymptote (7) is attained rather slowly, the deviation amounting to about 15% even for $g = 100$.

Figure 18 presents a comparison of the results obtained for $g \le 20$ to the data reported by other researchers.

8.2. The Possibility of Logarithmic Branching

Since the value of α differs only slightly from unity, a question arises as to whether the accuracy is sufficient to consider this deviation significant. Formally speaking, this is really so because the error was estimated objectively and there is no ground to expect it to be significantly understated. Nevertheless, the possibility that the equality $\alpha = 1$ is strict is not excluded, since asymptotics (7) may contain logarithmic corrections of the type

$$W(g) = W_\infty g^\alpha (\ln g)^{-\gamma}, \quad g \longrightarrow \infty. \tag{75}$$

For $\gamma > 0$, these corrections may inspire a small decrease in α. In this case, formula (20) contains an

Table 6. Asymptotic parameters for the φ^4 theory obtained for $\tilde{b}_0 = 4$ and various $\tilde{N}$ values by the interpolation with $L_0 = 2$, $L = 5$

Estimates based on	α for $\tilde{N}$									
	−0.5	−0.3	−0.2	−0.12	−0.1	0.0	0.1	0.2	0.3	0.5
First χ^2 minimum	0.863	0.920	0.945	0.962 ±0.005	0.964	0.975	0.974	0.931	–	–
Second χ^2 minimum	0.54	0.78	0.90	0.960	0.970	1.00	1.01	1.01	0.97	1.16
U_∞ changing sign	0.795	0.865	0.915	0.960 −0.961	0.973	1.035	1.105	1.175	1.255	1.415
$U_\infty(b_0)$ slope	0.907	0.90	0.929	0.961 ±0.001	0.971	1.022	1.082	1.147	1.218	1.371
	W_∞ for $\tilde{N}$									
	−0.5	−0.3	−0.2	−0.12	−0.1	0.0	0.1	0.2	0.3	0.5
First χ^2 minimum	4.67	5.22	5.75	6.36 ±0.16	6.63	8.26	11.82	30.9	–	–
$U_\infty(b_0)$ slope	3.02 15.9	5.58 10.0	6.55 7.85	7.35 7.55	7.34 7.61	7.18 9.07	6.78 11.3	6.45 16.5	5.91 17.3	5.05 12.3

additional factor $(\ln N)^{-\gamma}$ with unchanged W_∞, so that the results for U_N can be treated according to Eq. (75) with the parameters

$$\alpha = 1, \quad \gamma \approx 0.14, \quad W_\infty \approx 7.7 \tag{76}$$

without any increase in χ^2. Actually, the possibility of such a logarithmic branching seems to be quite probable for the following reasons.

1. It can be ascertained that the logarithmic branching in the case of strict equality $\alpha = 1$ is unavoidable. Indeed, let us write series (1) in the form of the Sommerfeld–Watson integral [2, 13]:

$$W(g) = \sum_{N=N_0}^{\infty} W_N(-g)^N = -\frac{1}{2i}\oint_C dz \frac{\tilde{W}(z)}{\sin \pi z} g^z, \tag{77}$$

where $\tilde{W}(z)$ is the analytical continuation of W_N onto the complex plane ($\tilde{W}(N) = W_N$) and C is the contour containing the points N_0, $N_0 + 1$, $N_0 + 2$, … (Fig. 19). If $z = \alpha$ is the extreme right-hand singularity of $\tilde{W}(z)/\sin \pi z$, we can modify the contour into the position C' and show that this singularity determines the behavior of $W(g)$ as $g \longrightarrow \infty$. The purely power law (7) corresponds to the presence of a simple pole at $z = \alpha$, while the law described by Eq. (75) corresponds to a singularity of the $(z - \alpha)^{\gamma - 1}$ type.[9]

[9] It is clear from the above considerations that the assumption of analyticity of the coefficient function on the real axis for $N \geq N_0$, which is necessary for interpolation, is confirmed in all cases by the results obtained.

Note that the first term β_0 is absent in the expansion of the β function (5) simply by its definition, while vanishing of the next coefficient β_1 is accidental. Indeed, in the $(4 - \epsilon)$-dimensional φ^4 theory, the latter term is nonzero and has a magnitude on the order of ϵ; accordingly,

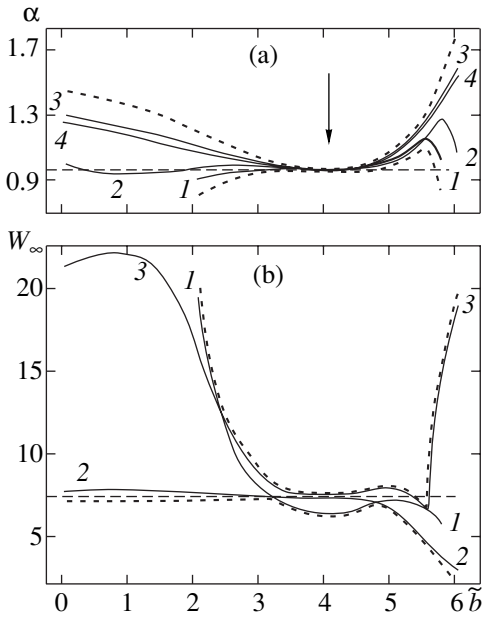

Fig. 17. The plots of various (a) α and (b) W_∞ estimates versus $\tilde{b}$ for the φ^4 theory. The notations are the same as in Fig. 12. Small-dash lines indicate the error interval expanded by a factor of 2 and 1.1 for α and W_∞ values, respectively.

Table 7. The Gell-Mann–Low function for the φ^4 theory (values in parentheses indicate the error estimated in units of the last decimal digit)

g	$\beta(g)$	g	$\beta(g)$
0.2	0.04993(2)	30	138.7(50)
0.4	0.18518(26)	40	193.2(75)
0.6	0.3939(10)	50	248.3(100)
0.8	0.6667(27)	60	303.9(127)
1	0.9952(51)	70	359.7(155)
2	3.272(33)	80	415.6(182)
3	6.278(85)	90	471.7(212)
4	9.758(157)	100	527.7(240)
5	13.57(25)	150	808.1(389)
6	17.64(36)	200	1087(54)
7	21.90(47)	250	1366(70)
8	26.32(60)	300	1644(86)
9	30.87(75)	350	1920(101)
10	35.53(90)	400	2196(127)
15	59.95(175)	450	2471(133)
20	85.59(275)	500	2745(149)
25	111.9(38)	$g \longrightarrow \infty$	$7.41g^{0.96}$

$\mathcal{W}(1) \sim \epsilon$. The limiting transition $\epsilon \longrightarrow 0$ shows that, in the four-dimensional case, $\mathcal{W}(1) = 0$ and a simple pole cannot take place at $\alpha = 1$. If the function $\mathcal{W}$ tends to zero as $z \longrightarrow 1$ by the law $\mathcal{W}(z) = \omega_0(z-1)^\gamma$, then

$$\beta(g) = \frac{\omega_0}{\Gamma(1-\gamma)} g(\ln g)^{-\gamma}, \quad g \longrightarrow \infty \qquad (78)$$

and the positive definiteness of γ has a quite clear origin.

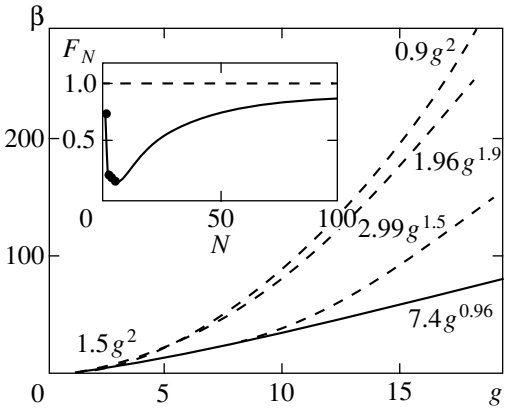

Fig. 18. A comparison of the Gell-Mann–Low function for the φ^4 theory calculated in this work (solid curve) to the results reported by other researchers (dashed curves top to bottom corresponding to [12, 13, 14], respectively). The inset shows a reduced coefficient function (in this scale, differences between the data obtained using various interpolation methods are insignificant).

2. Lipatov [29] considered the class of field theories (generalizing the four-dimensional φ^4 theory) with a nonlinearity of the φ^n type and a space dimension of $d = 2n/(n-2)$, for which a logarithmic situation takes place. For all such theories, $\beta_1 = 0$; however, this coefficient differs from zero when d decreases. Therefore, $\mathcal{W}(1) = 0$ by analogy with the cases considered above. In the limit $n \longrightarrow \infty$, the Gell-Mann–Low function is exactly calculated [29] and the extreme right-hand singularity of $\mathcal{W}(z)$ has the form of $(z-1)^{3/2}$, which leads to asymptotics of the type $\beta(g) \propto g(\ln g)^{-3/2}$. From the continuity considerations, we may expect for large but finite n values that a nonanalytical zero of the type $(z-1)^\gamma$ is retained and the singularity at $z = 1$ is still the extreme right-hand one. Therefore, asymptotics (78) is natural for such field theories and it is not surprising that it may be retained up to $n = 4$. Note that W_∞ is negative when $n \longrightarrow \infty$, so that the Gell-Mann–Low function possesses a zero; a direct extrapolation of the results to $n = 4$ leads to an analogous conclusion for the φ^4 theory [29]. In fact, with this extrapolation we must take into account that the index γ series from 3/2 to small values such as in (76); then the change in sign of the asymptotics naturally takes place according to (78) at $\gamma = 1$. The positiveness of ω_0 follows from the matching of $\mathcal{W}(2) \sim \omega_0$ and the positiveness of β_2 [29].

Anyhow, we have to select between two possibilities: (i) a purely power law (7) with a critical index α slightly below unity and (ii) an asymptotics of the type (78) with $\gamma > 0$. In both cases, the φ^4 theory turns out to be self-consistent.

8.3. On the Results Obtained in [12, 13]

The curves in Fig. 14 display for $N < 10$ a linear portion where $\tilde{U}_N \approx 1.1(N-1)$, which is stable with respect to changes both in b_0 and in the extrapolation procedure. This region might be considered as a true asymptotics for $\tilde{U}_N$ (assuming the results for $N > 10$ to be the interpolation artifacts), corresponding to the dependence $\beta(g) \approx 1.1g^2$, which is close to the result obtained in [12, 13].

In fact, stability of the above region has a different origin. This behavior is related to a characteristic "trough" in

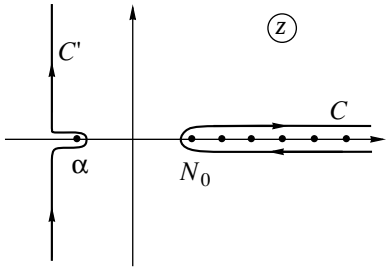

Fig. 19. Integration contour for Eq. (77).

the reduced coefficient function F_N at $N \lesssim 10$ (see the inset in Fig. 18). Modeling this trough by assuming $F_3 = F_4 = \ldots = F_{10} = 0$ and taking into account Eqs. (19) and (22), we obtain

$$\tilde{U}_N = c\Gamma(b_0 + 2) \times \sum_{K=1}^{N} F_K(-1)^K \frac{\Gamma(K+b)}{\Gamma(K+b_0)} C_{N-1}^{K-1}. \quad (79)$$

From this we obtain for $N \le 10$ and all b_0 the result $\tilde{U}_N = 1.5(N-1)$, which is determined by the first nonvanishing coefficient F_2 (see the curve for $b_0 = \infty$ in Fig. 14) and is close to the real situation. For the β function, this result implies that a single-loop law $1.5g^2$ is valid up to $g \sim 10$.

Upon modeling the trough in F_N more precisely by assuming $F_3 = F_4 = \ldots = F_{10} = \epsilon$ and using (26) we obtain, in the case when the ratio of gamma functions in (79) reduces to a polynomial for $b_0 = b - p$ with an integer p and N in the interval $p + 2 \le N \le 10$,

$$\tilde{U}_N = W_2\left\{\left(1 - \frac{\epsilon}{F_2}\right)(N-1) + \frac{\epsilon}{F_2}\frac{1+b_0}{1+b}\right\}. \quad (80)$$

This result indicates that the linear slope varies but remains independent of b_0. More complicated calculations show that Eq. (80) is valid for arbitrary b_0 to within corrections on the order of $\epsilon/(N+b_0)^{b+1}$; for $\epsilon = 0.2$ (see Fig. 18), we obtain $\tilde{U}_N = 1.1(N-1) + \text{const}$, where the last constant depends on b_0 but does not exceed a few tenths in the interval $0 < b_0 < 10$. Thus, a notion of the quadratic law with modified coefficient $\beta(g) = 1.5(1 - \epsilon/F_2)g^2$ is really meaningful in the interval $1 \lesssim g \lesssim 10$ but is a consequence of the trough in F_N.[10] The limited width of the trough indicates that this law is not related to a real asymptotics (whatever it is).

The above considerations clearly indicate that the result obtained in [12, 13] is by no means a computational error and objectively reflects the behavior of the β function for $g \lesssim 10$. This result is unavoidably obtained upon summing a series with a small number of expansion coefficients, since no other portion obeying the power law can be found in Fig. 14 for $N < 7$ (the points on the curves for $b_0 < 0$ are omitted for clarity, because their sharp oscillations would overload the pattern).

8.4. The Question of "Triviality" of the φ⁴ Theory

The situation when the β function possesses asymptotics of the g^α type with $\alpha > 1$ can be given a two-fold interpretation. From the standpoint of finiteness of a physical charge at large distances, the φ^4 theory is inconsistent: the effective charge $g(L)$ turns into infinity at a certain L_c (Landau pole), while for $L < L_c$ the $g(L)$ is undetermined. Considering the field theory as a limiting case of the lattice theories, the φ^4 theory is "trivial": the physical charge tends to zero for any value of the seeding charge.

In recent years, the problems related to the concept of triviality were actively discussed by several researchers (see [30, 31] and references therein). On the one hand, the existing indications of triviality of the φ^4 theory were emphasized; on the other hand, the φ^4 theory was declared verified (with a positive result) by numerical modeling on a lattice. Let us briefly discuss this problem as well.

The φ^4 theory is strictly proven to be trivial in a space with the dimensionality $d > 4$ and nontrivial for $d < 4$ [32, 33]. In the case of $d = 4$, the obtained inequalities were only slightly insufficient for the statement of triviality. Such situations are usually considered by mathematicians as insignificant and not worth special effort, which explains why triviality of the φ^4 theory is considered as "virtually proved." From the standpoint of physics, this optimism is by no means justified: on the modern level, the aforementioned results for $d \ne 4$ are rather primitive, being merely a consequence of the theory of renormalization and a single-loop renorm group. On the contrary, the situation with $d = 4$ is physically highly complicated and no analytical approaches to solving this problem have been developed so far.

In the author's opinion, the results of numerical experiments on the lattice revealed nothing unexpected. In view of the absence of zeros of the β function, the effective charge $g(L)$ always decreases with the distance. However, the numerical methods cannot answer the question as to whether the "charge zero" does exist, which is explained by limited lattice dimensions. There are many cases of misunderstanding related to the charge normalization: even in the "natural" normalization used in this work, the quadratic law is extended to $g \sim 10$ (see Section 8.3); traditional normalizations extend this interval even greater, for example, up to $g \sim 600$ when the interaction term is written in the form of $g\varphi^4/8$. Therefore, behavior of any quantities is indistinguishable from trivial in a broad range of parameters.

Among old publications, only the paper of Freedman *et al.* [34] is worth of mentioning where it was stated that $g(L)$ uniformly decreases in g_0, which is actually indicative of the "charge zero." However, judging by the results, the charge normalization employed in [34] differed by a factor of about 100 (an expression for the action obviously contains a misprint) from that used in this work and all results for finite g_0 fell within a region where the quadratic law is operative. Nontrivial results were only obtained for $g_0 = \infty$ by reduction to the Ising model. Although this reduction is apparently possible, there is no method (except for extrapolation)

[10] This law is more clearly pronounced for the Borel image and is somewhat distorted for the β function as a result of integration in Eq. (8); however, $\beta(g)$ remains downward-convex up to $g \sim 100$.

to establish a correspondence between normalization of the field variable in the Ising model and that in the initial φ^4 theory. This leads to uncertainty in the charge normalization, an allowance for which makes unjustified any conclusions concerning uniform convergence.

Now let us turn to the original results of [30, 31]. The main idea was illustrated by the example of a nonideal Bose gas possessing a well-known spectrum of the Bogolyubov type: $\epsilon(k) \sim k$ for small k and $\epsilon(k) \sim k^2$ for $k \longrightarrow \infty$. Let us pass to the "continuum limit" by allowing two characteristic scales of the problem (scattering length and interparticle distance) to tend to zero. If the first value tends to zero rather rapidly, a "quite trivial theory" appears and a quadratic spectrum of the ideal gas is restored. If the limiting transition is performed so as to maintain a certain relationship (ensuring constant sound velocity) between the two scales, a "trivial theory with nontrivial vacuum" appears and the spectrum becomes strictly linear (i.e., strongly different from that of the ideal gas), although no interaction of quasiparticles (phonons) takes place. The latter scenario was suggested for the continuum limit of the φ^4 theory, stating that it is logically self-consistent.

Even if the last statement is accepted, a question still remains unanswered as to why this limiting transition does physically take place. For a Bose gas of neutral atoms, there is no real possibility of simultaneously changing both the gas density and the scattering length. The situation required for the authors of [30, 31] may take place only in the case of a special long-range interaction, whereby a change in the density affects the Debye screening radius. However, this scenario is not arbitrary and can be predicted based on the initial Hamiltonian.

It was stated [30, 31] that the assumption concerning a nontrivial character of the continuum limit was confirmed by the results of numerical modeling on the lattice. However, this conclusion was based only on a particular interpretation of the "experimental" data, rater than on a direct experimental evidence: the numerical experiments were performed deep in the region of the single-loop law and could not contain any information concerning the triviality. The results, however unusual they might seem, must by explained within the framework of a weak coupling limit.

Triviality of the φ^4 theory leads to the non-renormalizability of the Higgs spectrum of the Standard Model. This results in violating one of the basic postulates, the principle of renormalizability. Thus, papers [30, 31] were stimulated by the wish to resolve the difficulties. According to the results obtained in this work, no such difficulties were inherent in the system studied.

9. CONCLUSION

This paper develops an algorithm for summing divergent series of the perturbation theory with arbitrary values of the coupling constant. Verification on the test examples showed that the algorithm is stable under conditions of strongly restricted information and confirmed reliability of the error estimation. The main physical result of this study consists in restoring the Gell-Mann–Low function of the φ^4 theory and demonstrating its self-consistency. The latter conclusion agrees with the absence of renormalon singularities established previously [9].

The proposed algorithm can be applied to solving many other problems as well, in particular, to restoring the Gell-Mann–Low functions in quantum electrodynamics and quantum chromodynamics. At present, solving this task is complicated by the absence of calculations of the full-scale Lipatov asymptotics in these theories, although the basis for such calculations is fully prepared [27, 35–39]. Application of the proposed algorithm to the theory of phase transitions may increase the accuracy of calculation of the critical indices by at least two–three orders of magnitude.

ACKNOWLEDGMENTS

This study was supported by the INTAS foundation (grant no. 99-1070) and by the Russian Foundation for Basic Research (project no. 00-02-17129).

REFERENCES

1. I. M. Suslov, Pis'ma Zh. Éksp. Teor. Fiz. **71**, 315 (2000) [JETP Lett. **71**, 217 (2000)].
2. L. N. Lipatov, Zh. Éksp. Teor. Fiz. **72**, 411 (1977) [Sov. Phys. JETP **45**, 216 (1977)].
3. *Large Order Behavior of Perturbation Theory*, Ed. by J. C. Le Guillou and J. Zinn-Justin (North-Holland, Amsterdam, 1990).
4. J. Zinn-Justin, Phys. Rep. **70**, 109 (1981).
5. E. B. Bogomolny, V. A. Fateyev, and L. N. Lipatov, Sov. Sci. Rev., Sect. A **2**, 247 (1980).
6. J. C. Le Guillou and J. Zinn-Justin, Phys. Rev. Lett. **39**, 95 (1977); Phys. Rev. B **21**, 3976 (1980).
7. G. A. Baker, Jr., B. G. Nickel, M. S. Green, and D. I. Meiron, Phys. Rev. Lett. **36**, 1351 (1976); Phys. Rev. B **17**, 1365 (1978).
8. J. C. Le Guillou and J. Zinn-Justin, J. Phys. Lett. **46**, L137 (1985); J. Phys. (Paris) **48**, 19 (1987); **50**, 1365 (1989).
9. I. M. Suslov, Zh. Éksp. Teor. Fiz. **116**, 369 (1999) [JETP **89**, 197 (1999)].
10. N. N. Bogoliubov and D. V. Shirkov, *Introduction to the Theory of Quantized Fields* (Nauka, Moscow, 1976; Wiley, New York, 1980).
11. V. S. Popov, V. L. Eletskiĭ, and A. V. Turbiner, Zh. Éksp. Teor. Fiz. **74**, 445 (1978) [Sov. Phys. JETP **47**, 232 (1978)].
12. D. I. Kazakov, O. V. Tarasov, and D. V. Shirkov, Teor. Mat. Fiz. **38**, 15 (1979).
13. Yu. A. Kubyshin, Teor. Mat. Fiz. **58**, 137 (1984).
14. A. N. Sissakian *et al.*, Phys. Lett. B **321**, 381 (1994).

15. A. A. Vladimirov and D. V. Shirkov, Usp. Fiz. Nauk **129**, 407 (1979) [Sov. Phys. Usp. **22**, 860 (1979)].
16. M. V. Sadovskiĭ, Usp. Fiz. Nauk **133**, 223 (1981) [Sov. Phys. Usp. **24**, 96 (1981)].
17. I. M. Suslov, Usp. Fiz. Nauk **168**, 503 (1998) [Phys. Usp. **41**, 441 (1998)].
18. F. M. Dittes, Yu. A. Kubyshin, and O. V. Tarasov, Teor. Mat. Fiz. **37**, 66 (1978).
19. Yu. A. Kubyshin, Teor. Mat. Fiz. **57**, 363 (1983).
20. G. H. Hardy, *Divergent Series* (Clarendon, Oxford, 1949; Inostrannaya Literatura, Moscow, 1951).
21. Yu. V. Sidorov, M. V. Fedoryuk, and M. I. Shabunin, *Lectures on Theory of Functions of Complex Variable* (Nauka, Moscow, 1976), Para. 32.
22. W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes* (Cambridge Univ. Press, Cambridge, 1988).
23. C. M. Bender and T. T. Wu, Phys. Rev. **184**, 1231 (1969); Phys. Rev. D **7**, 1620 (1973).
24. J. Cizek and E. R. Vrskay, Int. J. Quantum Chem. **21**, 27 (1982).
25. I. M. Suslov, Zh. Éksp. Teor. Fiz. **117**, 659 (2000) [JETP **90**, 571 (2000)].
26. A. I. Mudrov and K. B. Varnashev, Phys. Rev. E **58**, 5371 (1998).
27. S. V. Faleev and P. G. Silvestrov, Nucl. Phys. B **463**, 489 (1996).
28. S. V. Faleev and P. G. Silvestrov, Phys. Lett. A **197**, 372 (1995).
29. L. N. Lipatov, Zh. Éksp. Teor. Fiz. **71**, 2010 (1976) [Sov. Phys. JETP **44**, 1055 (1976)].
30. M. Consoli and P. M. Stevenson, Z. Phys. C **63**, 427 (1994).
31. A. Agodi, G. Andronico, P. Cea, *et al.*, Mod. Phys. Lett. A **12**, 1011 (1997).
32. J. Frolich, Nucl. Phys. B **200** (FS4), 281 (1982); M. Aizenman, Commun. Math. Soc. **86**, 1 (1982).
33. J. P. Eckmann and R. Epstein, Commun. Math. Phys. **64**, 95 (1979).
34. B. Freedman, P. Smolensky, and D. Weingarten, Phys. Lett. B **113B**, 481 (1982).
35. E. B. Bogomolny and V. A. Fateyev, Phys. Lett. B **71B**, 93 (1977); L. N. Lipatov, A. P. Bukhvostov, and E. I. Malkov, Phys. Rev. D **19**, 2974 (1979).
36. G. Parisi, Phys. Lett. B **66B**, 382 (1977).
37. C. Itzykson, G. Parisi, and J. B. Zuber, Phys. Rev. D **16**, 996 (1977); R. Balian, C. Itzykson, G. Parisi, and J. B. Zuber, Phys. Rev. D **17**, 1041 (1978).
38. E. B. Bogomolny and V. A. Fateyev, Phys. Lett. B **76B**, 210 (1978).
39. I. I. Balitsky, Phys. Lett. B **273**, 282 (1991).

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 93, No. 1, 2001, pp. 103–110.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 1, 2001, pp. 119–126.

FLUIDS

Theory of the Resonant Properties of Electrons Localized on the Surface of Liquid Helium

P. D. Grigor'ev[1, 2]* and A. M. Dyugaev[2, 3]

[1]*Landau Institute of Theoretical Physics, Russian Academy of Sciences, Chernogolovka, Moscow oblast, 142432 Russia*
[2]*Grenoble High Magnetic Field Laboratory, MPI-FRF and CNRS, BP166, Grenoble, F-38042 France*
[3]*Max-Plank-Institut für Physik Komplexer Systeme, D-01187, Dresden, Deutschland*
**e-mail: pashag@itp.ac.ru*
Received February 1, 2001

Abstract—The problem of the shape of the line of optical transition of an electron between bound states on the surface of liquid helium is solved within the independent boson model. Such bound states are realized, for example, in the potential of a positively charged impurity located on a substrate or in the field of a He^+ ion located beneath the surface. Reference is made to the importance of the relaxation processes of the dimple on the helium surface under the electron. The adiabatic approximation, in the case of which the dimple does not change during the time of electron transition, is not always valid. At low temperatures, two maxima may appear on the absorption line. It is demonstrated that the far tails of the optical absorption line feature a universal (Urbach rule) exponential dependence on the electron transition energy. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Electrons on the surface of liquid helium are subjected to numerous investigations [1]. They form a two-dimensional electron gas and are held on the surface by image forces and electric field. We will treat the problem of the line of optical transition of an electron between bound states that are realized in the impurity potential and in the magnetic field. If a charged impurity is placed on a substrate beneath a thin film of helium, this impurity develops on the surface of a liquid an attractive potential capable of holding a single electron. The lower levels of such a potential coincide with the levels of harmonic oscillator. A more exact determination of levels with due regard for the formation of dimple, i.e., static deformation of the surface under the electron, was obtained numerically [2] by varying the total electron energy and the liquid surface. Electrons may also be localized using the magnetic field perpendicular to the surface of helium. Here, the formation of a dimple is likewise of importance. Because the energy of dimple is different for different electron levels, a shift of the cyclotron resonance frequency occurs [3]. The effects of electron level broadening were treated only recently [4]. The shape of the optical absorption line or of the cyclotron resonance line is defined both by the natural width of electron levels and by the processes of relaxation of static deformation of the helium surface under the electron. The approximation, in the case of which the dimple does not have enough time to change during the electron transition from level to level, defines well the transition energy rather than the shape and width of the absorption line. Moreover, it is not always that this adiabatic approximation is valid. We will demonstrate that, under some conditions, two maxima may appear on the absorption line.

The system of electrons on discrete surface levels in question is of further interest because it is a good realization of the exactly solvable independent boson model which we are now going to treat.

2. MODEL

The effects of relaxation of a dimple under an electron in bound states may be described by the Hamiltonian of the independent boson model [5],

$$\hat{H} = \sum_i \varepsilon_i C_i^+ C_i + \sum_q \hbar\omega_q a_q^+ a_q + \sum_{i,q} M_{q,i}(a_q^+ + a_q) C_i^+ C_i, \quad (1)$$

where C_i^+ is the operator of electron production on a level of number i and energy ε_i; and a_q^+ is the operator of production of ripplon, i.e., of a quantum of a capillary-gravity wave on the helium surface with the wave vector q. The ripplon dispersion law is given by

$$\omega_q^2 = \frac{\alpha}{\rho} q(q^2 + \kappa^2)\tanh qd, \quad (2)$$

where α is the surface tension, ρ is the liquid density, κ is the inverse gravity-capillary length, and d is the thickness of helium film. The value of $\kappa \approx 20$ cm^{-1} is much less than the characteristic wave vectors in the problem being treated. Therefore, it is almost always

possible to assume that $\kappa = 0$. The matrix element of electron–ripplon interaction M_{qi} will be determined below.

The Hamiltonian given by Eq. (1) contains no terms of the $a_q^+ C_i^+ C_j$ type responsible for electron transitions with ripplon emission. This is attributed to the smallness of the matrix element

$$M_{qij} \propto \exp\left(-\frac{a^2 q^2}{4}\right),$$

where q is the wave vector of emitted ripplon with energy ω_q equal to that of electron transition, and a is the size of the wave function of electron in bound state. The ripplon spectrum is very soft; therefore, for the given vector q, it is always $aq \gg 1$. The sole exception is provided by transitions between very weakly split discrete levels [4]. The ground and first excited levels in the potential of positively charged impurity are not split. Ripplon transitions in a high magnetic field are likewise suppressed. The Hamiltonian given by Eq. (1) is diagonalized using the unitary transformation [5]

$$\bar{H} = e^S H e^{-S}, \tag{3}$$

where the operator S is defined by the expression

$$S = \sum_i C_i^+ C_i \sum_q \frac{M_{qi}}{\omega_q}(a_q^+ - a_q). \tag{4}$$

The new Hamiltonian $\bar{H}$ is diagonal,

$$\bar{H} = \sum_i (\varepsilon_i - \Delta_i) C_i^+ C_i + \sum_q \omega_q \bar{a}_q^+ \bar{a}_q. \tag{5}$$

The shift of electron levels Δ_i is nothing but the energy of formation of a dimple under the electron in the ith bound state,

$$\Delta_i = \sum_q \frac{M_{qi}^2}{\hbar\omega_q}. \tag{6}$$

The ripplon spectrum does not vary, but the position of equilibrium of ripplon modes is shifted,

$$\bar{a}_q = a_q - \sum_i \frac{M_{qi}}{\omega_q} C_i^+ C_i. \tag{7}$$

The new electron operators $\bar{C}_i$ are related to C_i by the unitary transformation

$$\bar{C}_i = e^S C_i e^{-S},$$

$$\bar{C}_i = C_i \exp\left\{-\sum_q \frac{M_{iq}}{\omega_q}(a_q^+ - a_q)\right\}. \tag{8}$$

The intensity of light absorption is proportional to the conductance of the electron system σ, which may be determined by Kubo's formula [5],

$$\mathrm{Re}\sigma_{\alpha\beta}(\omega) = \frac{1}{2\omega}\int_{-\infty}^{\infty} e^{i\omega t}\langle j_\alpha(t) j_\beta(0)\rangle dt, \tag{9}$$

where j_α is the current operator, $j_\alpha = \sum_{ij} P_{ij,\alpha}\bar{C}_i^+ \bar{C}_j$. The matrix element of optical transition P_{ij} is taken to be constant. The calculation of the current–current correlation function on the basis of Eqs. (8) and (9) leads to the expression [5]

$$\mathrm{Re}\sigma_{\alpha\beta}(\omega) = \frac{1}{2\omega}\sum_{ij} n_i(1-n_j)P_{ij,\alpha}P_{ij,\beta} \times \int_{-\infty}^{\infty} \exp\left\{\frac{it}{\hbar}(\omega\hbar + \varepsilon_i - \varepsilon_j - \Delta_i + \Delta_j) - \Phi_{ij}(t)\right\} dt, \tag{10}$$

where Δ_i is the energy of dimple given by Eq. (6), and the function $\Phi_{ij}(t)$ is defined by the relation

$$\Phi_{ij}(t) = \sum_q \frac{(M_{qi} - M_{qj})^2}{\omega_q^2} \times [(n_q + 1)(1 - e^{-i\omega_q t}) + n_q(1 - e^{i\omega_q t})], \tag{11}$$

where n_i and n_q are the occupation numbers of electrons and ripplons, respectively. At low temperatures, numbers n_i are equal to 0 or 1, and n_q represents the Bose function,

$$n_q = \frac{1}{e^{\hbar\omega_q/T} - 1}.$$

It is our objective to investigate the general solution given by Eqs. (10) and (11) as applied to optical electron transitions on the surface of liquid helium. Unlike the analogous problem on electron transitions from level to level in the impurity potential in a solid [5], the ripplon spectrum is continuous and given by expression (2). The optical phonon spectrum is independent of the wave vector [5].

The matrix element of electron–ripplon interaction M_{qi} is given by the expression [1]

$$M_{qi} = \int d^2 r |\psi_i|^2 e^{i\mathbf{qr}} Q(q) F(q),$$

$$Q^2(q) = \frac{\hbar q \tanh qd}{2\rho\omega_q},$$

$$F(q) = eF_\perp + \int_0^\infty \frac{(\epsilon - 1)e^2 q}{4(\epsilon + 1)Z}\left(\frac{1}{qZ} - K_1(qZ)\right)\varphi^2(Z) dZ, \tag{12}$$

where $\varphi(Z)$ is the electron wave function along the Z axis perpendicular to the helium surface, $\psi(r)$ is the

electron wave function along the surface, and ϵ is the permittivity of helium. The pressing electric field $E_\perp$ includes the external field E_{ext}, the impurity field, and the image field of the substrate,

$$E_\perp = E_{\text{ext}} + \frac{Z^* e}{d^2} + \frac{\delta}{4}\frac{e}{d^2}, \qquad Z^* = \frac{2Z}{1+\epsilon_S}, \quad \delta = \frac{\epsilon_S - 1}{\epsilon_S + 1}, \tag{13}$$

Z^* is the effective charge of impurity, and ϵ_S is the permeability of the substrate. The second term in $F(q)$ in Eq. (12) is the contribution made by the electron interaction with the surface of liquid. In our case, the average value of $\langle Z\rangle$ is much less than a, where a is the characteristic scale of the function $\psi(r)$. Because the ripplon momenta $q < a^{-1}$ will be important, one can expand the Bessel function

$$K_1(qZ) \approx \frac{1}{qZ} - qZ$$

at $qZ \ll 1$ (Eq. (12)), which simplifies the expression for $F(q)$,

$$F(q) = eE_\perp + \frac{(\epsilon - 1)e^2 q^2}{4(\epsilon + 1)}. \tag{14}$$

The contribution by these two terms to $F(q)$ is compared for the characteristic values of $qa \le 1$ at $E_\perp \sim$ 300 V/cm.

Because the problem is axisymmetric, it is convenient to expand the plane wave in Eq. (12) in terms of cylindrical functions. In so doing, the contribution to M_{qi} in Eq. (12) is made only by a harmonic with zero projection of momentum and Eq. (12) yields

$$M_{qi} = \int_0^\infty 2\pi r dr |\psi_i|^2 J_0(qr) Q(q) F(q), \tag{15}$$

where $J_0(qr)$ is the Bessel function. A charged impurity develops on the surface of liquid the potential,

$$V(r) = \frac{Z^* e^2}{\sqrt{r^2 + d^2}} \approx \frac{Z^* e^2}{d}\left(1 - \frac{r^2}{2d^2}\right),$$

which may be expanded in terms of the parameter $r/d \ll 1$. Therefore, in the impurity potential, as well as in the magnetic field, the squares of the wave functions of two lower levels are given by

$$\psi_0^2 = \frac{1}{\pi a^2} e^{-r^2/a^2}, \quad \psi_1^2 = \frac{1}{\pi a^2}\left(\frac{r}{a}\right)^2 e^{-r^2/a^2}, \tag{16}$$

where $a^2 = \hbar/m\omega_0$, and $\hbar\omega_0$ is the distance between the electron levels. From Eqs. (15) and (16) we find the matrix elements M_{q0} and M_{q1},

$$M_{q0} = \exp\left(-\frac{q^2 a^2}{4}\right) F(q) Q(q), \qquad M_{q1} = \left(1 - \frac{q^2 a^2}{4}\right)\exp\left(-\frac{q^2 a^2}{4}\right) F(q) Q(q). \tag{17}$$

We use Eqs. (6) and (17) to derive the expression for the energy of dimples in the ground state Δ_0 and in the first excited state Δ_1 of electron,

$$\Delta_0 = \int_0^\infty \frac{q dq}{4\pi\alpha} \exp\left(-\frac{q^2 a^2}{2}\right)\frac{F^2(q)}{q^2 + \kappa^2}, \qquad \Delta_1 = \int_0^\infty \frac{q dq}{4\pi\alpha} \exp\left(1 - \frac{q^2 a^2}{4}\right)^2 \exp\left(-\frac{q^2 a^2}{2}\right)\frac{F^2(q)}{q^2 + \kappa^2}. \tag{18}$$

It is only the difference $\Delta_0 - \Delta_1$ that is of interest. We assume, for simplicity, that $F(q) = \text{const} = F$ and $\kappa = 0$, to derive from Eq. (18)

$$\Delta_0 - \Delta_1 = \frac{3F^2}{32\pi\alpha}. \tag{19}$$

Equations (10) and (11) yield the expression for the intensity of optical absorption $I(\omega)$,

$$I(\omega) \propto \int_{-\infty}^{\infty} dt \cos\{(\omega - \omega_0)t - J_1(t)\} e^{-J_2(t)}, \tag{20}$$

where $\omega_0\hbar = \varepsilon_1 - \varepsilon_0 + \Delta_0 - \Delta_1$,

$$J_1(t) = \sum_q \frac{(M_{q0} - M_{q1})^2}{\hbar^2 \omega_q^2} \sin\omega_q t, \qquad J_2(t) = \sum_q \frac{(M_{q0} - M_{q1})^2}{\hbar^2 \omega_q^2} \coth\frac{\omega_q}{2T}(1 - \cos\omega_q t). \tag{21}$$

We derive from Eqs. (20) and (21), in view of (17),

$$J_1(t) = \int_0^\infty \frac{q dq (qa)^4}{64\alpha} \exp\left(-\frac{q^2 a^2}{2}\right) \times \frac{F^2}{(q^2 + \kappa^2)\hbar\omega_q} \sin\omega_q t, \qquad J_2(t) = \int_0^\infty \frac{q dq (qa)^4}{64\alpha} \exp\left(-\frac{q^2 a^2}{2}\right)\frac{F^2}{q^2 + \kappa^2} \times \frac{(1 - \cos\omega_q t)}{\hbar\omega_q} \coth\frac{\omega_q \hbar}{2T}. \tag{22}$$

The integrals defining the functions $J_1(t)$ and $J_2(t)$ converge rapidly at $qa > 1$ (22). The value of $a \sim 100$ Å corresponds to the transition energy $\hbar\omega_0 = \hbar^2/ma^2 \sim 10$ K. An important parameter of the problem is the ripplon frequency $\omega(q)$ at $q^2 = 2/a^2$. It defines approximately the inverse relaxation time of a dimple. At $a = 100$ Å, this frequency $\omega(a) \approx 10^9$ s^{-1}, and the ripplon energy is $\hbar\omega(a) \approx 0.03$ K. Ripplons with a high energy have an exponentially small matrix element of interaction with electron. The short-term asymptotic behavior of the functions $J_1(t)$ and $J_2(t)$ given by Eq. (22) is readily determined in the classical limit of $\hbar\omega(a) \ll T$, when the frequency $\omega(a)$, which is related to the wave vector q by a fairly complex relation (2), drops out completely during integration of Eq. (22) in the region of $\omega_q(t) \ll 1$,

$$J_1(t) = \frac{F^2 t}{32\pi\hbar\alpha}, \quad J_2(t) = J_1(t)\frac{tT}{\hbar}. \tag{23}$$

We restrict ourselves to this approximation and find, from Eqs. (20) and (23),

$$I(\omega) \sim \exp\left\{-\frac{(\omega\hbar + \varepsilon_0 - \varepsilon_1 - \Delta)^2}{\Delta T}\right\}. \tag{24}$$

The shift of transition energy Δ caused by the deformation of the helium surface is

$$\Delta = \Delta_0 - \Delta_1 + \frac{J_1(t)}{t} = \frac{F^2}{8\pi\alpha}. \tag{25}$$

This value of Δ coincides exactly with the result obtained numerically by Cheng and Platzman [6] in the same limit. It appears of interest to treat the case of thin film, $d < a$, which has not been experimentally investigated.

3. THIN FILM

In this limiting case, the ripplon frequency $\omega(q) \propto q^2$ (2). During integration in Eq. (22), it is convenient to introduce the variable $X = (qa)^2/2$ and define the characteristic frequency $\omega(a)$ by relation (2),

$$\omega(q) = \omega(a)X, \quad \omega^2(a) = 4\frac{\alpha}{\rho}\frac{d}{a^4}. \tag{26}$$

We use Eqs. (22) and (26) to find

$$J_1(t) = A\frac{t_*}{1 + t_*^2}, \quad t_* = \omega(a)t, \qquad A = \frac{F^2}{32\alpha\hbar\omega(a)}. \tag{27}$$

In Eq. (27), the parameter A is the ratio of the difference of the dimple energies $\Delta_1 - \Delta_0$ to $\hbar\omega(a)$, and the time t is reduced to the dimensionless form by multiplying it by the characteristic frequency $\omega(a)$.

Equations (22) and (26) yield the expression for the function $J_2(t_*)$,

$$J_2(t_*) = A\int_0^\infty e^{-X}(1 - \cos t_* X)\coth\frac{X}{2T_*}dX, \tag{28}$$

where T_* is the reduced temperature, $T_* = T/\hbar\omega(a)$. In the classical limit of $T_* \gg 1$ and from Eq. (28), we derive the dependence of the function J_2 on the reduced time t_*,

$$J_2(t_*) = AT_*\ln(1 + t_*^2). \tag{29}$$

In the same limit, we use Eqs. (20), (27), and (29) to find the dependence of the absorption intensity on the reduced frequency $\omega_* \equiv (\omega - \omega_0)/\omega(a)$,

$$I(\omega_*) \sim \int_0^\infty \cos\left(\omega_* t_* - A\frac{t_*}{1 + t_*^2}\right)\frac{dt_*}{(1 + t_*^2)^\nu}, \qquad \nu \equiv AT_*. \tag{30}$$

One must distinguish between two cases, namely, $A > 1$ and $A < 1$. The strong coupling limit, $A > 1$, is realized in a high electric field given by Eqs. (14) and (27), when only small times t_* are important in the integral in Eq. (30),

$$I(\omega^*) \sim \exp\left\{-\frac{(\omega_* - A)^2}{4\nu}\right\}. \tag{31}$$

In view of determining the reduced quantities T_* and ω_* and the parameter A, one can see that the expressions (31) and (24) for $I(\omega)$ coincide and are independent of the characteristic ripplon frequency $\omega(a)$. Consequently, the specific features of the thin film case being treated do not show up. In a low electric field, $A < 1$, and the shape of the line of optical absorption $I(\omega^*)$ is defined by the value of the parameter $\nu = AT_*$ in Eq. (30). Because $T_* \gg 1$, both limits are realized, namely, $\nu > 1$ and $\nu < 1$,

$$I(\omega_*) \sim \int_0^\infty \cos\omega_* t_* \frac{dt_*}{(1 + t_*^2)^\nu} = K_{\nu - 1/2}(\omega^*)\left(\frac{\omega_*}{2}\right)^{\nu - 1/2}\frac{\Gamma(1/2)}{\Gamma(\nu)}, \tag{32}$$

where $K_{\nu - 1/2}(\omega^*)$ is the cylindrical function and Γ is the gamma function [7].

Asymptotic expansions of $I(\omega_*)$ have the form [7, 8]

$$I(\omega_*) = I(0)\left(1 - \frac{\omega_*^2}{4(\nu - 3/2)}\right), \quad |\omega_*| \ll 1, \tag{33}$$

$$(I(\omega_*) \propto e^{-|\omega_*|})|\omega_*|^{\nu-1}, \quad |\omega_*| \gg \nu. \tag{34}$$

In the case of a high value of the parameter ν ($\nu \gg 1$), the intermediate asymptotic behavior at $1 < |\omega_*| < \nu$ is also realized [8],

$$I(\omega_*) = I(0)\frac{1}{(1+Z^2)^{1/4}}e^{-\chi(Z)\nu}, \tag{35}$$

where the function $\chi(Z)$ is defined by the expression

$$\chi(Z) = (1+Z^2)^{1/2} - 1 - \ln\frac{1+\sqrt{1+Z^2}}{2} \tag{36}$$

and the parameter Z is the ratio of ω_* to ν, $Z = \omega_*/\nu$. A comparison of Eqs. (33) and (36) reveals that the asymptotics $I(\omega_*)$ at $\omega_* \ll 1$ and $1 \ll \omega_* \ll \nu$ are joined, because the expansion of the function $\chi(Z)$ at $Z \ll 1$ has the form

$$\chi(Z) = \frac{1}{4}Z^2\left(1 - \frac{Z^2}{8}\right). \tag{37}$$

It is interesting to follow the transition from the strong coupling limit ($\nu \gg 1$) to the weak coupling limit ($\nu \leq 1$). In a high electric field at high temperatures ($\nu \gg 1$), the shape of the absorption line is almost Gaussian (see Eqs. (35) and (37). The exponential dependence $I(\omega^*)$ as given by Eq. (34) takes place at $|\omega_*| \gg \nu$, when the absorption intensity is very low. At $1/2 < \nu < 3/2$, the quantity $I(\omega_*)$ depends nonanalytically on ω_*: at $\omega^* = 0$, the function $I(\omega^*)$ exhibits a kink. At $\nu < 1/2$, the value of $I(0)$ is undetermined, because the integral in Eq. (32) converges if $\omega_* = 0$. It is apparently this particular case when the absorption line is neither Gaussian nor Lorentzian that is of primary interest. The convergence of the integral in Eq. (32) is provided for either by the inclusion of the natural width of the absorption line or by the inclusion of the interaction of the localized electron with helium vapor [1, 9] whose density is very strongly dependent on temperature. It is interesting that the inclusion of the level width affects only details of the dependence of the central part of the absorption line on the reduced frequency ω_*. The asymptotic behavior of $I(\omega^*)$ at $\omega_* > \nu$ in Eq. (34) is independent of ν, i.e., of temperature and electric field, only because of the pre-exponential factor $|\omega_*|^\nu$. In other words, even in the classical limit of $T_* \gg 1$, the exponential wings of the absorption line are independent of temperature. When treating the experimental data in low electric fields $E_\perp$, an indeterminacy appears in calculating the width of the absorption line. A universally accepted definition of this width exists for Lorentzian or Gaussian lines. In the region of low values of ν, i.e., in low electric fields, this definition may be invalid. The nonmonotonic dependence of the absorption line width on $E_\perp$ has been revealed by Edel'man [9] by the cyclotron resonance, which may be due to the nonanalytic dependence of $I(\omega^*)$ on ω^*at $\nu < 3/2$.

In studying the quantum limit of $T_* \ll 1$, it is convenient to identify the main part, independent of temperature, in expression (28) for the function $J_2(t_*)$:

$$J_2(t_*) = A\frac{t_*^2}{1+t_*^2} + 2A\int_0^\infty \frac{(1-\cos t_* X)e^{-X}}{e^{X/T_*}-1}dX. \tag{38}$$

The integral in Eq. (38) may be represented as the sum

$$\int_0^\infty dx = \sum_{m=1}^\infty \frac{t_*^2 T_*^3}{(T_*+n)[t_*^2T_*^2 + (T_*+n)^2]}. \tag{39}$$

When $T_* \ll 1$, but $t_*T_* \gg 1$, high values of n are important in the summation with respect to n in Eq. (39), and Eqs. (38) and (39) yield

$$J_2(t_*) = A\frac{t_*^2}{1+t_*^2} + \nu\ln(1+t_*^2T_*^2). \tag{40}$$

This expression is also valid in the other limit, $T_*t_* \ll 1$, because the second term in Eq. (40) is in this case unimportant. We use Eqs. (20), (27), and (40) to derive the final expression for the absorption intensity in the quantum limit of $T_* \ll 1$,

$$I(\omega_*) \propto \int_0^\infty \cos\left(\omega_* t_* - A\frac{t_*}{1+t_*^2}\right) \times \exp\left\{-A\frac{t_*^2}{1+t_*^2}\right\}(1+t_*^2T_*^2)^{-\nu}dt_*. \tag{41}$$

The parameter ν is defined above, $\nu = T_*A$.

In the region of strong coupling, $A \gg 1$, low values of t_* are important upon integration in Eq. (41) and the adiabatic approximation

$$I(\omega_*) \propto \exp\left\{\frac{-(\omega_*-A)^2}{4A}\right\} \tag{42}$$

is valid.

In the region of weak coupling, $A \ll 1$, the first term of expansion of $I(\omega_*)$in Eq. (41) in powers of A has the form

$$I(\omega_*) \propto A\left[\frac{T_*}{|\omega_*|}e^{-|\omega_*|/T_*} + e^{-\omega_*}\theta(\omega_*)\right], \tag{43}$$

where $\theta(\omega_*)$ is a step function: $\theta(\omega_*) = 1$ at $\omega_* > 0$ and $\theta(\omega_*) = 0$ at $\omega_* < 0$. Expression (43) does not allow the

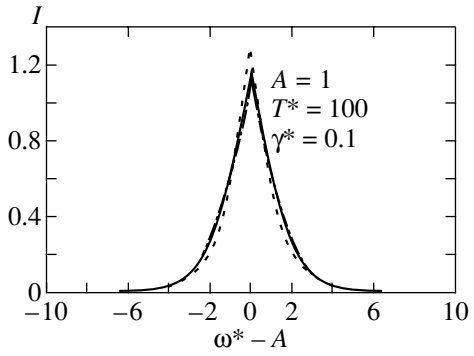

Fig. 1. The absorption line for different values of the helium film thickness. The solid line corresponds to $d/a = 1$, the dotted line corresponds to $d/a = 0.01$, and the dot-and-dash line corresponds to $d/a = 100$. In spite of the very wide range of values of d/a, the absorption lines for different values of d/a differ very little.

literal limit of $T_* = 0$ because, at $\nu = T_* A = 0$, the integral in Eq. (41) diverges. The nonphysical singularity of $I(\omega_*)$ at $\omega_* = 0$ is cut off by $\omega_* \sim \gamma/\omega(a)$, where γ is the natural level width of electron in an excited state with energy ε_1. One can see that, in the quantum limit of weak coupling ($T_* \ll 1$, $A \ll 1$), the absorption line of $I(\omega_*)$ has a maximum at $\omega_* = 0$ and is highly asymmetric. With the threshold energy $\hbar\omega_0 = \varepsilon_1 - \varepsilon_2 + \Delta_0 - \Delta_1$, a change of the absorption mode occurs. At $\omega_* > 0$, the exponential frequency dependence $I(\omega_*)$ is defined by a ripplon emission that is independent of temperature. However, at $\omega_* < 0$, this dependence is related to the absorption of real, i.e., thermal, ripplons, whose density is exponentially low at $T_* \ll 1$. The shift of the electron transition energy, i.e., the difference between $\hbar\omega_0$ and $\varepsilon_1 - \varepsilon_0$, differs from the adiabatic value of Δ from Eq. (25) only by a factor of 3/4,

$$\hbar\omega_0 = \varepsilon_1 - \varepsilon_0 + \Delta',$$
$$\Delta' = \Delta_0 - \Delta_1 = \frac{3F^2}{32\pi\alpha} = \frac{3}{4}\Delta. \quad (44)$$

The parameter $F \sim E_\perp$ (Eq. (14)); therefore, the transition from the case of strong coupling ($A \gg 1$) to weak coupling ($A \ll 1$) brings about a kink on the curve of dependence of the shift of absorption line $\hbar\omega$ on the square of electric field $E_\perp$. One can demonstrate that this result is general and independent of the temperature T_*. In the classical limit ($T_* \gg 1$), the adiabatic approximation is disturbed if $\nu = AT_* < 1/2$, when the integral in Eq. (30) diverges at $\omega_* = 0$. The above-mentioned kink was observed in the experiment with cyclotron resonance in low electric fields $E_\perp$ [9].

We have seen that, in the quantum case of $T_* \gg 1$, the positions of maxima on the absorption line in the strong and weak coupling limits are shifted by A. The intermediate case of $A \approx 1$ is of interest, when both maxima occur on the absorption line. While the first, adiabatic, maximum at $\omega_* = A$ arises as a result of integration in Eq. (41) over the region of $t_* \ll 1$, the second, "antiadiabatic," peak at $\omega_* = 0$ is associated with integration in Eq. (41) over the region of $t_* \gg 1$. Its value is proportional to e^{-A}. At first glance, it might appear that the second (antiadiabatic) peak cannot be observed, because the lifetime of the excited level is always much less than the relaxation time of the dimple. However, no global rearrangement of the dimple occurs during electron transition, with only its minor part changing. This is obvious even from the fact that the difference between the values of dimple energy for different electron levels is much less than the dimple energy proper. A dimple is formed mainly due to long-wave ripplons which deform the surface over scales of the order of $1/\kappa \gg a$. However, the long-wave part of a dimple is almost independent of the electron level number. Electron transitions cause a dimple to vary over scales of the order of a, which occurs much faster than the global rearrangement of the dimple.

4. THICK FILM AND NUMERICAL RESULTS

In this section, we will demonstrate that no qualitative distinction is observed between the cases of thick ($d \gg a$) and thin ($d \ll a$) films for the scope of our investigation. This is due to the similarity between the $\omega(q) \propto q^2 (d \ll a)$ and $\omega(q) \propto q^{3/2} (d \gg a)$ spectra. We will also give the results of numerical calculation by formulas (20), (22), and (2), which illustrate the foregoing conclusions.

Figure 1 gives the frequency dependence of $I(\omega^*)$ in the classical limit ($T_* = 100$) at $\nu = 1$ ($A = 0.01$) for different values of the helium film thickness, $d/a = 0.01$, 1, 100. In order to preclude the emergence of nonphysical singularities in calculations, the cut-off factor $e^{-\gamma t_*}$ is introduced in formula (20), where $\gamma = 0.02$. One can see that the absorption spectrum looks not at all Gaussian or Lorentzian, and the variation of the helium film thickness by four orders of magnitude fails to result in qualitative differences in $I(\omega_*)$. The plots are constructed on the basis of numerical count by formulas (20), (22), and (2).

In order to analyze the shape of the absorption line, its dependence on ω_* is given on a logarithmic scale (Fig. 2) for different values of $\nu = 0.5$, 1, 2. The temperature T_* was kept constant ($T_* = 100$), with the coupling constant A being varied. This corresponds to the experimental situation in which the pressing electric field is varied, with the temperature T maintained constant [9]. Portions of the linear dependence $\ln I(\omega_*)$ are observed in Fig. 2; these portions are the longer, the lower the values of ν. As ν increases, the dependence

$\ln I(\omega_*)$ changes to an inverted parabola, which corresponds to the Gaussian shape of the absorption line. On the contrary, as ν decreases, the dependence of I on ω_* comes to ever better satisfy the Urbach rule (see Eq. (34)) and the sharp peak at $\omega_* = 0$ becomes more explicit. The plot is constructed on the basis of numerical count by formulas (20), (22), and (2) for $d/a = 100$.

Figure 3 illustrates the quantum limit of $T_* < 1$, when the emergence of two maxima on the absorption line is possible. For this limit to be experimentally realized, the temperature must be reduced to $T \sim 10^{-2}$ K. The fact of the presence of two maxima does not depend on whether the helium film being treated is thin or thick. These two maxima are most pronounced at $A \approx 2$ for a thick helium film and at $A \approx 3$ for a thin helium film.

5. DISCUSSION OF THE RESULTS

We have demonstrated the validity of the exactly solvable independent boson model [5] for describing optical transitions between states of an electron localized on the liquid helium surface. The results obtained within this model are used to analyze the shape, width, and shift of the absorption line during such transitions. We have demonstrated that, even in the classical limit ($T^* \gg 1$), the absorption line in the region of weak coupling $\nu = AT_* < 1$ is neither Gaussian nor Lorentzian. In a wide range, $\omega_* > \nu$, the Urbach rule is valid (Eq. (34)).

Note that the exponential dependence of absorption on the electron transition energy $\hbar(\omega - \omega_0)$ is typical of numerous optical experiments, but it was never substantiated theoretically [10]. In this case, we are dealing with the Urbach law rather than with the empirical Urbach rule. If $\nu < 1$, the absorption line $I(\omega_*)$ has a peak at $\omega_* = 0$, which is the sharper, the lower the value of ν.

In the quantum limit in the region of weak coupling, the absorption line is highly asymmetric. Its shape corresponds to the Urbach rule (see Eq. (43). In the $A \sim 1$ region, the absorption line has two maxima (see Fig. 3). The shift of the electron transition energy is determined, which is associated with the deformation of the helium surface (Eqs. (2), (25), and (44)).

A detailed experimental study of the phenomena treated here for the case of cyclotron resonance of electrons on the helium surface was performed by Edel'man [3, 9] for $\varepsilon_1 - \varepsilon_0 = \hbar\omega_H$ (where ω_H is the cyclotron frequency). In high magnetic fields ($\hbar\omega_H \gg T$), electrons populate only one Landau level. Unfortunately, we are not ready to compare our results with the experimental results of [3, 9]. The thing is that the problem on cyclotron resonance of electrons on the surface of liquid helium does not reduce to the "trivial" problem of interaction between a two-level electron system and ripplons. This problem is essentially complicated by the infinitely multiple degeneracy of Landau levels in a high magnetic field. Rather than interpreting the results of old experiments [3, 9], it is our objective to provoke new experiments that are simpler from a theoretical standpoint.

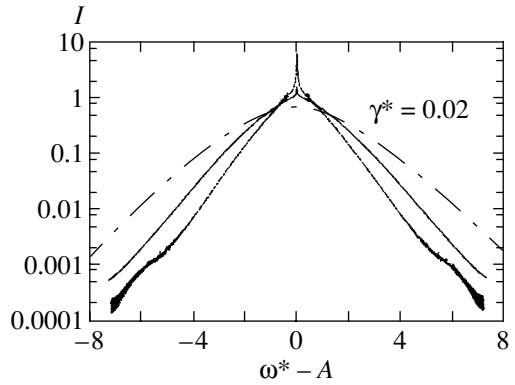

Fig. 2. The intensity of absorption on a logarithmic scale as a function of reduced frequency $\omega^* \equiv (\omega - \omega_0 - \Delta)/\omega(a)$. At a constant temperature $T^* = 100$, the coupling constant varies. The solid line corresponds to $\nu = 1$ ($A = 0.01$), the dot-and-dash line corresponds to $\nu = 2$, and the dotted line corresponds to $\nu = 0.5$ ($A = 0.005$). For a low value of ν, one can clearly see the linear portions of the dependence $\ln I(\omega^*)$ (Urbach rule, Eq. (34)). If $\nu < 1$, a sharp peak at $\omega^* = 0$ is also observed.

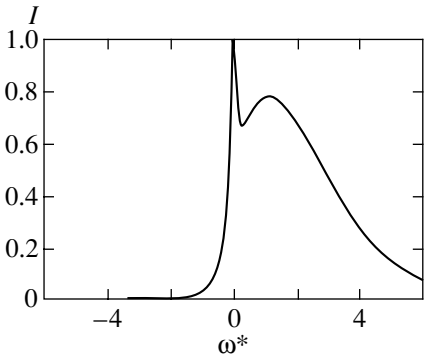

Fig. 3. The absorption line in the quantum limit: $T^* = 0.2$, $d/a = 100$. One can clearly see two maxima.

Optical transitions between the levels of diplons in the vicinity of the helium surface may also be regarded as an interesting realization of the phenomena being treated. A diplon is a bound state of an electron in the field of a He^+ ion located beneath the helium surface. The coupling energy of a diplon may be varied in a wide range by varying the pressing electric field [1].

Another possible application of the obtained results is in electronic devices coated with a film of liquid helium. Because all impurities are frozen out from helium, it is possible to realize small-sized ideally pure electronic systems using the method suggested in [11]. These are annular electronic structures, quantum dots, and electron filaments.

ACKNOWLEDGMENTS

In conclusion, it is a pleasure to express our gratitude to V.B. Shikin and A.S. Ioselevich for discussion of our results.

This study received financial support from the Russian Foundation for Basic Research (project no. 00-02-17729).

REFERENCES

1. V. B. Shikin and Yu. P. Monarkha, *Two-dimensional Charged Systems in Helium* (Nauka, Moscow, 1989).
2. G. A. Farias and F. M. Peeters, Phys. Rev. B **55**, 3763 (1997).
3. V. S. Édel'man, Pis'ma Zh. Éksp. Teor. Fiz. **24**, 510 (1976) [JETP Lett. **24**, 468 (1976)].
4. P. D. Grigor'ev, Pis'ma Zh. Éksp. Teor. Fiz. **66**, 630 (1997) [JETP Lett. **66**, 630 (1997)].
5. G. D. Mahan, *Many-Particle Physics* (Plenum, New York, 1990).
6. A. Cheng and P. M. Platzman, Solid State Commun. **25**, 813 (1978).
7. I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products* (Nauka, Moscow, 1971; Academic, New York, 1980).
8. *Higher Transcendental Functions (Bateman Manuscript Project)*, Ed. by A. Erdelyi (McGraw-Hill, New York, 1953; Nauka, Moscow, 1973), Vol. 1.
9. V. S. Édel'man, Zh. Éksp. Teor. Fiz. **77**, 673 (1979) [Sov. Phys. JETP **50**, 338 (1979)].
10. B. I. Shklovskiĭ and A. L. Efros, *Electronic Properties of Doped Semiconductors* (Nauka, Moscow, 1979; Springer-Verlag, New York, 1984), pp. 365, 366.
11. A. M. Dyugaev, A. S. Rozhavskiĭ, I. D. Vagner, and P. Wyder, Pis'ma Zh. Éksp. Teor. Fiz. **67**, 410 (1998) [JETP Lett. **67**, 434 (1998)].

Translated by H. Bronstein

Journal of Experimental and Theoretical Physics, Vol. 93, No. 1, 2001, pp. 111–116.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 1, 2001, pp. 127–133.

SOLIDS
Structure

An EPR Study of the Temperature Dependence of the Energy Gap in Ytterbium Dodecaboride

T. S. Altshuler[a]*, A. E. Altshuler[a], and M. S. Bresler[b]

[a]*Kazan Physicotechnical Institute, Russian Academy of Sciences, Kazan, Tatarstan, 420029 Russia*
[b]*Ioffe Physicotechnical Institute, Russian Academy of Sciences, St. Petersburg, 194021 Russia*
**e-mail: tatiana@dionis.kfti.kcn.ru*

Received November 14, 2000

Abstract—An EPR study of ytterbium dodecaboride (YbB_{12}) showed the presence of an energy gap with a width of $2\Delta = 12$ meV in the energy spectrum of this Kondo insulator. The temperature dependence of the energy gap was determined by interpreting the experimental data within the framework of the exciton dielectric model: $\Delta(T) = 72$ K at an absolute zero and $\Delta(T) = 0$ at ~115 K. The temperature dependence of the EPR linewidth exhibits a feature at 13–15 K, which is indicative of a finite density of states inside the gap. This can be related to the presence of impurity states or bound polaron excitations in the electron spectrum of YbB_{12}.

1. INTRODUCTION

The group of Kondo insulators (Kondo semiconductors) possessing a narrow gap in the energy spectrum have attracted the attention of researchers for several decades by exhibiting certain features in the ground state and unusual properties at low temperatures. A classical object for investigations in this field is samarium hexaboride (SmB_6), in which samarium possesses an average valence of 2.6. However, in recent years, attention has also been drawn to ytterbium dodecaboride (YbB_{12})—another compound with intermediate valence. The valence of an ytterbium ion is closer to an integer, amounting to 2.9. Despite being probed by a number of experimental techniques, YbB_{12} is yet studied to a much lower extent as compared to SmB_6. Available data refer to the electrical properties [1, 2], magnetic susceptibility [1, 2], heat capacity [3], and photoconductivity in the far IR range [4]. Also studied were the spectra of inelastic neutron scattering [5, 6], photoemission [7], and nuclear magnetic resonance [8].

In the papers cited above, the experimental data were interpreted based on the concept of intermediate valence of ytterbium ions. It was assumed that the spectrum of electron states contains a 10–25 meV energy gap opened at temperatures below 70 K.

The unusual low-temperature properties of Kondo insulators were described for a long time within the framework of a model of the *f–d* hybridization gap partly renormalized by the correlation effects [9]. However, this model failed to explain some fine details in the behavior of YbB_{12} at low temperatures. More involved theories of the ground state of Kondo insulators were developed by Kikoin and Mishchenko [10], Curnoe and Kikoin (exciton–polaron model) [11], and Kasuya (Wigner crystallization or Wigner liquid model) [12].

At present, the nature of the ground state of Kondo insulators cannot be considered as completely clear. In order to select between the models available, it is necessary to employ the whole variety of the available experimental methods. Among these, a very useful tool is offered by the electron paramagnetic resonance (EPR). The EPR measurements make it possible to establish the presence of an energy gap in the spectrum of electron excitations and to study the temperature dependence of the gap width (in order to judge on the nature of the ground state of this Kondo insulator). Previously [13] this program was fulfilled in the case of SmB_6—a classical object in the physics of Kondo insulators (Kondo semiconductors). Study of the EPR effect in SmB_6 led to the discovery of a dynamic and static Jahn–Teller effects in this material [14, 15], which is evidence in favor of the exciton–polaron model of the ground state of SmB_6.

The purpose of this study was to apply the EPR method to YbB_{12}. Based on the exciton dielectric model, the EPR data were used to determine the temperature dependence of the correlation gap width in YbB_{12}. We observed a finite density of electron states inside the gap, which is probably related to additional excitations in the electron spectrum that also possess a collective character.

The results of this study were preliminarily partly published [16] and reported at the All-Russia Conference (NT-32) [17].

2. EXPERIMENTAL

The EPR measurements were performed using powdered samples of YbB_{12}, both pure and doped with Gd^{3+} (spin marker) to a concentration of $c = 0.1$ at.%

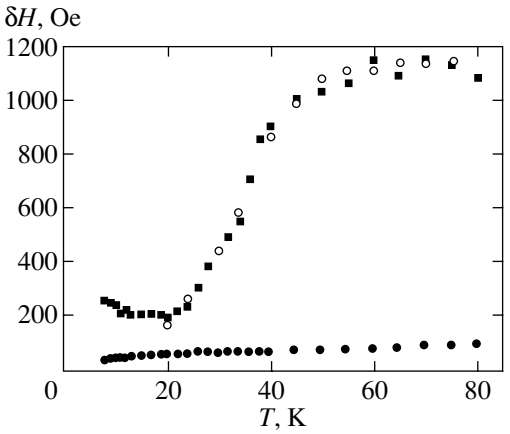

Fig. 1. Temperature dependence of the EPR linewidth for Gd^{3+} ions (c = 1 at.%) in (▪) YbB_{12} and (•) LuB_{12}; open circles show the results of theoretical calculations.

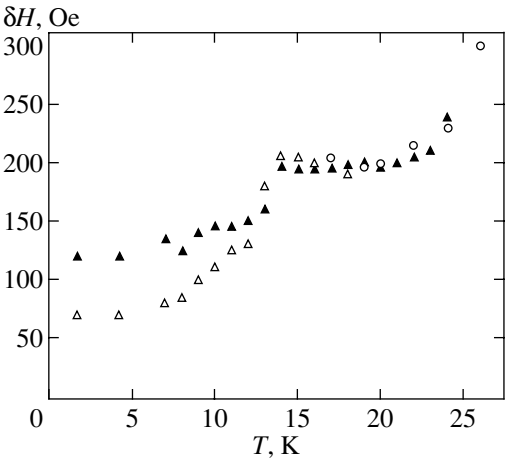

Fig. 2. Temperature dependence of the EPR linewidth for Gd^{3+} ions in YbB_{12} in the low-temperature region for the Gd^{3+} ion concentrations c = 0.1 (▵), 0.5 (▲), and 1.0 at.% (○).

($Yb_{0.999}Gd_{0.001}B_{12}$) 0.5 at.% ($Yb_{0.995}Gd_{0.005}B_{12}$), or 1 at.% ($Yb_{0.99}Gd_{0.01}B_{12}$). For comparison, we also studied a sample of lutetium dodecaboride ($Lu_{0.99}Gd_{0.01}B_{12}$) doped with Gd^{3+} to a concentration of 1 at.%. The YbB_{12} compound was synthesized in an induction furnace at 1700 K by the barometric reduction of Yb_2O_3 in vacuum. Then the compound was melted in an arc furnace and dissolved in nitric acid in order to remove residual YbB_6. A single-phase sample, appearing as a black powder, was checked by X-ray diffraction. An analogous procedure was used for the synthesis of lutetium dodecaboride. The powder grain size was 10–20 μm, which is smaller than the skin-layer thickness in both YbB_{12} (semiconductor) and LuB_{12} (metal). In order to provide for a better electromagnetic field penetration into the materials, the powders were dispersed in melted paraffin. The measurements were performed at a frequency of ν = 9.4 GHz in the temperature interval from 1.7 to 80 K.

Rare-earth dodecaborides possess simple cubic lattices of the NaCl type, with the metal atoms occupying Na positions and a cubic octahedron of 12 boron atoms occupying Cl positions [18]. All such dodecaborides, except for YbB_{12}, exhibit a metal conductivity. The rare-earth (R) ions in RB_{12} compounds are usually trivalent. In the crystal field of a cubic symmetry, the multiplet of $Yb^{3+}(4f^{13}, {}^8F_{7/2})$ splits into the Γ_6 doublet, Γ_8 quartet, and Γ_7 doublet. Under these conditions, the ground state of RB_{12} may represent a Kramers doublet Γ_6 or Γ_7, with the EPR signal observed from both of these. However, we did not detect an EPR signal from pure YbB_{12}. This can be explained by the metal valence fluctuating between the states of Yb^{3+} and Yb^{2+} at a frequency exceeding that of the EPR spectrometer ($\nu_{Yb} > 10^{10}$ Hz), which confirms that the metal actually occurs in the state of intermediate valence.

All the powdered samples of compounds doped with gadolinium exhibited intense EPR signals. Figure 1 shows the temperature dependence of the EPR linewidth $\delta H(T)$ for $Yb_{0.99}Gd_{0.01}B_{12}$ and $Lu_{0.99}Gd_{0.01}B_{12}$. As was noted above, the latter compound possesses a metal conductivity. The EPR spectrum of $Lu_{0.99}Gd_{0.01}B_{12}$ contained a signal from Gd^{3+}, which was characterized by a linear temperature dependence (typical of metals) of the linewidth $\delta H = a + bT$ with the slope $b = \delta H/\delta T$ = 1.25 Oe/K. The same linewidth in the spectrum of $Yb_{0.99}Gd_{0.01}B_{12}$ showed a more complicated variation with the temperature: below 10 K, the line exhibited a slight broadening; above 20 K, the linewidth almost exponentially increased with the temperature; in the region of 50 K, the linewidth exhibited saturation. As the temperature further increases, the EPR signal intensity drops and above 80 K the linewidth cannot be determined: the error of the linewidth determination increases from ±10 Oe at 15 K up to ±150 Oe at 70 K. In the samples with lower Gd content (Fig. 2), the EPR signal was observed below 17 K (c = 0.1 at.%) and below 25 K (c = 0.5 at.%). The residual linewidth (δH at T = 0) increased with the Gd concentration, but this concentration dependence was not manifested when the temperature increased to 13–14 K or above.

The δH values for all three Gd concentrations studied fit to the same curve. This behavior implies that the EPR linewidth at T > 14–15 K is determined by purely relaxational effects. It is important to note that the temperature dependence of δH in the samples studied exhibited a clearly pronounced singularity (a kink) at T = 13– 14 K.

Figure 3 shows the temperature variation of the g value for the signals of Gd^{3+} ions in LuB_{12} and YbB_{12}. Gadolinium ions (Gd^{3+}) possess magnetism of the purely spin type (ground state: $4f^7$, ${}^8S_{7/2}$), and hence, their g value must be close to 2.00. The spectra of Gd-doped LuB_{12} (Korringa metal) samples actually contain a signal with

the temperature-independent g value close to 2.00 ($g_{Lu} = 1.990 \pm 0.05$).

In the spectra of YbB_{12}, behavior of the g value of gadolinium ions at low temperatures (<12 K) is completely correlated with the linewidth variation. As the Gd concentration decreases, both the g value and the linewidth become less dependent on the temperature. This correlation is explained by the ferromagnetic ordering of Gd ions at $c = 1$ at. %.

At $T > 12$ K, the g value depends neither on the temperature nor on the Gd content in the samples and is equal to 1.945 ± 0.015. As is seen, the g value in YbB_{12} exhibits a rather strong shift ($\delta g_{Yb} = -0.045$) as compared to that in usual metals ($\delta g_{Lu} = -0.01$ in $Lu_{0.99}Gd_{0.01}B_{12}$). As is well known, the density of states $N(\varepsilon_F)$ at the Fermi level is significantly higher in Kondo insulators (due to the s–f hybridization) than in usual metals. Since $\delta g \sim N(\varepsilon_F)$, this shift is indicative of the intermediate valence manifestations in YbB_{12}.

Indeed, there is a close correlation between the average valence of the rare-earth ion and the shift of its g value in Kondo insulators: for Yb in YbB_{12}, the average valence is 2.90 and the g value shift is $\delta g = -0.045$, whereas for Sm in SmB_6, the corresponding values are 2.65 and $\delta g = -0.080$ [19]. In addition to this fact, the existence of intermediate valence in YbB_{12} is confirmed by a "semiconductor" character of the temperature variation of the EPR linewidth $\delta H(T)$. This curve is similar to that observed in SmB_6 (a classical Kondo insulator [19]), differing sharply from a linear dependence $\delta H(T)$ (typical of metals) observed in dodecaborides of other rare-earth elements. Finally, as noted above, the absence of the EPR signal from Yb^{3+} ions also serves an additional sign of a Kondo insulator.

3. DISCUSSION OF RESULTS

Ytterbium dodecaboride YbB_{12}, in contrast to the isostructural compound LuB_{12} possessing a metal conductivity, is a cubic Kondo insulator with a narrow energy gap (Kondo semiconductor). The temperature dependence of the EPR linewidth for Gd^{3+} ions in these compounds are sharply different (Fig 1). In LuB_{12}, $\delta H(T)$ is a linear function showing a relatively small Korringa relaxation (typical of metals). In YbB_{12}, the $\delta H(T)$ curve reveals exponential line broadening in the temperature interval from 14 to 80 K, which is related to the presence of an energy gap in the spectrum of electron excitations of this semiconductor compound. This gap may arise, for example, as a result of hybridization between s electrons of the conduction band and f electrons of the valence band of Yb. Interaction of the spin of Gd^{3+} ions with the spin of electrons and holes leads to broadening of the EPR linewidth. Since the level of the s–f excitation increases exponentially with the temperature, the $\delta H(T)$ curve also exhibits an exponential character.

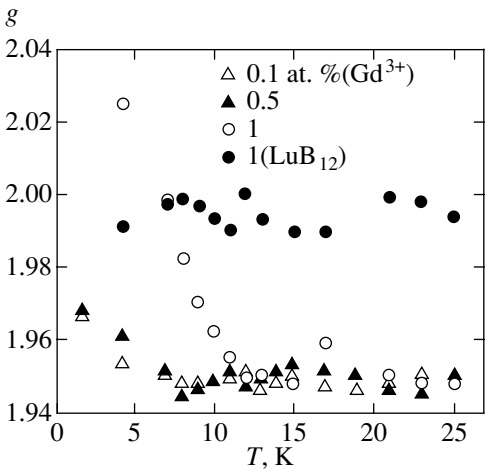

Fig. 3. Temperature dependence of the g value for Gd^{3+} ions in YbB_{12} and LuB_{12}.

However, there are different points of view concerning the origin of the energy gap in Yb. In particular, the appearance of this gap was explained by the s–f hybridization [9], the Wigner crystallization [12], and the exciton coupling of d electrons with f holes [13]. There are significant discrepancies between theoretical results in the region of low temperatures. Indeed, the "hybridization" gap is not temperature-dependent, while the "exciton" gap appears as a collective effect and disappears at temperatures on the order of the gap width. Calculations performed within the framework of the s–f hybridization model with a constant gap width did not provide adequate description of our experimental results.

Since no detailed theory explaining the ground state and the unusual low-temperature properties of YbB_{12} have been elaborated on the level comparable to that accessible for SmB_6, our considerations will be based on the theory developed for samarium hexaboride.

Apparently, the unusual properties of SmB_6, including anomalies in the phonon spectrum [20], the dispersion of magnetic excitations studied by inelastic neutron scattering [21], and the optical absorption and dispersion in the submillimeter wavelength range [22] (as well as the transport properties [23, 24]), are most adequately described by the exciton–polaron model of Kikoin and Mishchenko [10, 11]. The basic concept of the model is that the ground state of the system is a superposition of the f^6 state corresponding to Sm^{2+} ion and the f^5p state corresponding to an intermediate-coupling exciton, comprising a hole in the samarium f shell (Sm^{3+}) and an electron on the orbit representing a linear combination of the p states of boron atoms surrounding (in the first coordination sphere) samarium. On the whole, the symmetry of this linear combination is the same as that of the hole state. The exciton occurs in a singlet (i.e., nonmagnetic) state. The valence fluctuations are essentially the quantum beats between the two system states described above. A correlation in the state of excitons at various lattice sites is probably established

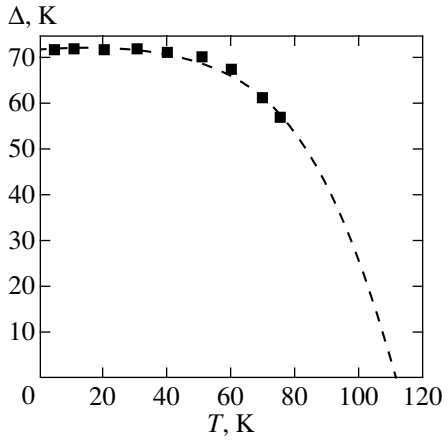

Fig. 4. Temperature dependence of the energy gap width $\Delta(T)$ in YbB_{12}.

as a result of the exciton interaction (not arising in the mean-field approximation). This interaction leads to a macroscopically coherent state. The exciton formation leads to the appearance of a gap in the electron spectrum. Since the valence fluctuates at a "phonon" frequency (10^{12}–10^{13} Hz), it is not surprising that these fluctuations are coupled to the lattice oscillations. This coupling leads to "softening" of the phonon modes and renormalization of the electron (exciton) states, which results in a mixed exciton–polaron state formation. The phonon modes mix both with charge excitations (polarons) and spin excitations (spin-polarons). The latter are related to the Jahn–Teller effect observed in SmB_6 [14, 15] (see also [25]).

Unfortunately, the spin relaxation rate in a Kondo insulator was not calculated within the framework of the Kikoin–Mishchenko model. For this reason, we had to interpret our experimental data using the results of calculations performed by Khaliulin and Khomskii [13] based on an exciton dielectric model with large-size (Mott–Wannier) excitons composed of a *d* electron and an *f* hole. The Khaliulin–Khomskii calculations followed in the general part the electron pairwise correlation model for superconductors. Naturally, this theory (as well as that for superconductors) led to the appearance of a temperature-dependent energy gap in the electron spectrum of the exciton dielectric. This temperature dependence can be determined by the comparison with experiment.

According to this theory, the spin relaxation rate is described by the formula [13]:

$$T_2^{-1} = 2\pi T f(\Delta)(b_d^2 + b_f^2) \times \left\{1 + \alpha[1 - f(\Delta)]\frac{\Delta}{2T}\ln 2\Delta\tau\right\},$$

$$f(\Delta) = \left(1 + \exp\frac{\Delta}{T}\right)^{-1}, \quad b_i = J_i N_i,$$

$$\alpha = \frac{(b_d + b_f)^2}{b_d^2 + b_f^2},$$

where $\Delta = \Delta(T)$ is the exciton gap; J_d and J_f are the integrals of the exchange interaction of Gd with *d* electron and *f* hole; N_d and N_f are the corresponding densities of states at the Fermi level in YbB_{12}; and τ is the momentum relaxation time (characteristic correlation loss time).

We obtained a satisfactory theoretical description of the experimental T_2^{-1} values (see the results of calculations in Fig. 1) for the following values of parameters:

$$b_d = -0.701 \times 10^{-2}, \quad b_f = -1.902 \times 10^{-2},$$

$$\alpha = 1.65, \quad \tau = 1.00\ \mathrm{K}^{-1}.$$

Although the quantities b_d, b_f, and τ are independent fitting parameters, these values must be also consistent with the other experimental data. As is known, the πb_d^2 value is on the order of the temperature slope of the EPR linewidth for Gd in LuB_{12}: $\pi b_d^2 = 1.54$ Oe/K and $\delta H/\delta T(\mathrm{exp.}) = 1.25$ Oe/K. A shift of the *g* value in Kondo insulators is proportional to the sum of b_d and b_f:

$$\delta g = b_d + \gamma b_f, \quad \gamma = \left[\frac{4}{3}J(J+1)\right]^{1/2},$$

$$\delta g = -0.09, \quad \delta g_{\mathrm{exp}} = -0.045.$$

As is seen from this comparison, the theory agrees quite well with the experiment. Probably, a difference between the effects of the intermediate-coupling excitons and the Mott–Wannier excitons on the Gd^{3+} spin relaxation rate T_2^{-1} is not significantly manifested in the comparison of theory and experiment.

Figure 4 shows the temperature dependence of the energy gap width in YbB_{12} determined from the results of our experiments. As is seen, the gap width $\Delta(T)$ at the lowest temperatures amounts to 72 K. This value remains almost unchanged when the temperature increases up to 40 K, but then decreases and, probably, completely disappears at 115 K. The full gap width of $2\Delta = 140$ K is close to $T_{cr} \approx 115$ K, in agreement with theoretical predictions. Our data ($2\Delta = 12$ meV) can be compared to the experimental vales obtained from the photoelectron emission [7] and photoconductivity [4] measurements in YbB_{12}. In the photoemission spectrum [7], the gap with a width of 10 meV appeared below 75 K (on approaching from higher temperatures). In the photoconductivity response [4], the gap had a width of about 25 meV and appeared at temperatures below

70 K (the optical gap can be greater than the transport gap, since the main contribution in the optical absorption is due to direct transitions, whereas the minimum gap can be due to indirect transitions in the momentum space). Thus, our results are in satisfactory agreement with the data obtained by independent methods.

Note a kink in the temperature dependence of the EPR linewidth at low temperatures (Fig. 2). This feature is observed for two levels of doping YbB_{12} samples with gadolinium. The kink is not distinguished in the $\delta H(T)$ curve for a sample containing 1 at.% Gd because of the aforementioned ferromagnetic ordering. It is interesting to note that a similar kink was observed for a Gd-doped SmB_6: the feature occurred in the same temperature interval and was independent of the impurity concentration. Apparently, an increase in the linewidth at a temperature of 13-14 K can be related to the presence of a density of electron states inside the correlation gap and, hence, is of considerable interest. The most simple explanation consists in the presence of impurity states in the gap. However, this assumption disagrees with the fact that the observed feature is independent of the dopant (Gd) concentration. A more interesting hypothesis is that this density of states is related to certain features of the ground state of the system, for example, to the formation of a bound polaron analogous to that considered by Curnoe and Kikoin [11]. The states inside the energy gap of SmB_6 were also reported by Sluchanko *et al.* [24], where the appearance of these states was related to a correlation between the exciton–polaron complexes formed at different lattice sites. Nyhus *et al.* [25] observed magnetic excitations in SmB_6 with the energies falling inside the energy gap of this Kondo insulator. These excitations were explained by interaction of the ground sate of Sn^{3+} with the mixed modes of phonons and valence fluctuations (the Jahn–Teller effect). However, the available experimental data are insufficient to make any final judgments: the question concerning the nature of the observed additional excitations still remains open, offering a subject for further detailed investigation.

In addition, the ground state of the Kondo insulator was considered, besides the exciton dielectric model, within the framework of the Wigner crystallization model, also leading to the correlation gap formation in the electron excitation spectrum [12]. This model was criticized by Curnoe and Kikoin [11] in application to SmB_6. One of the arguments against the Kasuya model was that the state introduced is incompletely symmetric, which must lead (at a sufficiently low temperature) to a ferroelectric ordering not observed in experiment; in addition, the ground state in the Kasuya model is not homogeneous and implies a charge ordering of the impurity ions closest to Sm^{2+} and Sm^{3+}, which has also not yet been confirmed in experiment.

The results of our EPR measurements showed that the energy gap in the Kondo insulators YbB_{12} (this work) and SmB_6 [13] has a collective rather than a single-particle character (such as in the case of a simple *s–f* hybridization). Unfortunately, based on the EPR data, it is impossible to judge between the exciton dielectric and Wigner crystallization models. Nevertheless, we may ascertain that the gap width variation is satisfactorily described within the framework of the exciton dielectric model. Taking into account the absence of experimental evidence for the Wigner crystallization, the exciton dielectric model seems to be advantageous.

4. CONCLUSION

1. We have measured the temperature dependence of the EPR linewidth and g value of a Kondo insulator YbB_{12} doped with Gd^{3+} ions.

2. The temperature variation of the energy gap width in the electron spectrum of YbB_{12} and the Kondo constants for electrons and holes (products of exchange integrals by the density of states in the corresponding bands) were determined within the framework of the exciton dielectric model.

3. The results agree with the data on the energy gap width and the temperature of the gap vanishing obtained by methods of photoemission spectroscopy, inelastic neutron scattering, and photoconductivity in the far IR range.

4. An additional EPR line broadening at a temperature of 13–14 K was observed, which is indicative of the presence of as finite density of states inside the correlation gap. This effect may be caused either by the impurity levels appearing as a result of doping or by the coupled polaron excitations characteristic of the ground state of the system studied.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research, project no. 00-02-16080.

REFERENCES

1. T. Takabatake, F. Iga, T. Yoshino, *et al.*, J. Magn. Magn. Mater. **177-181**, 277 (1998).
2. F. Iga, N. Shimizu, and T. Takabatake, J. Magn. Magn. Mater. **177-181**, 337 (1998).
3. F. Iga, S. Hiura, J. Klijn, *et al.*, Physica B (Amsterdam) **259-261**, 312 (1999).
4. H. Okamura, S. Kimura, H. Shinozaki, *et al.*, Phys. Rev. B **58**, R7496 (1998); H. Okamura, S. Kimura, H. Shinozaki, *et al.*, Physica B (Amsterdam) **259-261**, 317 (1999).
5. A. Bouvet, T. Kasuya, M. Bonnet, *et al.*, J. Phys.: Condens. Matter **10**, 5667 (1998); F. Iga, A. Bouvet, L. P. Regnault, *et al.*, J. Phys. Chem. Solids **60**, 1193 (1999).
6. E. V. Nefeodova, P. A. Alekseev, J.-M. Mignot, *et al.*, Phys. Rev. B **60**, 13507 (1999); E. V. Nefeodova, P. A. Alekseev, J.-M. Mignot, *et al.*, Physica B (Amsterdam) **276-278**, 770 (2000).
7. T. Suzaki, Y. Takeda, M. Arita, *et al.*, Phys. Rev. Lett. **82**, 992 (1999).

8. K. Ikushima, Y. Kato, M. Takigawa, *et al.*, Physica B (Amsterdam) **281-282**, 274 (2000).
9. G. Aeppli and Z. Fisk, Comments Condens. Matter Phys. **16**, 155 (1992).
10. K. A. Kikoin and A. S. Mishchenko, Zh. Éksp. Teor. Fiz. **104**, 3810 (1993) [JETP **77**, 828 (1993)]; K. A. Kikoin and A. S. Mishchenko, J. Phys.: Condens. Matter **7**, 307 (1995).
11. S. Curnoe and K. A. Kikoin, Phys. Rev. B **61**, 15714 (2000).
12. T. Kasuya, J. Phys. Soc. Jpn. **65**, 2548 (1996).
13. T. S. Al'tshuler, G. G. Khaliullin, and D. I. Khomskii, Zh. Éksp. Teor. Fiz. **90**, 2104 (1986) [Sov. Phys. JETP **63**, 1234 (1986)].
14. H. Sturm, B. Elschner, and K. H. Hoeck, Phys. Rev. Lett. **54**, 1291 (1985); C. Weber, E. Sigmund, and M. Wagner, Phys. Rev. Lett. **55**, 1645 (1985).
15. T. S. Al'tshuler and M. S. Bresler, Zh. Éksp. Teor. Fiz. **115**, 1860 (1999) [JETP **88**, 1019 (1999)].
16. A. E. Altshuler and T. S. Altshuler, Physica B (Amsterdam) **281-282**, 276 (2000).
17. A. E. Al'tshuler, T. S. Al'tshuler, and M. S. Bresler, in *Proceedings of the All-Russia Conference NT-32, Kazan, 2000.*
18. F. Bertaut and P. Blum, C. R. Hebd. Seances Acad. Sci. **34**, 666 (1949).
19. T. S. Al'tshuler, V. N. Mironov, G. G. Khaliullin, and D. I. Khomskii, Pis'ma Zh. Éksp. Teor. Fiz. **40**, 28 (1984) [JETP Lett. **40**, 754 (1984)].
20. P. A. Alekseev, A. S. Ivanov, B. Dorner, *et al.*, Europhys. Lett. **10**, 457 (1989).
21. P. A. Alekseev, J.-M. Mignot, J. Rossat-Mignot, *et al.*, J. Phys.: Condens. Matter **7**, 289 (1995).
22. B. Gorshunov, N. Sluchanko, A. Volkov, *et al.*, Phys. Rev. B **59**, 1808 (1999).
23. N. E. Sluchanko, A. A. Volkov, V. V. Glushkov, *et al.*, Zh. Éksp. Teor. Fiz. **115**, 970 (1999) [JETP **88**, 533 (1999)].
24. N. E. Sluchanko, V. V. Glushkov, B. P. Gorshunov, *et al.*, Phys. Rev. B **61**, 9906 (2000).
25. T. P. Nyhus, S. L. Cooper, Z. Fisk, and J. Sarrao, Phys. Rev. B **55**, 12488 (1997).

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 93, No. 1, 2001, pp. 117–120.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 1, 2001, pp. 134–138.

SOLIDS
Structure

The Effect of a Weak Magnetic Field on the Mobility of Dislocations in Silicon

A. A. Skvortsov*, A. M. Orlov, and L. I. Gonchar
Ul'yanovsk State University, Ul'yanovsk, 432700 Russia
**e-mail: scvor@sv.uven.ru*
Received December 5, 2000

Abstract—The magnetoplastic effect in dislocation silicon is discovered. It is shown that in the presence of tensile stresses (up to 20 MPa), the mechanically activated path of surface dislocation half-loops is limited mainly by the dynamics of defects in various slip systems relative to the applied load. The activation barriers for the motion of dislocations controlled by various conditions in the temperature range T = 850–950 K are $E_{aF} = 2.1 \pm 0.1$ eV and $E_{aS} = 1.8 \pm 0.1$ eV. An increase in the path of surface dislocation half-loops and a change in the activation barriers are detected ($E_{aF} = 1.4 \pm 0.1$ eV and $E_{aS} = 1.6 \pm 0.1$ eV) after subjecting silicon to a magnetic field (B = 0.7 T) for 30 min. Possible reasons behind the observed effects are discussed. © *2001 MAIK "Nauka/Interperiodica".*

An analysis of plastic properties of condensed media in an external magnetic field revealed [1] the existence of a magnetosensitive response of the dislocation structure of ionic crystals. In this and subsequent publications [2–5], the emergence of the magnetoplastic effect is attributed to the detachment of dislocations from paramagnetic impurities as a result of spin-dependent transitions in the system formed by a dislocation and an impurity center, followed by the relaxation of the defect structure. In ionic crystals [2, 3] and metals [4, 5], this is manifested in a change in the mobility of individual dislocations after the exposure to a magnetic field. As regards semiconductors, the number of publications devoted to magnetosensitive interactions in these materials is much smaller [6–9]. The information obtained by us [9] indicates the effect of a weak magnetic field on the acoustic emission of silicon, reflecting the dislocation-type rearrangement of the internal structure of the semiconductor, although this has not been confirmed yet by direct visual observations. The present work aims at filling this gap.

In our experiments, we used *n*-type silicon wafers of diameter 76 mm and resistivity ρ = 0.01–0.1 Ω m, which were oriented in the [111] direction. Cutting these wafers along $[\bar{1}10]$, $[11\bar{2}]$, and [111], we obtained samples having a size of $30 \times 10 \times 0.5$ mm, respectively, in these directions. Dislocations were introduced according to the method developed in [10, 11] by scratching the polished surface of silicon by a diamond tip along the $[\bar{1}10]$ direction. For this purpose, up to six parallel scratches were made on all the samples cut from the nearest regions of the same wafer. The separation between these scratches was much larger than the maximum path of dislocation half-loops, whose density was determined according to the standard silicon technology [9]. The edge regions were not analyzed.

The motion of dislocations was induced by tensile stresses up to 20 MPa emerging during bending relative to the $[11\bar{2}]$ axis by the four-support method [10, 11] for 120 min in the temperature range 850–950 K. Their displacement was controlled by the method of repeated chemical etching. The path lengths of dislocations segments (half-loops) was recorded using a MII-4 microscope with the help of an eyepiece–micrometer to within approximately 0.5 μm and was detected from 5×10^2–10^3 individual dislocations. It is typical that the motion of dislocation in the absence of loading was not observed. A similar result was obtained using multiple chemical etching of the surface of the samples under investigation. The observed effect indicates that the influence of surface stoppers on the dislocation displacement dynamics is insignificant, which is typical of crystals with a high values of Peierls barrier, including silicon (≈1.3 eV).

Dislocation wafers were treated between the poles of a permanent electromagnet at room temperature for 30 min. The direction of the magnetic field (B = 0.7 T) coincided with the $[\bar{1}10]$ direction. The time between the magnetic exposure and mechanical deformation did not exceed 3 min.

The experimental results were analyzed from the histograms of the path length distribution of the detected dislocation half-loops (n_i) carrying information on the presence of obstacles overcome by dislocations during their motion. Since the spectrum of stoppers in the crystals under investigation is quite broad, the experimental histogram carries integrated information on overcoming various types of stoppers by a lin-

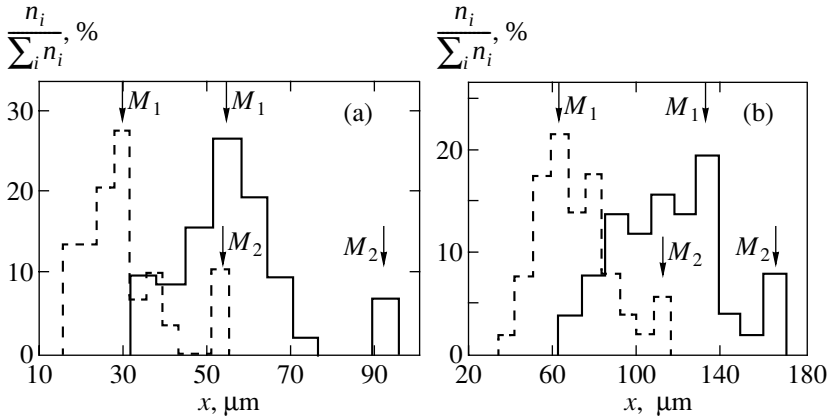

Fig. 1. Histograms of dislocation path length distribution under tensile stresses of 20 MPa at T = 600 K (a) and 625 K (b): dashed plots were obtained before the magnetic field treatment, while solid plots were recorded after the magnetic exposure for 30 min in field B = 0.7 T.

ear defect. Indeed, during the stimulated motion of defects in a Peierls relief, the mobility of a dislocation segment is determined by the time of its detachment from a stopper, which, in turn, depends on the effective decelerating ability of the stopper in the given slip plane. When a sample is subjected to a mechanical action, the slip planes present in it can be divided into favorable and unfavorable from the point of view of dislocation mobility. Such a division is dictated by the mutual spatial orientation of the allowed priority slip planes relative to the direction of mechanical action exerted on the sample. In the Si sample under investigation, whose lattice has the diamond-type structure, there is only one favorable slip plane (the (111) plane). The remaining slip planes from the {111} family, as well as {110}, are unfavorable since they are arranged at various angles to the direction of external action. Consequently, the stoppers encountered on the path of dislocations moving in different slip planes have different decelerating abilities, which directly affects the spatial distribution of dislocations in a moving dislocation ensemble. Hence, the clearly manifested peaks detected on histograms must be associated with dislocations overcoming of stoppers with various effective decelerating ability.

Typical histograms of the dislocation path distributions in samples exposed to a magnetic field (solid curve) and in control samples (dashed curve) presented in Fig. 1 serve as a convincing illustration of what has been said above. Obviously, both peaks on the histograms (indicated by arrows) are determined by the dynamics of interaction between dislocations and various stoppers in the case of mechanically activated motion of the dislocations at velocity V.

Indeed, at finite temperatures, an equilibrium concentrations of kinks is established on dislocations. If a kink on a dislocation performs random motion under the action of thermal fluctuations, its diffusion coefficient has the form [12]

$$D = \beta a^2 \omega. \quad (1)$$

Here, a is the space parameter (the separation between the grooves of the Peierls relief), β is the correlation factor, and ω is the frequency of jumps in the corresponding direction. After time t, the defect can move over a distance

$$x \approx \sqrt{\frac{a^2 n}{\tau}} = \mu_m F t, \quad (2)$$

where $n = t/\tau$ is the number of jumps during time t, τ is the time of settled life of the kink determined by the delay time of the dislocation (or its segment) at a stopper, F is the force acting on the defect, and μ_m is the mechanical mobility.

For low stresses, kinks diffuse directionally in the field of external forces, causing the displacement of the dislocation as a whole with velocity

$$V = \mu_m F. \quad (3)$$

For this reason, the motion of the most rapid dislocations is characterized by the smallest values of τ and is limited by their interaction with low-energy obstacles possessing a small delay time, whose concentration on the dislocation line must be dominating. The predominance of one type of centers over others determines the strong difference between the amplitudes of the peaks on experimental histograms[1] (Fig. 1).

An increase in temperature facilitating the movement of dislocations must lead to a natural change in

[1] The "low" peak M_2 practically does not appear in 15% of the histograms recorded from different scratches under identical conditions of deforming, while the main peak M_1 is present in all histograms.

the velocity of migration for a fixed mechanical load and, hence, to an increase in the path length L.

An analysis of the temperature dependence proved that mechanical deformation of dislocation samples unexposed to the magnetic field lead to changes in the dynamics of linear defects in the indicated temperature range (see Fig. 1). For example, the change in temperature by only 25 K leads to an increase in the path lengths of both "slow" (by a factor exceeding 2) and "fast" (by a factor of 1.8) dislocations.

The temperature dependences of the path lengths before the exposure to the magnetic field are presented in Fig. 2. It follows from these dependences that the path length in the initial samples is of the thermally activated type for all types of stoppers:

$$L = L_0 \exp(-E_a/kT). \quad (4)$$

Here and below, L_0 is the preexponential factor, T is the temperature, and E_a is the apparent activation energy of the process of defect displacement.

The considerable difference between the path lengths of fast and slow dislocations at a fixed temperature presumes different origins of stoppers and, hence, the difference is the activation barriers overcome by them. Indeed, these values for fast and slow dislocations were found to be $E_{aF} = 2.1 \pm 0.1$ eV and $E_{aS} = 1.8 \pm 0.1$ eV, respectively. The obtained value of activation energy E_{aF} is in good agreement with the available results of similar experiments [9]. We attribute the observed differences in the activation energy mainly to point defects limiting the dislocation dynamics in a defect crystal [9]. The relation between the amplitudes of the M_1 and M_2 peaks on the histograms determines the difference in the concentrations of low- and high-energy stoppers overcome by dislocations.

In another series of experiments, we investigated the effect of magnetic perturbations on the motion of linear defects under mechanical loading. For this purpose, we repeated the experiments described above on samples preliminarily exposed to a magnetic field. The field itself did not lead to a noticeable displacement of dislocation lines if no external load was subsequently applied. Its role was manifested only after the application of a mechanical perturbation reflecting the "magnetic" past history of the sample.

A convincing illustration of this fact is the considerable difference of the histograms of dislocation path lengths (see Figs. 1a and 1b). It can be seen that as in ionic crystals [2, 13], the dislocation structure of silicon "remembers" the fact of magnetic exposure, which eases the motion of linear defects after the withdrawal of the magnetic field. The role of the field in this case is reduced to a change in the energy state of a dislocation core and point defects facilitating the weakening of their interaction during thermally activated motion.

In order to verify the effect of magnetically stimulated transitions on the energy barriers for the motion of dislocations, the temperature variations of the path lengths in silicon after the magnetic exposure were recorded. The results of investigations indicated an increase in the dislocation path lengths after the magnetic treatment in the entire temperature range, indicating the effect of the magnetic field on both types of stoppers. A distinguishing feature of this effect is manifested in different values of activation energy recorded for slow ($E_{aS} = 1.6 \pm 0.1$ eV) and fast ($E_{aF} = 1.4 \pm 0.1$ eV) dislocations (curves *3* and *4* in Fig. 3). Thus, the magnetic field facilitates a certain decrease in the energy barrier heights in the case of mechanically activated motion of defects, which is in accord with the results of publications on the deformation of LiF and InSb crystals in a constant magnetic field [8, 13]. The reason behind such a change in the activation energy of the motion of dislocations is associated with spin-dependent detachment of a dislocation from a paramagnetic center (phosphorus ion). The magnetic field "neutralizes" a part of the paramagnetic stoppers and hence lowers the energy barriers for the motion of a linear defect.

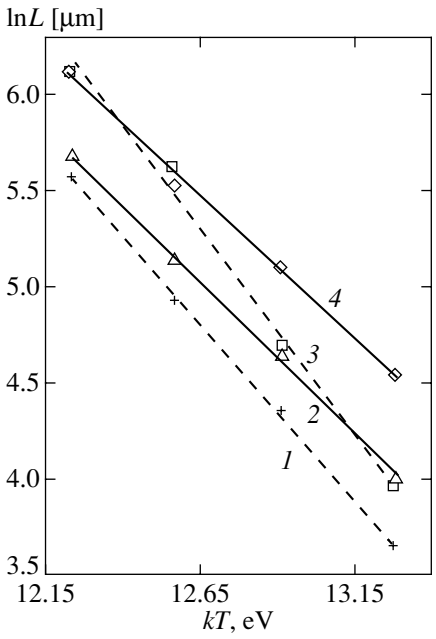

Fig. 2. Activation dependences before (dashed lines) and after (solid lines) magnetic treatment of slow (*1*, *2*) and fast (*3*, *4*) dislocations.

The kinetics of displacement of dislocations under the combined action of temperature and magnetic field is determined by thermally activated and magnetically stimulated contributions. The thermally activated component increases with temperature, while the magnetically stimulated component remains unchanged. Consequently, the effect of magnetic field on the dislocation path length is manifested more clearly in the low-temperature range, while the strong difference in the dynamics of slow and fast dislocations subjected to magnetic field is deter-

mined by the higher concentration of paramagnetic stoppers encountered by fast defects during their motion.

Thus, we detected the magnetoplastic effect in silicon single crystals, which is manifested in the change in the velocity of dislocation half-loops under mechanical action following the magnetic exposure. The aftereffect of the magnetic field is associated with residual changes in the state of point defects and dislocation cores.

It was found that in the silicon crystals under investigation, the magnetic field may lower the activation barriers in the motion of dislocations. The plasticization of crystals after annealing in a magnetic field indicates the formation of stoppers with different sensitivities to the magnetic field in silicon. This is manifested in different heights of the activation barriers emerging in the motion of dislocations for two types of stoppers after exposure to a magnetic field.

ACKNOWLEDGMENTS

This work was supported financially by the Ministry of Education of the Russian Federation under the program "Degradation Processes in Multilayered Thin-Film Structures" and by the Russian Foundation for Basic Research (project no. 98-02-03335).

REFERENCES

1. V. I. Al'shits, E. V. Darinskaya, T. M. Perekalina, *et al.*, Fiz. Tverd. Tela (Leningrad) **29**, 467 (1987) [Sov. Phys. Solid State **29**, 265 (1987)].
2. Yu. I. Golovin, R. B. Morgunov, and S. E. Zhulikov, Fiz. Tverd. Tela (St. Petersburg) **39**, 495 (1997) [Phys. Solid State **39**, 430 (1997)].
3. V. I. Al'shits, E. V. Darinskaya, E. Yu. Mikhina, *et al.*, Fiz. Tverd. Tela (St. Petersburg) **38**, 2426 (1996) [Phys. Solid State **38**, 1333 (1996)].
4. V. I. Al'shits, E. V. Darinskaya, and E. A. Petrzhik, Fiz. Tverd. Tela (St. Petersburg) **34**, 155 (1992) [Sov. Phys. Solid State **34**, 81 (1992)].
5. O. I. Datsko and V. I. Alekseenko, Fiz. Tverd. Tela (St. Petersburg) **39**, 1234 (1997) [Phys. Solid State **39**, 1094 (1997)].
6. V. N. Davydov, E. A. Loskutova, and E. N. Naĭden, Fiz. Tekh. Poluprovodn. (Leningrad) **23**, 1596 (1989) [Sov. Phys. Semicond. **23**, 989 (1989)].
7. Yu. I. Golovin, R. V. Morgunov, A. A. Baskakov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **70**, 114 (1999) [JETP Lett. **69**, 127 (1999)].
8. E. V. Darinskaya, E. A. Petrzhik, S. A. Erofeeva, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **70**, 298 (1999) [JETP Lett. **70**, 309 (1999)].
9. A. A. Skvortsov, A. M. Orlov, L. I. Gonchar, *et al.*, Fiz. Tverd. Tela (St. Petersburg) **42** (2000) [Phys. Solid State **42**, 1861 (2000)].
10. I. V. Ostrovskiĭ, L. P. Steblenko, and A. B. Nadtochiĭ, Fiz. Tverd. Tela (St. Petersburg) **42**, 478 (2000) [Phys. Solid State **42**, 488 (2000)].
11. V. A. Makara, L. P. Steblenko, V. V. Obukhovskiĭ, *et al.*, Fiz. Tverd. Tela (St. Petersburg) **42**, 854 (2000) [Phys. Solid State **42**, 877 (2000)].
12. J. P. Hirth and J. Lothe, *Theory of Dislocations* (McGraw-Hill, New York, 1967; Atomizdat, Moscow, 1972).
13. V. I. Al'shits, A. A. Urusovskaya, A. E. Smirnov, *et al.*, Fiz. Tverd. Tela (St. Petersburg) **42**, 270 (2000) [Phys. Solid State **42**, 277 (2000)].

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 93, No. 1, 2001, pp. 121–125.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 1, 2001, pp. 139–144.

SOLIDS
Structure

The Sound Velocity, Internal Friction, and Thermal Expansion in a Single Crystal of $La_{0.85}Sr_{0.15}MnO_3$

R. I. Zainullina[a], N. G. Bebenin[a, *], V. V. Mashkautsan[a], A. M. Burkhanov[a], V. S. Gaviko[a], V. V. Ustinov[a], Ya. M. Mukovskii[b], D. A. Shulyatev[b], and V. G. Vasil'ev[c]

[a]*Institute of Physics of Metals, Ural Division, Russian Academy of Sciences, Yekaterinburg, 620219 Russia*

[b]*Moscow State Institute of Steel and Alloys, Moscow, 117936 Russia*

[c]*Institute of Chemistry of Solids, Ural Division, Russian Academy of Sciences, Yekaterinburg, 620219 Russia*
**e-mail: bebenin@imp.uran.ru*

Received December 22, 2000

Abstract—The results are given of the investigation of the temperature dependence of the sound velocity v, internal friction Q^{-1}, and thermal expansion $\Delta L/L$ of a single crystal of $La_{0.85}Sr_{0.15}MnO_3$ in the temperature range from 5 to 400 K. Clearly defined singularities of the elastic properties at a temperature of charge ordering $T_{co} \approx 200$ K are revealed. The results of X-ray diffraction studies performed at room temperature are used to determine the orientations and estimate the sizes of twins. The correlation between magnetic and structure inhomogeneities is established. Based on the results of analysis of the temperature dependence of internal friction and thermal expansion, an assumption is made of the presence in a single crystal of $La_{0.85}Sr_{0.15}MnO_3$ of a structural transition in the temperature range from 15 to 60 K that has not been observed previously. *© 2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Lanthanum manganites with colossal magnetoresistance (CMR) exhibit a strong correlation between the lattice, electron, and magnetic subsystems of crystal. This brings about the presence of phases differing by their crystal structure, pattern of conductance, and type of magnetic ordering. Transitions between these phases may be induced by the variation of temperature, pressure, or magnetic field.

A series of phase (x, T)-diagrams are suggested for the $La_{1-x}Sr_xMnO_3$ system [1–3], which represent magnetic and structural transitions, as well as the regions of charge and orbital ordering. The most substantial differences between diagrams of different authors pertain to the range of $x < x_c$ ($x_c = 0.17$ is the concentration at which the metal–dielectric transition occurs in the ferromagnetic phase). In view of this, we have investigated the temperature dependence of the sound velocity, internal friction, and thermal expansion of a single crystal of $La_{0.85}Sr_{0.15}MnO_3$. The study of the elastic properties of this compound is of interest *per se*, because no such data are available in the literature.

Unlike most publications on the subject of CMR manganites, we give the results of a detailed X-ray diffraction study of the above-identified single crystal.

2. SAMPLES AND EXPERIMENTAL PROCEDURE

A single crystal of $La_{0.85}Sr_{0.15}MnO_3$ was grown by the floating-zone method with radiation heating; it had the form of a cylindrical rod 3.5 mm in diameter and 32 mm long. This rod was used to investigate the elastic properties. For other measurements, the samples were cut out from different parts of the initial single crystal.

The samples were oriented in an RKV-86A chamber under conditions of Mo radiation and investigated in an RKU-114M chamber and in a DRON diffractometer under conditions of monochromatized K_αCr radiation at room temperature. In the case of the diffractometer, the samples were precrushed. The calculation of powder diffraction patterns involved the use of FullProff computer codes [4].

The elastic properties and thermal expansion were investigated in the range from 5 to 400 K. The method of composite vibrator [5] at frequencies of about 70 kHz was used to determine the sound velocity and internal friction. This method involves measurements of the resonance frequency and Q factor of a mechanical system consisting of the sample being investigated and a piezoelectric transducer glued on to it. The sensor was provided by a quartz-crystal vibrator exciting longitudinal oscillation. The description of the procedure for calculating the sound velocity and internal friction of a sample is found, for example, in [6].

The thermal expansion was determined by the tensometric method using a dc bridge circuit. One strain gage with a base of 3 mm was glued onto the flat surface of the sample parallel to the direction of crystal growth, and the other strain gage was glued onto a quartz plate. High-purity aluminum was used for check measurements in the investigated temperature range.

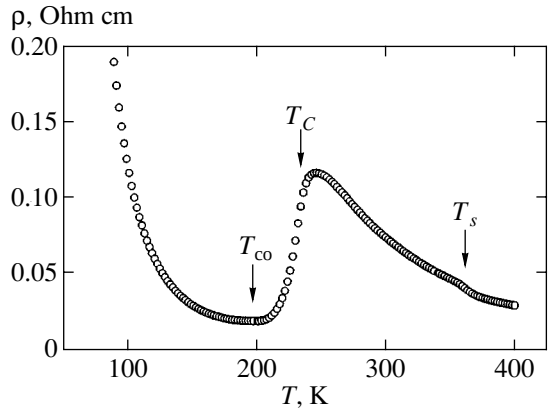

Fig. 1. The temperature dependence of the resistivity of single crystal of $La_{0.85}Sr_{0.15}MnO_3$.

All measurements were performed in an atmosphere of helium gas with the temperature varying at an average rate of 30 deg/h.

Also performed were resistivity measurements common with GMR manganites. The Curie temperature was determined by the Below-Arrott method of thermodynamic coefficients by treating the magnetization isotherms measured using a vibrating-coil magnetometer. For the single crystal investigated by us, T_C = 232 K, which is consistent with the literature data [7].

3. MEASUREMENT RESULTS

According to the data of X-ray studies at room temperature, the single crystal has a *Pnma* orthorhombic crystal structure with the parameters

$$a = 5.546\ \text{Å}, \quad b = 7.175\ \text{Å} \quad (b/\sqrt{2} = 5.073\ \text{Å})$$

$$c = 5.509\ \text{Å}.$$

These values agree well with the data of De Leon-Guevara *et al.* [8]. The table gives the results of calculating diffraction patterns obtained at room temperature, namely, the positions of atoms in an elementary cell, the isotropic temperature factor *B*, and the occupation of atomic positions. One can see that the composition of the single crystal being studied corresponds to the chemical formula of $La_{0.85}Sr_{0.15}MnO_3$.

It has been found that the single-crystal sample is broken up into structural domains in which the [010], [$\bar{1}$01], and [101] orientations of the axes of the orthorhombic lattice coincide. These data agree with the results of neutron diffraction studies of $LaMnO_3$ [9]. In a cubic description, the above-identified directions correspond to the [001], [010], and [100] directions. Structural domains with the [201], [161], and [323] orientations (for a cubic lattice, it is a direction of the [310] type) are arranged along the direction of single crystal growth. X-ray diffraction patterns for a specially prepared needle-shaped sample with a thickness of 0.5 mm or less exhibit reflexes from three orientations of twins of almost the same intensity; therefore, one can assume that the twin size is appreciably less than 0.1 mm.

The curve of temperature dependence of the resistivity of single crystal being investigated, $\rho(T)$ (Fig. 1), has a shape that is typical of lanthanum manganite samples of this composition. In the region of $T_s \approx 360$ K, a hysteresis is observed, caused by the structural transition from the $R\bar{3}c$ high-temperature rhombohedral phase to the *Pnma* orthorhombic phase [10]. The Curie temperature is found as a maximum of the derivative $d\rho/dT$. In the ferromagnetic region, a minimum is observed on the $\rho(T)$ curve at a temperature of about 198 K; it is agreed that the temperature of minimum resistivity in $La_{0.85}Sr_{0.15}MnO_3$ is the temperature T_{co} of charge (polaron) ordering [2, 11, 12].

Figure 2 gives the temperature dependence of the velocity v of longitudinal ultrasonic waves in a single crystal of $La_{0.85}Sr_{0.15}MnO_3$. The transitions mentioned above show up as minima at 377, 232, and 208 K, respectively. As the temperature decreases, a considerable decrease in the velocity of sound is observed in the interval between T_s and T_C. Below T_C, the velocity of sound continues to decrease, and at $T \approx T_{co}$ it reaches a minimum; with a further decrease in temperature, the velocity of sound increases monotonically. The interval from 50 to 100 K stands out because of the small value of the derivative dv/dT.

The positions of atoms, the temperature factor *B*, and the occupation of atomic positions of single crystal of $La_{0.85}Sr_{0.15}MnO_3$ at room temperature, obtained as a result of calculation of X-ray diffraction patterns

	x	*y*	*z*	*B*, Å^2	Occupation
La	−0.01708	0.25000	−0.00801	0.04047	0.8498
Sr	−0.01708	0.25000	−0.00801	0.04047	0.1500
Mn	0.00000	0.00000	0.50000	0.07944	1.0000*
O(1)	0.49666	0.25000	0.05306	0.11620	1.0078
O(2)	0.24762	−0.02380	−0.26813	1.26803	0.9935

Note: * The value was not refined. Reliability factors: R_p = 4.52%, R_{wp} = 5.32%, R_{exp} = 6.2%, R_{Bragg} = 3.7%

In the temperature range from 180 to 240 K, which included T_C and T_{co}, we observed, in addition to the fundamental resonance on a lower frequency, a resonance whose intensity was maximal in the vicinity of T_{co}. The solid line in Fig. 2 indicates the temperature dependence of the sound velocity for the additional mode. One can see that the $v(T)$ curves for the fundamental and additional modes are similar. The emergence of the additional mode was previously observed during investigation of the elastic properties of ceramic HTSC samples [6].

Figure 3 gives the curve of temperature dependence of internal friction, $Q^{-1}(T)$. Peaks of internal friction are observed at 377, 232, and 205 K. Note that the peak at the Curie point is relatively low. The internal friction in the range from 120 to 190 K is low and depends little on temperature. A considerable rise of internal friction is observed with a further decrease in temperature. At $T = 60$ K, the value of Q^{-1} reaches a maximum, after which it starts to decrease.

The temperature dependence of thermal expansion $\Delta L/L$ for samples cut out from different parts of the rod proved to be somewhat different; however, the position of the singularities remained unvaried. Figure 4 gives the $\Delta L/L(T)$ curve for one of the samples. At $T \approx 367$ K, a jump is observed associated with the structural transition from the $R\bar{3}c$ phase to the *Pnma* phase. The transition from the paramagnetic to ferromagnetic state shows up as a characteristic kink on the $\Delta L/L(T)$ curve at $T \approx 234$ K. Singularities are observed in the neighborhood of T_{co}, whose form turned out to be different for samples cut out from different parts of the initial rod. With a further decrease in temperature, a smooth decrease in $\Delta L/L$ is observed up to $T = 60$ K. Then, $\Delta L/L$ increases to reach a maximum at $T \approx 15$ K, after which it decreases rapidly.

In the neighborhood of T_s, a temperature hysteresis is observed (see the inset in Fig. 4) that is characteristic of first-order phase transitions. Outside of the region of structural transition, the curves recorded upon heating and cooling almost coincide.

4. DISCUSSION OF THE MEASUREMENT RESULTS

By and large, the anomalies observed in the temperature dependence of the velocity of sound agree well with the temperatures of the known magnetic and structural transitions and are indicative of the strong variation of the elastic moduli of the manganite being investigated during phase transformations. The most radical changes of elastic moduli occur during structural transitions at temperatures T_s and T_{co}, which points to a strong rearrangement of the phonon spectrum. Note that the singularities at $T = T_{co}$ do not show up on the curve of temperature dependence of lattice parameters determined by X-ray techniques.

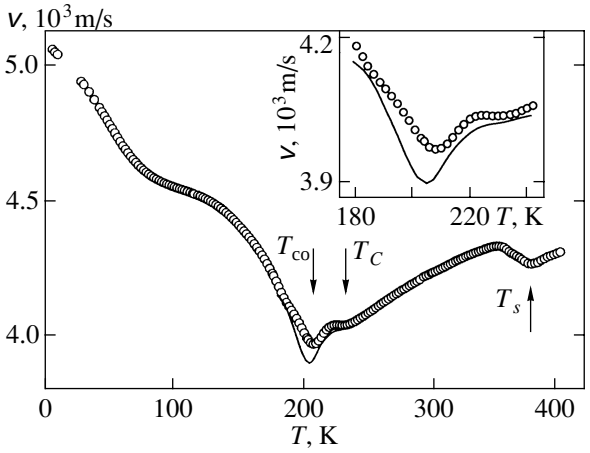

Fig. 2. The temperature dependence of the velocity of ultrasonic waves for a single crystal of $La_{0.85}Sr_{0.15}MnO_3$. The solid line indicates the velocity of sound calculated for the additional mode.

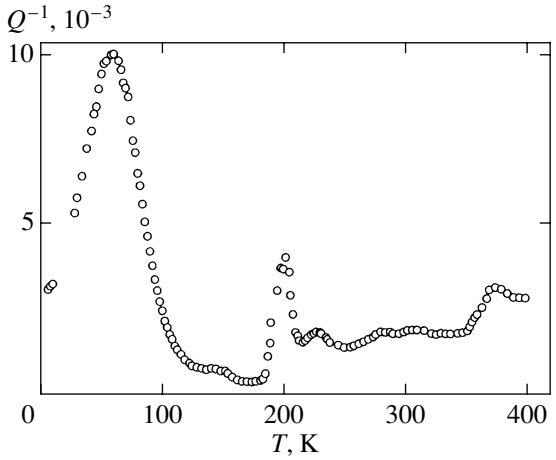

Fig. 3. The temperature dependence of internal friction of a single crystal of $La_{0.85}Sr_{0.15}MnO_3$.

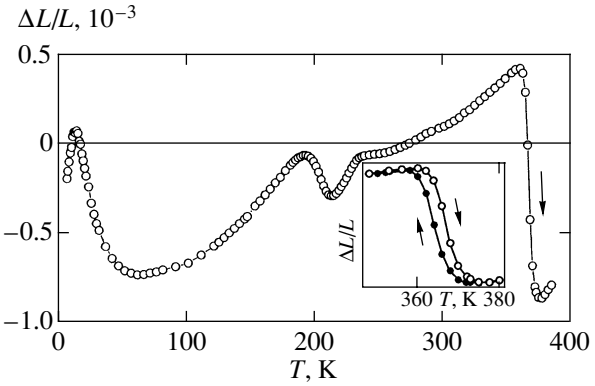

Fig. 4. The temperature dependence of thermal expansion. The inset shows temperature hysteresis in the region of temperature T_s of structural transition.

Note further the increase in the sound velocity with the temperature in the paramagnetic region increasing up to the neighborhood of T_s. Because the conductance of $La_{0.85}Sr_{0.15}MnO_3$ in this temperature range increases markedly with temperature (see Fig. 1) due to the increase in the number of current carriers in delocalized states [13], one can assume that this increase in the sound velocity is associated with the increase in the contribution made by the delocalized carriers to elastic moduli.

The presence of additional resonance in the vicinity of T_C and T_{co} points to the interaction between normal modes in the rod. Such interaction may be due to the deviation of the direction of sound propagation from the principal crystallographic axes. Indeed, the [310] direction of crystal growth, which coincides in our case with the direction of sound propagation, deviates somewhat from the [100] axis (in a cubic approximation). However, additional resonance is absent from the paramagnetic region. This is indicative of the fact that the foregoing reason is not the main one.

Another possible reason for the interaction between normal modes is the presence of inhomogeneities in the sample [14]. Inhomogeneities of several types may be identified in manganites, which are realized on different scales. First, the magnetic [15] and elastic [16] properties, as well as the thermal expansion, differ somewhat for samples cut out from different parts of the initial single crystals. Second, a sample is broken up into small-sized structural domains. And, finally, microdomains of the monoclinic phase may exist [17]. Because no additional resonance is present at $T > T_c$, the presence of structure inhomogeneities *per se* is insufficient for explaining the additional resonance. Because the additional resonance is observed only in the ferromagnetic region in the vicinity of T_C, the reason for its formation is apparently associated with the nonuniform distribution of magnetization. This nonuniformity, however, cannot be due to the magnetic domain structure, because, at a fairly large distance from T_C, the additional resonance disappears and, in addition, the singularity on the $Q^{-1}(T)$ curve at $T = T_C$ is defined weakly.

The additional resonance may be caused by the interrelation between structure and magnetic inhomogeneities. According to Kadomtseva *et al.* [12], the magnetostriction in $La_{1-x}Sr_xMnO_3$ at $x < x_c$ is high in the vicinity of T_C and T_{co}. In the neighborhood of magnetic phase transition, spontaneous magnetization in different parts of the crystal differs appreciably because of the scatter of the values of the Curie temperature: in lanthanum manganites, this scatter is of the order of at least several degrees [15]. Because the sample is broken up into small-sized structural domains which are apparently characterized by different values of T_C, the magnetostriction in the vicinity of the temperature of transition from the paramagnetic to ferromagnetic state must give rise to inhomogeneous internal stresses. Further away from T_{co} deep into the ferromagnetic region, the magnetostriction constants and the scatter of magnetization values decrease, as a result of which the additional resonance must attenuate (as is the case observed).

The presence of clearly defined anomalies on the $\Delta L/L(T)$ and $Q^{-1}(T)$ curves in the range from 15 to 60 K leads one to assume that a structural phase transition, which was not previously observed, occurs in this temperature range. Note that, according to the results of neutron diffraction studies of single crystal of $La_{0.85}Sr_{0.15}MnO_3$ [11], the curve of temperature dependence of the intensity of line $(2, 2, -0.5)_0$ also exhibits a singularity in the range from 10 to 50 K. Manganite of similar composition ($La_{0.9}Sr_{0.1}MnO_3$) at $T_1 \approx 100$ K is characterized by transition from the orthorhombic to pseudocubic phase, accompanied by a decrease in volume [18]. Because the temperature of transition from the rhombohedral to orthorhombic phase in $La_{1-x}Sr_xMnO_3$ decreases with the increase in the bivalent ion concentration [1, 10], one can expect that, as x increases, the temperature of transition from the orthorhombic to pseudocubic phase decreases as well. Therefore, one can assume that the low-temperature transition in $La_{0.85}Sr_{0.15}MnO_3$ is, as in $La_{0.9}Sr_{0.1}MnO_3$, the transition from the orthorhombic to pseudocubic phase. In order to validate or disprove this reasoning, one must perform X-ray studies in the above-identified temperature range.

5. CONCLUSIONS

The investigated single crystal of $La_{0.85}Sr_{0.15}MnO_3$ is broken up into structural domains with a characteristic size of less than 0.1 mm.

The presence of a clearly defined minimum on the curve of temperature dependence of longitudinal sound velocity at $T = T_{co}$ points to a substantial rearrangement of the lattice upon charge ordering. Delocalized charge carriers seem to produce an appreciable contribution to elastic moduli.

The observed interaction between the normal modes in the ferromagnetic region in the vicinity of T_C and T_{co} points to relation between structure and magnetic inhomogeneities.

The results of analyzing temperature dependence of internal friction and thermal expansion leads one to assume that a structural transition occurs in $La_{0.85}Sr_{0.15}MnO_3$ in the range from 15 to 60 K.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research (project nos. 00-02-17544 and 00-15-96745).

REFERENCES

1. A. Urushubara, Y. Moritomo, T. Arima, *et al.*, Phys. Rev. B **51**, 14103 (1995).

2. A. A. Mukhin, V. Yu. Ivanov, V. D. Travkin, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **68**, 331 (1998) [JETP Lett. **68**, 356 (1998)].
3. M. Paraskevopoulos, F. Mayer, C. Hartinger, *et al.*, J. Magn. Magn. Mater. **211**, 118 (2000).
4. J. Rodríguez-Carvajal, Physica B (Amsterdam) **192**, 55 (1993).
5. H. J. McSkimin, in *Physical Acoustics: Principles and Methods*, Ed. by W. P. Mason (Academic, New York, 1964; Mir, Moscow, 1966), Vol. I, Part A, p. 272.
6. A. M. Burkhanov, V. V. Gudkov, I. V. Zhevstovskikh, *et al.*, Fiz. Met. Metalloved., No. 1, 68 (1990).
7. L. Vasiliu-Doloc, J. W. Lynn, A. H. Moudden, *et al.*, J. Appl. Phys. **81**, 5491 (1997).
8. A. M. De Leon-Guevara, P. Berthet, J. Berthon, *et al.*, Phys. Rev. B **56**, 6031 (1997).
9. F. Mousa, M. Hennion, J. Rodríguez-Carvajal, *et al.*, Phys. Rev. B **54**, 15149 (1996).
10. A. Asamitsu, Y. Moritomo, R. Kumai, *et al.*, Phys. Rev. B **54**, 1716 (1996).
11. Y. Yamada, O. Hino, S. Nohdo, *et al.*, Phys. Rev. Lett. **77**, 904 (1996).
12. A. M. Kadomtseva, Yu. F. Popov, G. P. Vorob'ev, *et al.*, Fiz. Tverd. Tela (St. Petersburg) **42**, 1077 (2000) [Phys. Solid State **42**, 1110 (2000)].
13. N. G. Bebenin, R. I. Zaĭnullina, V. V. Mashkautsan, *et al.*, Zh. Éksp. Teor. Fiz. **117**, 1181 (2000) [JETP **90**, 1027 (2000)].
14. T. R. Meeker and A. H. Meitzler, in *Physical Acoustics: Principles and Methods*, Ed. by W. P. Mason (Academic, New York, 1964; Mir, Moscow, 1966), Vol. I, Part A, p. 112.
15. S. E. Lofland, S. M. Bhagat, K. Ghosh, *et al.*, Phys. Rev. B **56**, 13705 (1997).
16. Yu. P. Gaĭdukov, N. P. Danilova, A. A. Mukhin, and A. M. Balbashov, Pis'ma Zh. Éksp. Teor. Fiz. **68**, 141 (1998) [JETP Lett. **68**, 153 (1998)].
17. M. Hervien, G. Van Tendeloo, V. Caignaert, *et al.*, Phys. Rev. B **53**, 14274 (1996).
18. A. V. Korolyov, V. Ye. Arkhipov, V. S. Gaviko, *et al.*, J. Magn. Magn. Mater. **213**, 63 (2000).

Translated by H. Bronstein

Journal of Experimental and Theoretical Physics, Vol. 93, No. 1, 2001, pp. 126–135.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 1, 2001, pp. 145–155.

SOLIDS
Structure

"Phonons" in Two-Dimensional Vortex Lattices

V. V. Smirnov and K. V. Chukbar*
Russian Research Center "Kurchatov Institute", pl. Kurchatova 1, Moscow, 123182 Russia
**e-mail: chukbar@dap.kiae.ru*
Received January 18, 2001

Abstract—The macroscopic dynamics of regular lattices formed by 2D vortices of various physical origin is considered. The effective equations describing this dynamics are derived and their properties are analyzed. The general feature of the evolution of such systems and their peculiarities distinguishing qualitatively vortex ensembles from ordinary crystals are considered. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

In spite of its long history, the dynamics of two-dimensional pointlike vortices remains a popular object of investigations. It is inevitably considered in special chapters of all monographs and reviews devoted to vortex motion in various physical media (see, for example, [1–5]). This is obviously due to the importance of the associated problems (for example, a transition from smooth distributions of the "vorticity" in space to its local concentration at individual points is directly related to the strategic question on finite-dimensional approximations of continuous media) and with the novelty and peculiarity of the "mechanical" behavior of such systems (the representation of vortices in the form of individual particles makes it possible to visualize apparently mathematical abstractions like the phase space). At the same time, in spite of the persistent attention that has been paid to this problem for a long time, its analysis is confined, as a rule, to a certain type of vortices (these types are different in different fields) with a fixed "flow function" (see below), which strongly exaggerates certain (sometimes accidental) features of the system and masks general regularities. As a matter of fact, however, the vortex dynamics in various physical "fluid" media must be considered from unified positions since the mathematical basis for this dynamics is a fairly universal freezing-in equation [6, 7].

This work aims at studying dynamic behavior of large (infinite in our case) regular ensembles (lattices) of identical vortices of different types, which possess different symmetries (hexagonal, square, or triangular). We are speaking of the evolution of "long-wave" perturbations of their regularity, which are similar to acoustic phonons in ordinary crystals, but differ from them in some properties. The relevant equations are derived (which describe, so to speak, the "secondary" hydrodynamics of a vortex medium) and their properties are considered. In spite of the observed high sensitivity of the behavior of vortex "crystals" to the lattice symmetry (see classical works [8–10]), as well as to the origin of vortices (flow function), the generality established qualitatively is also beyond all question. It can be noted, however, that such a sensitivity to "details" is also a generalizing and universal feature inherent precisely in vortex "mechanics."

2. BASIC CONCEPTS AND FORMULATION OF THE PROBLEM

First of all, we recall the key factors determining the vortex dynamics. A 2D vortex as a point particle is characterized, apart from its coordinates $\mathbf{r}_0$ in the x, y plane, by an individual intensity (charge) q_0 and the flow function $\psi(\mathbf{r} - \mathbf{r}_0)$, which is universal for a given system (for vortices of the given physical origin). Its main feature is that it creates in the surrounding space a circular incompressible flow of the initial (primary) continuous medium in accordance with formula

$$\mathbf{v} = q_0 \mathbf{e}_z \times \nabla \psi$$

($\mathrm{div}\mathbf{v} \equiv 0$), which carries away all the remaining vortices (frozen in the flow). In other words, the dynamics of an ensemble of vortices is described by the "mechanical" equations

$$\dot{\mathbf{r}}_i = \mathbf{e}_z \times \nabla \sum_{j \neq i} q_j \psi(\mathbf{r}_j - \mathbf{r}_i) \qquad (1)$$

with the Hamiltonian

$$\sum_{i > j} q_i q_j \psi(\mathbf{r}_i - \mathbf{r}_j)$$

and with canonical variables $\{q_i y_i, q_i x_i\}$; i.e., the configuration space of the given system indeed coincides with the phase space (see above).

In view of the initial isotropy of the plane, the flow function depends, as a rule, only on the distance to the vortex: $\psi(|\mathbf{r} - \mathbf{r}_0|)$. Vortices in the Vlasov equation (in real phase spaces), which are referred to as the Bernstein–Green–Kruskal waves (see, for example, [6]) and will not be considered here, form a noticeable exception. As regards the obvious violation of the "chiral" symmetry in (1), the physical reasons behind this effect

may be different for different vortex media. Two classes of vortices can be singled out. In the first class (ordinary vortices), they themselves are a source of gyrotropy since they have the form of a δ-functional distribution of the curl of the generalized momentum of liquid particles (i.e., $\mathrm{curl}\mathbf{P} \propto \mathbf{e}_z \cdot \sum q_i \delta(\mathbf{r} - \mathbf{r}_i)$ and, hence, charges q_i are pseudoscalar quantities). In actual practice, $\mathrm{curl}\mathbf{P}$ is naturally not concentrated at a point but is smeared over a finite region known as the core of a vortex; however, this circumstance virtually does not affect Eq. (1) in the case when the size of this region is smaller than the separation between neighbors. In the second class (drift vortices), cores are ordinary force centers (e.g., the concentration of the electric charge or gravitating mass with truly scalar q_i), but the medium itself is gyrotropic (e.g., due to a strong magnetic field applied to it or due to a rapid rotation). In this case, dominating Lorentz or Coriolis forces make other particles (force centers experiencing the action of neighboring vortices) to drift (precess) at right angles to the applied force.

The 2D nature of the motion can be due to the absence of longitudinal (directed along z) perturbations in actually three-dimensional infinitely long vortex filaments as well as due to actual two-dimensionality of the medium having the form of an infinitely thin layer (film). In this case, flow functions describing the response of the medium to a vortex perturbation may depend on the dimension of the medium (2D or 3D); i.e., function ψ is generally determined not only by the physical origin of vortices, but also by their geometry.

The behavior of ensembles consisting of just identical vortices with $q_i = q_0$ is often quite interesting. The actual reason behind such an identity is often associated with quantum effects (see, for example, [3, 4, 8–12]), but even in the case of essentially classical system, such a formulation of the problem is quite popular (see, for example, [1, 2, 5, 12]) apparently in view of a certain analogy with ordinary (Newtonian) particles and the convenience of numerical simulation. We can assume that the assumption made in this case leads to a considerable advance in the theoretical study of the vortex system which exhibits, as before, a peculiar (and even exotic) behavior.

In our formulation, such identical vortices are initially located at the sites of an infinite regular hexagonal, square, or triangular lattice (we mean that vortices are the apexes of regular n-gons covering the plane as a tiling) at a distance a from nearest neighbors (whose number is three, four, or six, respectively). It can easily be seen that this ensures the steady state ($\dot{\mathbf{r}}_i \equiv 0$) of such a crystalline state. However, even a small displacement of vortices from their positions sets the system in motion. We will be interested here (see above) only in large-scale macroscopic violations of the initial symmetry (with the characteristic spatial scale $\lambda \gg a$), which are similar to acoustic phonons in ordinary crystals, when it is as if pointlike vortices form a certain secondary continuous medium. Obviously this limit must correspond to the minimum effect of microscopic parameters of the crystal on its dynamics; nevertheless, this effect is significant. In particular, the decisive contribution always comes from the initial discreteness of the lattice (see below).

The typical distinguishing features of the functions $\psi(r)$ under investigation require a more detailed analysis. The main feature is the law describing the decrease of the function as $r \longrightarrow \infty$. It should be noted that the requirement of stability of finite-size vortices (with a nonzero core) leads to a monotonic decrease in $\psi(r)$, i.e., to a sign-definite value of $\psi' < 0$ (the constancy of the direction of the produced circulation of the medium) [6, 7, 12]. Besides, the flow function for all known physical types of vortices is also convex; i.e., $\psi'' > 0$ (the circulation rate also decreases monotonically). The physical reasons behind this property are not quite clear (in this case, an analogy with the absence of expansion shock waves in ordinary substances is possible), but its consequences are quite significant (see below).

Quite often, ψ is a certain power function of the distance ($\psi = 1/r^\alpha$, and the version with $\alpha = 0$ is included here through $-\ln r$). Such are the cases with an ideal liquid as well as electron plasma and superconductors for small r($\alpha = 0$), or superconducting films (plasma layers) for large r ($\alpha = 1$; this case also includes drift vortices in rapidly rotating dust gravitizing disks or in charged plasma layers in an external magnetic field) (see the reviews cited above). This version is interesting due to the absence of its own characteristic scale for ψ (to be more precise, this scale is variable and is of the order of r). On the other hand, functions ψ with a quite definite intrinsic scale b also exists. For instance, function ψ for vortex lines in bulk superconductors (in infinite electron plasma) has the form of the Macdonald function $K_0(r/b)$ with the London (collisionless skin) screening length $b = c/\omega_{pe}$. It is interesting to note that $\psi|_{r \to 0} \longrightarrow \infty$ again in all the cases known to us.

Thus, in accordance with what has been said above, we are interested in secondary hydrodynamics or mechanics of regular vortex media, which can be described in terms of the displacement of each vortex from the lattice site corresponding to it; i.e., through the initially discrete function $\boldsymbol{\xi}_i(t)$. If this function satisfies the inequality $|\boldsymbol{\xi}_i - \boldsymbol{\xi}_j| \ll a$ in the linear (for the time being) approximation, the crystal dynamics is described by the equation (cf. (1))

$$\dot{\boldsymbol{\xi}}_i = -\mathbf{e}_z \times q_0 \sum_{j \neq i} \frac{\partial}{\partial \mathbf{r}_i}\left[(\boldsymbol{\xi}_i - \boldsymbol{\xi}_j)\frac{\partial \psi(|\mathbf{r}_i - \mathbf{r}_j|)}{\partial \mathbf{r}_i}\right]. \quad (2)$$

In spite of the apparent simplicity of this formula, the summation over the infinite (although regular) lattice considerably required in this case complicates its analysis for arbitrary ψ.

This problem was apparently formulated and rigorously solved for the first time by Tkachenko [8–10] for vortices in an ideal liquid (to be more precise, in superfluid helium). However, the degeneracy of the case under investigation (roughly speaking, the fact that $\Delta\ln|\mathbf{r}| \propto \delta(\mathbf{r})$; see below), which considerably facilitated the analytic advance, did not allow us to estimate properly the specific features of the obtained results and even led to some not quite correct conclusions which still continue to appear in reviews without noticeable changes.

As a matter of fact, in the case of an arbitrary function ψ, the flow produced by pointlike vortices, which is incompressible ($\mathrm{div}\mathbf{v} = 0$), is generally not a vortex-free flow ($\mathrm{curl}\mathbf{v} \neq 0$!). Consequently, the powerful methods in the theory of analytic functions (a 2D vector field $\mathbf{v}$ can be presented in this case in the form of a complex-valued holomorphic function), which are suitable for an ideal liquid (for which $\mathbf{P} \propto \mathbf{v}$; i.e., the condition $\mathrm{curl}\mathbf{P} \equiv 0$ satisfied outside the vortex core is automatically extended to the velocity field: $\Delta\ln r = 0$) and which were elegantly used by Tkachenko, cannot be used for a universal analysis. Moreover, a unified approach to the problem does not exist at all since the physical properties of vortex crystals differ significantly depending on the behavior of $\psi(r)$. Each specific case requires its own approximation, which, unfortunately, does not always lead to exact numerical values of the sought parameters in Tkachenko's sense. For this reason, many formulas can be derived only to within a numerical factor. Nevertheless, their functional dependences and qualitative features of the solutions are quite rigorous.

We single out three specific cases of this kind. The first is typical of rapidly attenuating functions $\psi(r)$ (which correspond to the version with $\alpha > 2$, i.e., for the convergence of $\int \psi\, d^2\mathbf{r}$ for $r \longrightarrow \infty$, for power flow functions or $b \ll a$ for screened functions), when the dominating contribution to the dynamics of each vortex comes only from its nearest neighbors. Conversely, the second case is associated with small values of α (or with the case $b \gg \lambda$), when the motion of vortices is determined by a large aggregate of neighbors located at macroscopic distances $\sim\lambda$. In this case, nonlocal interaction effects play a decisive role. Finally, the version corresponding to $a \ll b \ll \lambda$ can be classified as locally macroscopic. In any case, an analysis leads to effective continual equations for $\boldsymbol{\xi}$ (which is now regarded as a continuous function of $\mathbf{r}$ and t, describing acoustic deformation of the secondary vortex continuous medium ($\lambda \gg a$!)).

3. NEAREST NEIGHBOR APPROXIMATION

Thus, we assume that the contribution to the velocity field perturbation causing a displacement of a given vortex comes only from the neighbors located exactly at a distance a from it (three neighbors for a hexagonal lattice, four for a quadratic lattice, and six for a triangular lattice). The accuracy of this model can easily be estimated for each specific case. For a rapid (e.g., exponential) law of attenuation for $\psi(r)$, the accuracy may be quite high. In accordance with (2), this contribution, which appears due to nonuniformity of lattice deformation ($\boldsymbol{\xi}_i \neq \boldsymbol{\xi}_j$ for $i \neq j$), amounts to

$$\delta\mathbf{v} = -\mathbf{e}_z \times q_0\left[\delta\boldsymbol{\xi}\frac{\psi'}{r} + \frac{\mathbf{r}}{r}(\delta\boldsymbol{\xi}\cdot\mathbf{r})\left(\frac{\psi'}{r}\right)'\right]\Bigg|_{r=a}, \qquad (3)$$

where $\delta\boldsymbol{\xi}$ is the difference in the displacements of a neighbor and a given vortex and $\mathbf{r}$ is the vector directed from the given lattice site to a neighboring site. It can easily be seen that the entire lattice dynamics is determined in this case by only two positive (see above) physical parameters $A = -\psi'/r|_{r=a}$ and $B = \psi''|_{r=a}$.

In order to go over to a continuous distribution of $\boldsymbol{\xi}$, it is sufficient to present the discrete difference in the form of the first two terms of the Taylor series (it should be recalled that $\lambda \gg a$):

$$\delta\boldsymbol{\xi} = x\frac{\partial\boldsymbol{\xi}}{\partial x} + y\frac{\partial\boldsymbol{\xi}}{\partial y} + \frac{x^2}{2}\frac{\partial^2\boldsymbol{\xi}}{\partial x^2} + xy\frac{\partial^2\boldsymbol{\xi}}{\partial x\partial y} + \frac{y^2}{2}\frac{\partial^2\boldsymbol{\xi}}{\partial y^2} \qquad (4)$$

and sum Eq. (3) over three, four, or six possible directions of $\mathbf{r}$. In the subsequent formulas, we assume everywhere that one of the nearest sites in all types of lattices (true, this is essential only for a quadratic lattice) lies at a point with coordinates $(a, 0)$. The answers differ significantly, demonstrating the considerable role of the geometrical parameters of the problem.

In the case of a hexagonal lattice, the decisive contribution comes from the linear terms in (4), while quadratic terms (in parameter $a/\lambda \longrightarrow 0$) can be neglected, and the effective equation in the continual limit has the form

$$\frac{\partial\boldsymbol{\xi}}{\partial t} = \frac{3}{4}q_0 a(A+B) \times\left[-\left(\frac{\partial\xi_x}{\partial y} + \frac{\partial\xi_y}{\partial x}\right)\mathbf{e}_x + \left(-\frac{\partial\xi_x}{\partial x} + \frac{\partial\xi_y}{\partial y}\right)\mathbf{e}_y\right], \qquad (5)$$

leading to the following (purely acoustic in spite of the nonstandard form of relation (5)!) dispersion relation for long-wave phonons:

$$\omega^2 = \frac{9}{16}q_0^2 a^2 (A+B)^2 k^2. \qquad (6)$$

However, in the cases with the central symmetry of the square and triangular lattices, the linear terms in the Taylor series are mutually cancelled out, thus necessitating the inclusion of quadratic corrections. This considerably affects the functional dependence in the equations being derived. For example, in a square lattice, the

evolution of acoustic "phonons" is described by the equation

$$\frac{\partial \boldsymbol{\xi}}{\partial t} = q_0 a^2 \left[\left(B\frac{\partial^2 \xi_y}{\partial y^2} - A\frac{\partial^2 \xi_y}{\partial x^2} \right) \mathbf{e}_x + \left(A\frac{\partial^2 \xi_x}{\partial y^2} - B\frac{\partial^2 \xi_x}{\partial x^2} \right) \mathbf{e}_y \right], \quad (7)$$

and their frequency is now a quadratic homogeneous function of the wave vector components:

$$\omega^2 = q_0^2 a^4 [(A+B)^2 k_x^2 k_y^2 - ABk^4]. \quad (8)$$

It can easily be seen that this version of the "secondary" medium is always unstable, e.g., relative to perturbations with $k_x = 0$ or $k_y = 0$.

It is interesting to note that in spite of the high lattice symmetry, phonons behave anisotropically. Their normal coordinates (in which the matrix on the right-hand side of the dynamic equation is diagonal) are ξ_x and ξ_y (i.e., $\partial\xi_x/\partial t \propto \xi_y$ and $\partial\xi_y/\partial t \propto \xi_x$). A similar behavior, which differs strongly from the behavior of standard phonons, is still encountered for certain types of waves in ordinary crystals with a considerable spin–orbit interaction. This is not very surprising since the symmetry properties of rotation in the quantum and classical regions are similar.

In view of the large number of neighbors, a triangular lattice requires more cumbersome calculations, but the obtained result is nevertheless compact (and even more symmetric):

$$\frac{\partial \boldsymbol{\xi}}{\partial t} = \frac{3}{8} q_0 a^2 \times [-(A-3B)\mathbf{e}_z \times \nabla \hat{d}\boldsymbol{\xi} + (3A-B)\nabla \hat{r}\boldsymbol{\xi}]. \quad (9)$$

Here, the following notation has been introduced for two scalar characteristics of the 2D vector field $\boldsymbol{\xi}$: $\hat{d}\boldsymbol{\xi} = \mathrm{div}\boldsymbol{\xi}$ and $\hat{r}\boldsymbol{\xi} = \mathbf{e}_z \cdot \mathrm{curl}\boldsymbol{\xi}$ (this notation will be used below). The "phonon" spectrum turns out to be isotropic again:

$$\omega^2 = \frac{9}{64} q_0^2 a^4 (A-3B)(3A-B)k^4. \quad (10)$$

It is stable both for power functions ψ with $\alpha > 2$ and for exponentially decreasing functions with $b \ll a$ (for which $B \gg A$), but differs considerably from the ordinary acoustic spectrum $\omega \propto k$. For normal coordinates, it is convenient to choose just the quantities $\hat{d}\boldsymbol{\xi}$ and $\hat{r}\boldsymbol{\xi}$ characterizing the bulk compression and torsion (shear) of the lattice.

Thus, even in the nearest neighbors approximation, vortex crystals demonstrate very peculiar polarization and dispersion properties, distinguishing them from classical Newtonian elastic analogues. The qualitative features of the behavior in this case are virtually independent of the details of the $\psi(r)$ dependence (if it falls into the required class), but the lattice geometry is extremely important.

4. NONLOCAL APPROXIMATION

In the case when macroscopic aggregates of vortices simultaneously affect the motion, the shot effects which are associated with the discreteness of the lattice (and, hence, its symmetry) and manifested explicitly in the previous section should not make a significant contribution to the phonon dynamics of vortex crystals. This situation is indeed encountered, but only in the zeroth approximation (in parameter a/λ), which is not always sufficient due to a specific degeneracy described below.[1] The continual limit in this case it attained as a result of transition from the summation of relation (3) in accordance with formula (2) to its integration. In the convolution-type integrals emerging in this case, it is convenient to remove both derivatives of ψ, integrating once by parts and taking the second derivative outside the integration symbol:

$$\frac{\partial \boldsymbol{\xi}}{\partial t} = -\frac{q_0}{S}\mathbf{e}_z \times \nabla \int \hat{d}\boldsymbol{\xi}(\mathbf{r}')\psi(|\mathbf{r}-\mathbf{r}'|)d^2\mathbf{r}'. \quad (11)$$

Here, S is the area of the unit cell enclosing each site (a triangle with side $\sqrt{3}\,a$; i.e., $S = 3\sqrt{3}\,a^2/4$, for a hexagonal lattice; a square with side a ($S = a^2$) for a square lattice; and a hexagon with side $2a/\sqrt{3}$ ($S = \sqrt{3}\,a^2$) for a triangular lattice). The difference $\delta\boldsymbol{\xi}$ in the sum leads to an integral in the sense of principal value.

Indeed, expression (11) does not contain any specific feature of the lattice (except the trivial coefficient S), while the specific features of ψ are very important. This expression is quite universal and isotropic, but phonons possess dispersion $\omega = 0$ in accordance with the expression derived above. Nevertheless, in contrast to the dynamics of ordinary crystal lattices, this law does not indicate the absence of the system evolution ($\partial\boldsymbol{\xi}/\partial t \neq 0$!), but only points to its power, and not exponential, nature ($\boldsymbol{\xi} \propto t$).

The reason behind this degeneracy is quite clear: Eq. (11) has the integral $\hat{d}\boldsymbol{\xi} = \mathrm{const}$ (i.e., its right-hand side is the divergence-free vector function of $\mathbf{r}$ only, which is defined by the initial strain distribution) since the macroscopic flow, which is incompressible in the given approximation (of zeroth order in a/λ) does not change the vortex density frozen in it. We can consider

[1] It is interesting to note that in [8–10], as well as in subsequent reviews, there are no indications of this fact, apparently, because the relation $\omega = 0$ does not appear as a solution in analogy with ordinary crystals.

another approach to this problem also. From the macroscopic point of view, the secondary nonlocal vortex medium under investigation is a certain smooth distribution of vorticity (curl$\mathbf{P}$ for ordinary vortices), which is not concentrated at the cores of individual vortices, but is as if "smeared" over the plane; consequently, the calculation of the macroscopic velocity field requires the convolution of the flow function with the continuous density $\rho(\mathbf{r}, t)\mathbf{e}_z$ of this vorticity. The initial homogeneous background $\rho_0 = q_0/S$ is stationary and gives $\mathbf{v}_{\text{macro}} \equiv 0$ (the actual velocity field between vortices naturally differs from zero, but the contributions from the neighbors are compensated at the sites of their location), and the situation is determined by its inhomogeneities, which are given in the linear approximation, in accordance with the continuity equation, by

$$\delta\rho = -\text{div}(\rho_0\boldsymbol{\xi}) = -\rho_0\hat{d}\boldsymbol{\xi} \tag{12}$$

(cf. (11)); i.e., only regions with nonzero bulk compression are sources of macroscopic flows. Since the motion produced by these inhomogeneities is incompressible (rotational), it leads only to a slow uniform ($\propto t$) increase in shear strains which do not affect primary sources.

Thus, the coupling between the compression and torsion of vortex lattices is broken macroscopically in the continual limit, and the corresponding evolution of long-wave perturbations has the form of inhomogeneous flows rather than running waves (which means that the term secondary hydrodynamics used here is quite adequate).

5. EFFECT OF LOW COMPRESSIBILITY

Naturally, the inclusion of the initial discreteness of the crystal structure in the next approximation changes the situation. Indeed, although the incompressibility of the field of the true microscopic velocity is a necessary feature of any system of pointlike vortices, this does not prevent the mutual convergence of pointlike neighbors and, hence, the possible increase or decrease in the macroscopic density ρ of the vorticity, i.e., effective compressibility of the macroscopic flow. This effect restores the relation between $\hat{d}\boldsymbol{\xi}$ and $\hat{r}\boldsymbol{\xi}$ so that the low (in parameter $(a/\lambda)^2$) compressibility leads to the emergence of additional terms in Eq. (11), which ensure the wave type of the phonon evolution. Obviously, these terms have a high sensitivity to the microscopic lattice symmetry due to a decisive role of discreteness, thus modifying the universal nature of Eq. (11) considerably.

The method of calculating the corresponding corrections is mathematically rather simple: it is sufficient to compare the integral

$$(1/S)\iint \mathbf{f}(x, y)dxdy$$

with the series

$$\sum_{i,j} \mathbf{f}(\mathbf{r}_{ij})$$

interpolating it (here, the vector function $\mathbf{f}$ is defined by relation (3), and i and j label now not individual vortices, but the two-dimensional array of sites; it is convenient to assume that the vortex whose motion is investigated belongs to site (0, 0)). In the first case, we are dealing with the algebraic volume (normalized to S) of the infinite (in x, y) figure bounded in z by the planes $z = 0$ and $z = f_x(x, y)$ or $z = f_y(x, y)$, while in the second case, we consider a similar volume occupied by an infinite set of right triangular, quadrangular, and hexagonal prisms whose bases (unit cells in the sense indicated above) enclose lattice sites and whose heights are equal to the values of components of $\mathbf{f}$ at the center of the base.

The difference of these quantities is obviously given by

$$\frac{1}{S}\sum_{i,j}\iint[\mathbf{f}(\mathbf{r}_{ij}) - \mathbf{f}(\mathbf{r})]dxdy, \tag{13}$$

where the integrals are taken over each unit cell. In the case of a quite smooth (as compared to a) distribution of $\mathbf{f}$ [it should be recalled that it is a combination of the actually smooth (of the scale of λ) function $\boldsymbol{\xi}$ and ψ], we can again use the expansion into the Taylor series and estimate the terms of series (13) as

$$-\frac{\Delta\mathbf{f}|_{ij}}{2S}\iint(x - x_{ij})^2dxdy = -C\Delta\mathbf{f}|_{ij}$$

(where $C = a^2/2^4$ for an elementary triangle, $a^2/(3 \times 2^3)$ for a square, and $5a^2/(3^2 \times 2^5)$ for a hexagon; the other first terms of the Taylor expansion make zero contribution due to the symmetry of unit cells). Consequently, we have

$$\sum_{i,j}\mathbf{f} \approx \frac{1}{S}\iint\mathbf{f}dxdy - C\sum_{i,j}\Delta\mathbf{f}. \tag{14}$$

On the right-hand side of this relation, we can again go over, in the continual limit, from the sum to the double integral (the error introduced in this case is of the next order of smallness). However, the correction for localized perturbations ($|\mathbf{f}| \longrightarrow 0$ for $|\mathbf{r}| \longrightarrow \infty$), as well as higher-order corrections in the parameter a/λ or a/b, identically becomes zero due to the presence of the total derivative in the integrand.

This means that in the case of a smooth distribution of $\mathbf{f}$ (which requires the smoothness of the flow function whose physical examples are unknown to us; see Section 2), the macroscopic compressibility of the lattice is small exponentially not according to a power law. This

can be seen by using another method of expansion also. Indeed, the initial discrete vorticity distribution

$$\rho_0 = q_0 \sum_{i,j} \delta(\mathbf{r} - \mathbf{r}_{ij})$$

can be presented in the form of a two-dimensional Fourier series

$$\rho_0 = \frac{q_0}{S} \sum_{i,j} \cos(\mathbf{k}_{ij}\mathbf{r} + \varphi_{ij})$$

with a discrete set of $\mathbf{k}_{ij}$ (which naturally depends on the type of the lattice) with lengths proportional to n/a ($n = 0, 1, 2, \ldots$). In this representation, the continual limit is attained by discarding the higher-order terms of the series (which corresponds to the expansion in parameter $a/n\lambda$ and not in $(a/\lambda)^n$, as before; the convergence of this expansion is worse). It was mentioned above that when expression (12) is used, the term with $n = 0$ gives Eq. (11), while the next terms lead to an exponentially small compressibility upon the convolution with the smooth function ψ.

However, all physical examples known to us are characterized by an irregular behavior of ψ at point $r = 0$. On the one hand, this circumstance ensures a power-type smallness of the compressibility of a discrete lattice in macroscopic processes, and on the other hand, it rules out the replacement of the sum on the right-hand side of (14) by an integral. Moreover, in the unit cells closest to zero (which just play the decisive role), the contribution of the discarded higher-order terms in the Taylor series for $\mathbf{f}$ is not small, and a rigorous mathematical approach requires direct summation. Unfortunately, we failed to find an analytic procedure ensuring a compact result in this case (for all possible values of ψ which differ significantly).

For this reason, we have to confine our analysis to an estimate (~1) of the correction leading to an adequate functional expression for the phonon equation, but not permitting the exact calculation of the coefficients in this expression. In other words, we obtain an order-of-magnitude estimate with true sign (which is important for the stability problem).

The following procedure appears to be simplest. In the zeroth cell with a singular point of ψ at the center, a Taylor expansion cannot be obtained, and the integral

$$-\iint \left[\delta\boldsymbol{\xi} \frac{\psi'}{r} + \frac{\mathbf{r}}{r}(\delta\boldsymbol{\xi} \cdot \mathbf{r}) \left(\frac{\psi'}{r}\right)' \right] dxdy \qquad (15)$$

must be calculated genuinely (the height of the corresponding prism is identically equal to zero). Naturally, this integral exists for power-type functions ψ only for $\alpha < 2$ (the hexagonal lattice must be treated with special care in this case), but the violation of this condition automatically transforms the physical system to the case considered in Section 3.

Further, only close cells make a contribution to the discrete correction, which is comparable with the given contribution [this also follows from the fact that the application of the Laplace operator to a slowly decreasing power function (in our terminology) transforms it into a rapidly decreasing function], for which formulas of type (14) can still be used for obtaining estimates. This allows us in our calculations, first, to expand $\delta\boldsymbol{\xi}$ into a Taylor series, confining ourselves only to quadratic terms or to linear terms in the case of a hexagonal lattice whose unit cells do not possess the central symmetry (naturally, this circumstance is also used while evaluating integral (15)), and, second, to sum again the contributions only from the nearest neighbors in correction (14).[2] This gives rise to corrections to the twisting term already present in Eq. (11), which should be omitted in comparison with this term. Naturally, the retained terms responsible for the finite compressibility of the lattice possess the same symmetry properties as the right-hand sides of Eqs. (5), (7), and (9) for the above reasons. In other words, the secondary hydrodynamics of vortex lattices possesses a high sensitivity to their local structure even in the nonlocal limit (see also [9]).

The formulas existing for hexagonal and square lattices are not presented here in view of their cumbersome appearance (since the additional terms in these formulas differ in structure from the isotropic term $\propto \hat{d}\boldsymbol{\xi}$). The quadratic lattice remains, as before, unstable and has the phonon spectrum (cf. Eq. (8))

$$\omega^2 = \psi_{\mathbf{k}}(C_1 k_x^2 k_y^2 - C_2 k^4), \qquad (16)$$

where C_1 and C_2 are positive constants $\sim q_0^2 |\psi(a)|/a^2$ (which, by the way, is well known, at least for $\psi = -\ln r$; i.e., $\psi_{\mathbf{k}} = 2\pi/k^2$; see [8, 9]). Such a dynamics is not interesting for our analysis.

On the contrary, the expression obtained for the most symmetric triangular lattice is quite isotropic and does not explicitly contain the information on the intrinsic geometry of the crystal:

$$\frac{\partial \boldsymbol{\xi}}{\partial t} = -\frac{q_0}{S}\mathbf{e}_z \times \nabla \int \hat{d}\boldsymbol{\xi}(\mathbf{r}')\psi(|\mathbf{r} - \mathbf{r}'|)d^2\mathbf{r}' + Kq_0 \nabla \hat{r}\boldsymbol{\xi}, \qquad (17)$$

where $K \sim |\psi(a)|$; i.e., its phonon spectrum is

$$\omega^2 = \frac{q_0^2}{S} K \psi_{\mathbf{k}} k^4, \qquad (18)$$

[2] We also used other methods of estimation like the replacement of $\sum \Delta f$ by the integral over a plane with deleted zero cell (whose contribution was again determined by (15)), which, in accordance with the Gauss divergence theorem, was transformed into a line integral over the boundaries of this cell. The results display no qualitative difference.

which gives $\omega \propto k$ for a power flow function with $\alpha = 0$ ($\psi_{\mathbf{k}} = 2\pi/k^2$) and $\omega \propto k^{3/2}$ for $\alpha = 1$ ($\psi_{\mathbf{k}} = 2\pi/k$). Naturally, in the former case, the result coincides with the spectrum of "Tkachenko waves" determined earlier. It was mentioned above that in contrast to the present work, the coefficient in [9, 10] was calculated exactly: it turns out that in the terms used by us, here $K = 1/8$ (however, the result is different for any other value of $\alpha \neq 0$).[3]

However, the phonon equations derived by us and presented in [9, 10] differ considerably. Proceeding from the obvious coincidence of the spectrum (18) for an ideal liquid with the standard acoustic spectrum $\omega^2 = k^2 c_S^2$, Tkachenko [9, 10] proposed that the long-wave dynamics of a triangular vortex lattice in superfluid helium can be described by the ordinary wave equation

$$\frac{\partial^2 \boldsymbol{\xi}}{\partial t^2} + c_S^2 \Delta \boldsymbol{\xi} = 0 \qquad (19)$$

(this proposition is still encountered in reviews). It can easily be seen that in the very formulation of the vortex problem itself (se Eq. (1)), the effective equation must be of the first and not second order in $\partial/\partial t$. Besides, it is obvious (see Eq. (11)) that for a given class of functions ψ, this equation must be nonlocal. It is interesting, however, that for the normal coordinates $\hat{d}\boldsymbol{\xi}$, $\hat{r}\boldsymbol{\xi}$ and for $\psi = -\ln r$, it becomes local since $\Delta \int \hat{d}\boldsymbol{\xi}\, \psi d^2\mathbf{r}' = -2\pi \hat{d}\boldsymbol{\xi}$! In other words, Eq. (19) can be rightfully used, but only for $\hat{d}\boldsymbol{\xi}$ and $\hat{r}\boldsymbol{\xi}$ and not for $\boldsymbol{\xi}$ itself. The linearity of effective equations does not imply that they can be transformed arbitrarily if their spectrum remains unchanged. From the mathematical point of view, the specific initial and boundary conditions of the problem, which are different for Eqs. (17) and (19), are equally important. Otherwise, the detailed analysis of the polarization of normal lattice vibrations carried out in [9] would not be required since Eq. (19), in contrast to Eq. (17), contains no relation between ξ_x and ξ_y.

Summing up, we may conclude that the qualitative difference in the behavior of vortex lattices is preserved in the nonlocal limit also (even with new features). In this case, their acoustic dynamics considerably depends on the form of the flow function as well as on the intrinsic geometry of the crystal.

[3] In the calculations made by Tkachenko, apart from the method in the theory of functions of complex variables, which is inapplicable to arbitrary ψ, there was one more (in our opinion, insignificant) difference from our approach. He considered waves against the background of the lattice rotation as a single entity. This is associated with the specific features of an extremely slow decrease in the function –ln r: the existence of infinitely remote boundaries of the lattice leads to its "solid-state" rotation with the angular velocity $\Omega = \pi q_0/S$. Other functions ψ, as well as –ln$r$ truncated by the Macdonald function, do not possess this property. In any case, a stationary regular state (i.e., $\delta\boldsymbol{\xi} \equiv 0$) is the exact solution of the problem for infinitely large lattices.

6. LOCAL MACROSCOPIC APPROXIMATION: INCLUSION OF NONLINEARITY

The phonon equation for a triangular vortex lattice with a slowly decreasing power function ψ truncated at distances b (say, $\psi = K_0(r/b)$) belonging to the interval $a \ll b \ll \lambda$ appears as simplest and most universal. It can be written straightaway on the basis of Eq. (17):

$$\frac{\partial \boldsymbol{\xi}}{\partial t} = -R\mathbf{e}_z \times \nabla \hat{d}\boldsymbol{\xi} + D\nabla \hat{r}\boldsymbol{\xi}. \qquad (20)$$

Here, R and D are real coefficients of the same sign, which determine the dynamics of torsion and compression, respectively, where $R = (q_0/S)\int \psi d^2\mathbf{r}$ and $D/R \sim (a/b)^2 \ll 1$. It can easily be verified that in the given case the property of virtual independence of the behavior of the details of $\psi(r)$ is restored again.

In order to determine the qualitative difference between secondary hydrodynamics (mechanics of vortex lattices) and simple propagation of acoustic waves in ordinary crystals, we analyze in greater detail the properties of this equation, which is the most symmetric and compact among the effective equations presented above (it should be noted, however, that some of these properties coincide with those of Eq. (17)). In view of the presence in this equation of small terms mentioned above, the evolution of the initial perturbations of regularity for a fairly long time is of a slow power type with a linear increase in the torsion strain of the lattice with time. However, this increase subsequently terminates, and a self-consistent oscillatory motion with $\omega = \pm\sqrt{RD}\,k^2$ sets in, in which the characteristic values of shear strain $\hat{r}\boldsymbol{\xi}$ exceed the corresponding values for "bulk compression" $\hat{d}\boldsymbol{\xi}$ by a factor of $\sqrt{R/D}$.

In the continual approach developed here for $|\boldsymbol{\xi}| \ll a$ (to be more precise, for $|\delta\boldsymbol{\xi}| \ll a$; i.e., for a small difference between the displacements of vortices separated in the given case by distances much larger than b; it is only these vortices that make a contribution to Eq. (20)), we can easily take into account nonlinear effects also. It is sufficient to supplement expansion (2) with the next (quadratic in displacement) term and repeat (using the expansion of $\delta\boldsymbol{\xi}$ into a Taylor series) the procedure of replacement of the sum by the integral, which formerly led to Eq. (11). As a result, the right-hand side of Eq. (20) acquires a new term:

$$\frac{R}{2}\mathbf{e}_z \times \nabla\left[(\hat{d}\boldsymbol{\xi})^2 + \frac{\partial \xi_\beta}{\partial x_\gamma}\frac{\partial \xi_\gamma}{\partial x_\beta}\right] \qquad (21)$$

(in this expression, the summation over recurring Greek indices assuming the values 1 and 2 is presumed).

In view of the smallness of coefficient D, it may so happen in actual practice that the role of compressibility in the lattice dynamics is much less significant than the role of nonlinearity (when the condition $b \gg \sqrt{a\lambda}$

is satisfied). In this connection, an interesting question arises: is the growth in $\hat{r}\boldsymbol{\xi}$ produced by the first (linear) term in Eq. (20) terminated in this case? In the general case, the answer to this question is negative; moreover, term (21) may even accelerate this process, bringing it independently to the explosive mode.

Indeed, for purely shear strains $\boldsymbol{\xi} = \mathbf{e}_z \times \nabla\phi$ (i.e., with $\hat{d}\boldsymbol{\xi} \equiv 0$), expressions (11) and (21) after single integration with respect to $\mathbf{r}$ (eliminating $\mathbf{e}_z \times \nabla$) lead to the equation

$$\frac{\partial\phi}{\partial t} = R\left[\frac{\partial^2\phi}{\partial x^2}\frac{\partial^2\phi}{\partial y^2} - \left(\frac{\partial^2\phi}{\partial x\partial y}\right)^2\right], \qquad (22)$$

describing the following effect in which the perturbation amplitude attains infinitely large value during a finite time (for $t \longrightarrow t_0$) near the saddle point (located at (0, 0)) of field $\boldsymbol{\xi}$:

$$\phi = -\frac{x^2y^2}{12R(t_0 - t)}.$$

Naturally, the assumption ($|\boldsymbol{\xi}| \ll a$) made while deriving the above equation ceases to hold near $t = t_0$, and, hence, we are speaking of the peculiarities of an effective equation rather than a physical system. However, first, the margin of the applicability of the formulas can be quite large and, second, purely mathematical properties of secondary hydrodynamics are also of interest to us.

The equation derived above also makes it possible to calculate a quantity important for the evolution of perturbations such as the nonlinear dispersion of the phonons under investigation. A one-dimensional travelling wave $\boldsymbol{\xi}(x - ut)$ is described by the system of equations

$$-u\frac{d\hat{d}\boldsymbol{\xi}}{dx} = D\frac{d^2\hat{r}\boldsymbol{\xi}}{dx^2},$$

$$u\frac{d\hat{r}\boldsymbol{\xi}}{dx} = R\frac{d^2[\hat{d}\boldsymbol{\xi} - (\hat{d}\boldsymbol{\xi})^2]}{dx^2},$$

which, after the elimination of two derivatives, gives

$$[\hat{d}\boldsymbol{\xi} - (\hat{d}\boldsymbol{\xi})^2]'' + \frac{u^2}{RD}\hat{d}\boldsymbol{\xi} = 0. \qquad (23)$$

Carrying out the standard procedure of expansion into a power series in the small amplitude to determine the corrections to a monochromatic wave (see, for example, [13]),

$$\hat{d}\boldsymbol{\xi} = d\cos(kx - \omega t), \quad \omega = ku = k^2\sqrt{RD} \quad (d \ll 1),$$

we can easily find that the first nonvanishing correction to frequency is $\delta\omega/\omega = -d^2/3$.

Further, discarding the assumptions of negligibly small nonlinearity, we can construct the solution in the form of a cnoidal wave. Moreover, Eq. (23) can be integrated completely in quadratures:

$$(\hat{d}\boldsymbol{\xi})'^2 = \frac{u^2}{RD}\frac{(4/3)(\hat{d}\boldsymbol{\xi})^3 - (\hat{d}\boldsymbol{\xi})^2 + \text{const}}{(2\hat{d}\boldsymbol{\xi} - 1)^2}.$$

The subsequent definition of $\hat{r}\boldsymbol{\xi}$ and $\boldsymbol{\xi}$ itself can be obtained by the simple integration of the obtained equations.

It is interesting to note that an equation with the structure similar to Eq. (23) (but with a cubic nonlinearity) was derived in [14] for a (potential!) electric field of high-frequency (near ω_{pi}) ion–sound waves in a plasma (these waves possess a completely different dispersion: in fact, the temporal (or frequency) and dispersion terms exchange places, and it is only the rather peculiar nonlinearity (21) that restores *status quo*, naturally, in the mathematical sense)

The simple examples considered above show that in spite of its nonstandard form, the effective equation proposed here is quite suitable for analytic investigations and sufficiently informative.

7. INTERACTION OF THE LATTICE WITH NONUNIFORM FLOWS

We continue the study of peculiarities of secondary hydrodynamics by analyzing the interaction between an ensemble of point vortices and distributed macroscopic vortices. The latter produce nonuniform zero-divergence flows of a viscous medium, which obviously entrain their pointlike analogues and, hence, perturb the initial regularity of lattices. These flows must be additively introduced into the right-hand side of Eq. (20) (the time derivative of the displacement of each vortex is equal to the velocity of the flow in the region of its location irrespective of the source of $\mathbf{v}$). We disregard the opposite effect of the lattice dynamics on a distributed vortex, assuming that its vorticity density ρ and, hence, the flow created by it ($\mathbf{v} = \mathbf{e}_z \times \nabla\int\rho\psi d^2\mathbf{r}' \propto \mathbf{e}_z \times \nabla\rho$; it is interesting to note that any smooth distribution of ρ in our case is quite stationary; cf. the evolution of the magnetic field in the electron plasma [7]) are defined by functions of $\mathbf{r}$. Ultimately, we are interested in the typical features of the dynamics of vortex crystals, and their phonon response to nonuniform flows appears in concordance with the problem formulated here. In a multicomponent plasma, such a mode with present and "external" incompressible flows can be realized in pure form using the beams of charged particles piercing the plasma (see [7]).

We choose the simple geometry of such a flow:

$$\mathbf{v} = v(y)\mathbf{e}_x = v_0\cos(ky)\mathbf{e}_x.$$

The linear (which means that the general case can be analyzed using the expansion into a Fourier integral) equation (20) with the modification indicated above,

rewritten in normal coordinates, is transformed into the system

$$\frac{\partial \hat{d}\boldsymbol{\xi}}{\partial t} = D\Delta\hat{r}\boldsymbol{\xi}, \tag{24}$$

$$\frac{\partial \hat{r}\boldsymbol{\xi}}{\partial t} = -R\Delta\hat{d}\boldsymbol{\xi} + kv_0\sin(ky). \tag{25}$$

For zero initial conditions (i.e., for zero lattice deformation at $t = 0$), this system has the following simple solution:

$$\hat{d}\boldsymbol{\xi} = \frac{v_0}{kR}[\cos(\omega t) - 1]\sin(ky),$$

$$\hat{r}\boldsymbol{\xi} = \frac{v_0}{k\sqrt{RD}}\sin(\omega t)\sin(ky),$$

whence

$$\boldsymbol{\xi} = \frac{v_0}{\omega}\left\{\sin(\omega t)\mathbf{e}_x + \sqrt{\frac{D}{R}}[1 - \cos(\omega t)]\mathbf{e}_y\right\}\cos(ky),$$

where, naturally, $\omega = \sqrt{RD}\,k^2$. A similar type of solution with strains proportional to $v(y)$ is also observed for $v(y) \propto \exp(-ky)$. For arbitrary initial conditions, the answers are slightly more complicated.

Thus, nonuniform steady-state flows of the initial continuous medium indeed excite periodic in time vibrations of the point-vortex lattices frozen in it. Thus, although the concept of elasticity of vortex crystals, which was introduced by Tkachenko [10] (naturally, for the nonlocal case), is admissible, it differs significantly from conventional analogues.

8. DYNAMICS OF VORTEX CHAINS

Another example confirming the generality of the behavior of vortex ensembles observed along with specific features determined by their physical origin (flow function) is the evolution of linear one-dimensional chains, i.e., infinitely long rows of identical vortices (discrete analogues of tangential discontinuities in conventional hydrodynamics) arranged, for instance, along the straight line $y = 0$ at a distance a from one another. The incompressibility of the macroscopic flow in this case does not lead to a degeneracy even in the nonlocal limit (the two-dimensional vector $\boldsymbol{\xi}(x, t)$ associated with the abscissa axis is not necessarily a zero-divergence vector); for this reason, the cases with different functions $\psi(r)$ are more alike than in the case of two-dimensional ensembles.

The version with the dominating effect of the two nearest neighbors is now described by the system of equations (the condition $\lambda \gg a$ is assumed to be satisfied as before)

$$\frac{\partial \xi_x}{\partial t} = q_0 a^2 A\frac{\partial^2 \xi_y}{\partial x^2}, \tag{26}$$

$$\frac{\partial \xi_y}{\partial t} = q_0 a^2 B\frac{\partial^2 \xi_x}{\partial x^2} \tag{27}$$

(it is assumed that $\partial^2\psi/\partial y^2|_{y=0} = (1/x)\partial\psi/\partial x|_{y=0}$), while the nonlocal case (in which the boundary between different modes passes through the value $\alpha = 1$) is described by

$$\frac{\partial \xi_x}{\partial t} = -q_0 a\int \xi_y(x - x')\frac{1}{x'}\frac{\partial \psi(x')}{\partial x'}\bigg|_{y=0} dx', \tag{28}$$

$$\frac{\partial \xi_y}{\partial t} = -q_0 a\int \xi_x(x - x')\frac{\partial^2 \psi(x')}{\partial x'^2}\bigg|_{y=0} dx' \tag{29}$$

(in contrast to the previous formulas, the substitution $x \longrightarrow r$ is not used in $\psi|_{y=0}$ since x may assume negative values also), where the integrals are considered as the principal values in Cauchy's sense.

A transition from Eqs. (28), (29) to the local macroscopic version is trivial:

$$\frac{\partial \xi_x}{\partial t} = q_0 a C_3 \xi_y, \tag{30}$$

$$\frac{\partial \xi_y}{\partial t} = -q_0 a C_4\frac{\partial^2 \xi_x}{\partial x^2}. \tag{31}$$

Here,

$$C_3 = -\int\frac{1}{x}\frac{\partial \psi}{\partial x}\bigg|_{y=0} dx, \quad C_4 = \int \psi|_{y=0} dx$$

are two positive constants connected through the relation $C_4 \sim C_3 b^2$ (the possible divergence at zero is removed by the standard mathematical interpretations of the "principal value").

In all cases, chains are unstable formations. For $\psi = -\ln r$, the increment following from Eqs. (28), (29) naturally coincides with the classical hydrodynamic expression (see, for example, [1]; in fact the methods of the theory of functions of complex variables provide answers for an arbitrary value of λ).

9. CONCLUSIONS

Let us summarize the main results obtained in this work.

1. Linear equations describing the evolution of long-wave perturbations in various two-dimensional lattices are derived for various classes of vortices (defined by peculiarities of flow functions). These equations are used to analyze the dispersion and symmetry properties

of these perturbations, which are similar to acoustic waves in ordinary crystals.

2. The reasons behind the decisive role of microscopic discreteness of secondary vortex media in the formation of acoustic spectra even in the macroscopic limit are determined.

3. For the stablest and most symmetric triangular vortex lattice with flow functions from the local macroscopic class, a nonlinear phonon equation (which possesses a peculiar hierarchy of nonlinearity and dispersion) is derived and its simplest properties are investigated.

4. The response of vortex lattices to the attempt of their deformation by nonuniform flows of the primary medium, which is characterized by the phonon generation even in the case of a steady-state external action in analogy with the sounding of a violin string caused by a uniform motion of the bow.

In order to avoid misunderstanding, it should be emphasized once again that the approximate estimate obtained by us for one of the coefficients in Eqs. (17) and (20) does not affect in any way their appearance or qualitative or quantitative properties of these equations. For any specific function ψ, this coefficient can be calculated quite rapidly with the help of a computer.

Thus, we have demonstrated that the developed ideology, which declares a unified approach to vortex problems, indeed makes it possible to determine the general regularities of the ensemble of pointlike vortices, simultaneously indicating the high sensitivity of the pattern of the processes to physical and geometrical features of specific systems. The effective equations derived by us for the macroscopic evolution of regular lattices may be successfully used for studying the specific features of various phenomena associated with external effects exerted on these peculiar crystals. The corresponding dynamics is found to differ considerably (and often qualitatively) from the phonon dynamics of conventional solids.

ACKNOWLEDGMENTS

The authors are deeply indebted to I.A. Ivonin for fruitful discussions stimulating our investigations.

This work was partly financed by the Ministry of Science of the Russian Federation (program "Nonlinear Dynamics") and the INTAS Foundation (grant no. 97-0021).

REFERENCES

1. P. G. Saffman, *Vortex Dynamics* (Nauchnyĭ Mir, Moscow, 2000; Gambridge University Press, 1992).
2. V. V. Kozlov, *General Theory of Vortexes* (Udmurtskiĭ Univ., Izhevsk, 1998), p. 1.
3. G. Blatter, M. V. Feigel'man, V. B. Geshkenbein, *et al.*, Rev. Mod. Phys. **66**, 1125 (1994).
4. B. Nienhuis, in *Phase Transitions and Critical Phenomena*, Ed. by C. Domb and J. L. Lebowitz (Academic, New York, 1987), Vol. 11, p. 1.
5. R. N. Kraichnan and D. Montgomery, Rep. Prog. Phys. **43**, 547 (1980).
6. V. I. Petviashvili and V. V. Yan'kov, in *Reviews of Plasma Physics*, Ed. by B. B. Kadomtsev (Énergoatomizdat, Moscow, 1985; Consultants Bureau, New York, 1987), Vol. 14, p. 3.
7. A. S. Kingsep, K. V. Chukbar, and V. V. Yan'kov, *Reviews of Plasma Physics*, Ed. by B. B. Kadomtsev (Énergoatomizdat, Moscow, 1987; Consultants Bureau, New York, 1990), Vol. 16, p. 209.
8. V. K. Tkachenko, Zh. Éksp. Teor. Fiz. **49**, 1875 (1965) [Sov. Phys. JETP **22**, 1282 (1966)].
9. V. K. Tkachenko, Zh. Éksp. Teor. Fiz. **50**, 1573 (1966) [Sov. Phys. JETP **23**, 1049 (1966)].
10. V. K. Tkachenko, Zh. Éksp. Teor. Fiz. **56**, 1763 (1969) [Sov. Phys. JETP **29**, 945 (1969)].
11. A. A. Abrikosov, *Fundamentals of the Theory of Metals* (Nauka, Moscow, 1987; North-Holland, Amsterdam, 1988), p. 1.
12. K. V. Chukbar, Fiz. Plazmy **25**, 83 (1999) [Plasma Phys. Rep. **25**, 77 (1999)].
13. G. B. Whitham, *Linear and Nonlinear Waves* (Wiley, New York, 1974; Mir, Moscow, 1977), Chap. 13.
14. L. I. Rudakov and V. N. Tsytovich, Zh. Éksp. Teor. Fiz. **75**, 1618 (1978) [Sov. Phys. JETP **48**, 816 (1978)].

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 93, No. 1, 2001, pp. 136–142.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 1, 2001, pp. 156–163.

SOLIDS
Structure

Distribution Function for Random Interaction Fields in Disordered Magnets: Spin and Macrospin Glass

V. I. Belokon* and K. V. Nefedev**

Far East State University, ul. Sukhanova 8, Vladivostok, 690000 Russia

**e-mail: belokon@ifit.phys.dvgu.ru*

***e-mail: nefedev@ifit.phys.dvgu.ru*

Received February 27, 2001

Abstract—Systems with an arbitrary dependence of exchange integral on the distance between atoms which are randomly scattered in an amorphous substance are investigated by averaging over random fields of interaction in the framework of the Ising model. This method is also used for describing long-term magnetization relaxation in a system of single-domain particles scattered in a nonmagnetic matrix. Random field distribution functions are obtained for the dipole–dipole and the Ruderman–Kittel–Kasuya–Yoshida (RKKY) interactions. Long-term relaxation in macrospin glasses is investigated. *© 2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

It is generally accepted [1–5] that the most typical features of the spin-glass state are the irreversibility and long-term relaxation of the magnetization, which are associated with nonergodicity of the spin glass phase. Various versions of constructing the physics of such systems are based on the assumption concerning the hierarchical structure of the valley space leading to the hierarchical structure of relaxation times. Theoretical descriptions of the properties of spin glasses are often based on the assumption that exchange integrals are random quantities obeying the normal distribution law. Most results were obtained for lattice models, and their extension to amorphous systems appears at first glance unjustified.

In this paper, we consider the possibility of applying the method of random interaction fields to amorphous media with an arbitrary law of interaction of particles; systems of particles with the Ruderman–Kittel–Kasuya–Yoshida (RKKY) interaction and the dipole–dipole interaction are also considered. The irreversibility and long-term relaxation in systems of interacting single-domain particles (macrospin glass) are considered separately without using the ideas of the hierarchical structure of states. This work is a continuation of the analysis of the possibility of applying the random field method in the theory of ferromagnetism of inhomogeneous systems started in [6, 7].

2. DISTRIBUTION FUNCTION FOR A RANDOM INTERACTION FIELD

The general scheme of computation of the random field distribution function was considered by several authors [8–10]. However, we repeat here the main points of its calculation, taking into account the intrinsic volume of particles constituting a system. We assume that the component H_i of the field created at the origin by a single arbitrary particle (located at the point with coordinate $\mathbf{r}_i$ and having spin $\mathbf{S}_i$) along the z axis (symmetry axis in the Ising model) is defined by the law

$$H_i = \varphi(\mathbf{r}_i, \mathbf{S}_i). \quad (1)$$

If the distribution of particles over $\mathbf{r}_i$ and $\mathbf{S}_i$ is known, the distribution function of the interaction field at a particle located at the origin is a δ-function of the type

$$\delta\left[H_i - \sum_i \varphi(\mathbf{r}_i, \mathbf{S}_i)\right].$$

In turn, the probability of this coordinate distribution of particles in an amorphous body taking into account the intrinsic volume of the particles is defined by the product

$$\frac{dV_1}{V}\frac{dV_2}{V - V_0}\cdots\frac{dV_N}{V - NV_0} \approx \frac{1}{V^N}\prod_{i=1}^{N} dV_i.$$

Here,

$$\frac{1}{V^N}\prod_i \frac{1}{1 - iV_0/V} \approx V^{-N}\prod_i (1 + i(V_0/V))$$

$$= \frac{1}{V^N}\frac{(V_0/V)^{N+1}\Gamma(N + 1 + V/V_0)}{(V/V_0)\Gamma(1 + V/V_0)}.$$

1063-7761/01/9301-0136$21.00

For $V/V_0 \gg 1$, we can use the Stirling formula, which readily gives the following expression for $N \longrightarrow \infty$:

$$\frac{1}{V^N}\prod_k \frac{1}{1-kV_0/V} \approx \left(\frac{V}{1+c}\right)^N = \frac{1}{v^N},$$

$$c = \frac{NV_0}{V} \ll 1,$$

where V_0 is the intrinsic volume of a particle, which must be taken into account for systems with a considerable concentration of scattered ferromagnetic particles.

The particle distribution $\prod_{i=1}^N \tau(\mathbf{S}_i)d\mathbf{S}_i$ over spin directions in the Ising model is also regarded as independent; in this case,

$$\tau(\mathbf{S}_i) = \frac{1}{S_i^2 \sin\gamma}\frac{1}{2\pi}\delta(S_i - S_0)[\alpha\delta(\gamma_i) + \beta\delta(\gamma_i - \pi)],$$

$$\alpha + \beta = 1.$$

Taking into account the above formulas, we can write the distribution function for field H in the form

$$W(H) = \frac{1}{v^N}\int\int\delta\left(H - \sum_{i=1}^N \varphi_i\right)\prod_{i=1}^N \tau(\mathbf{S}_i)d\mathbf{S}_i dV_i. \quad (2)$$

The characteristic function

$$A(\rho) = \int W(H)\exp(i\rho H)dH$$

can be written in the form

$$A(\rho) = \frac{1}{v^N}\int\exp\left(i\rho\sum_i \varphi_i\right)\prod_{i=1}^N \tau(\mathbf{S}_i)d\mathbf{S}_i dV_i.$$

Taking into account normalization conditions, we can write $A(\rho)$ in the form

$$A(\rho) = \left\{1 - \frac{n^*}{N}\int[1 - \exp(i\rho\varphi)]\tau(\mathbf{S})d\mathbf{S}dV\right\}^N,$$

where

$$n^* = \frac{N}{V} = \frac{N(1+c)}{V} = n(1+c)$$

and n is the number of particles per unit volume. In the limit $N \longrightarrow \infty$, we have

$$A(\rho) \longrightarrow \exp\{-F(\rho)\}.$$

Thus, the random interaction field distribution function has the form

$$W(H) = \int_{-\infty}^{\infty}\exp\{-F(\rho)\}\exp(-i\rho H)d\rho,$$
$$F(\rho) = n^*\int dV\int[1 - \exp(i\rho\varphi)]\tau(\mathbf{S})d\mathbf{S} \quad (3)$$
$$= \alpha I_- + \beta I_+,$$

where

$$I_\mp = n^*\int_V [1 - \exp(\mp i\rho\varphi)]dV.$$

Here, the sign reversal of the field upon the spin flip from direction α (spin up) to direction β (spin down) is taken into account. Integration in Eq. (3) should be carried out between $r = r_0$ (particle size) to $r = R$ (sample size). If the function $\varphi(r)$ decreases rapidly, we can proceed to the limit $R \longrightarrow \infty$, while in the case of its slow decrease, the shape of the sample (demagnetizing factor) must taken into consideration.

The structure of $F(\rho)$ is such that the main contribution to the integral in Eq. (3) comes from values of $F(\rho)$ close to zero. Indeed,

$$F(\rho) = i(\alpha - \beta)H_0\rho - \frac{B^2}{4}\rho^2 + \ldots, \quad (4)$$

where

$$H_0 = -n^*\int_V \varphi(r)dV, \quad \frac{B^2}{4} = \frac{n^*}{2!}\int_V \varphi^2(r)dV,$$

and so on. Consequently, rapid oscillations of the integrand for $\rho \gg 1$ lead to a considerable decrease of the contribution to the integral.

Confining the analysis to the first three terms in the expansion of the exponential, we obtain from formula (3)

$$A(\rho) = \exp\left[-i(\alpha - \beta)H_0\rho - \frac{B^2}{4}\rho^2\right], \quad (5)$$

whence

$$W(H) = \frac{1}{\sqrt{\pi}B}\exp\left\{-\frac{[H - H_0(\alpha - \beta)]^2}{B^2}\right\}. \quad (6)$$

3. SELF-CONSISTENT EQUATION FOR MAGNETIZATION

In the thermodynamic equilibrium state, the mean value of magnetization $M = 0$ in the Ising model is obviously obtained by the Gibbs-distribution averaging and the averaging over configurations:

$$M = \int\tanh\left(\frac{mH}{kT}\right)W(H)dH, \quad (7)$$

where m is the magnetic moment of a particle. Thus, in zero external field, the equation for magnetization in the equilibrium state has the form

$$M = \frac{1}{\sqrt{\pi}B}\int \tanh\left(\frac{m(H + H_0M)}{kT}\right)\exp\left(-\frac{H^2}{B^2}\right)dH. \quad (8)$$

Simple estimates can be obtained by replacing the Gaussian distribution function by the approximate function $f(H)$:

$$f(H) = \begin{cases} 0, & H > B, \quad H < -B, \\ \dfrac{1}{2B}, & -B < H < B. \end{cases}$$

For small values of M, we obtain

$$M \approx \frac{1}{2B}\int_{-B}^{B} \tanh\left(\frac{m(H + H_0M)}{kT}\right)dH \approx M\frac{H_0}{B}\tanh\frac{mB}{kT}. \quad (9)$$

This means that Eq. (8) has a nonzero solution (ferroglass) when

$$\frac{H_0}{B}\tanh\frac{mB}{kT} > 1, \quad \frac{H_0}{B} > 1. \quad (10)$$

For $H_0/B < 1$, the initial susceptibility in field h is given by

$$\chi = \frac{\partial M}{\partial h} \approx \frac{1}{B}\tanh\frac{mB}{kT}. \quad (11)$$

For high temperatures, for $mB/kT \ll 1$, we have

$$\chi \approx \frac{m}{kT},$$

which corresponds to paramagnetic susceptibility. For $mB/kT \gg 1$, susceptibility is given by

$$\chi \approx \frac{1}{B}$$

and is independent of temperature. In this case, spins are "frozen" in the random exchange interaction fields. The maximum rate of susceptibility variation is observed at $T^* = mB/k$; T^* can be treated as the temperature of transition to the spin glass phase.

4. RKKY INTERACTION AND SPIN GLASS

We apply this approach to a system of atoms coupled through the RKKY interaction [11]

$$\varphi(r) = -b\frac{k_Fr\cos(k_Fr) - \sin(k_Fr)}{(k_Fr)^4}, \quad (12)$$

where b is a certain coefficient having the dimensions of field, k_F is the momentum on the Fermi surface, and r is the separation between interacting atoms. Integration is carried out from 1 to k_FR. In this case,

$$I_{\mp} = \frac{n^*}{k_F^3} \times \int_1^{k_FR}\int_0^{\pi}\int_0^{2\pi}\left\{1 - \exp\left(\mp i\rho b\frac{k_Fr\cos(k_Fr) - \sin(k_Fr)}{(k_Fr)^4}\right)\right\} \times (k_Fr)^2 d(k_Fr)\sin\vartheta d\vartheta d\varphi.$$

After integration, we obtain the following relations for H_0 and B^2:

$$H_0 = -\frac{4\pi n^* b}{k_F^3}\frac{\sin(k_FR) - 0.84k_FR}{k_FR},$$

$$B^2 \approx \frac{4n^*b^2}{15k_F^8R^5}\{-5k_F^2R^2 + 6k_FR\sin(2k_FR) - 3 + 3\cos(2k_FR) - k_F^2R^2\cos(2k_FR) + k_F^3R^3\sin(2k_FR) + 2k_F^4R^4\cos(2k_FR) + 4k_F^5R^5\mathrm{Si}(2k_FR) - 3k_F^5R^5\},$$

where Si is the sine integral. Proceeding to the limit $k_FR \longrightarrow \infty$, we obtain

$$H_0 \approx 10\frac{n^*b}{k_F^3}, \quad B^2 \approx 2.5\frac{n^*b^2}{k_F^3}, \quad \frac{H_0}{B} \approx 6\sqrt{\frac{n^*}{k_F^3}}.$$

The last relation together with relations (10) makes it possible to construct the theoretical magnetic phase diagram.

5. DIPOLE–DIPOLE INTERACTION IN A SYSTEM OF SCATTERED MAGNETIC GRAINS: MACROSPIN GLASS

Let us consider a system of small ferromagnetic particles scattered in a nonmagnetic matrix. The magnetic moment distribution function for such particles has the form

$$\tau(\mathbf{m})d\mathbf{m} = f(m)w(\gamma, \psi)d\gamma d\psi,$$

where angles γ and ψ specify the orientation of $\mathbf{m}$ relative to a chosen direction. In this case,

$$\varphi = -\frac{m\cos\gamma}{r^3} + \frac{3m(\sin\gamma\cos\psi\sin\vartheta + \cos\gamma\cos\vartheta)\cos\vartheta}{r^3}.$$

In the Ising model, we have

$$w(\gamma, \psi) = \frac{1}{2\pi}\{\alpha\delta(\gamma) + \beta\delta(\gamma - \pi)\}. \quad (13)$$

and angle ϑ determines the orientation of vector $\mathbf{r}$ in the spherical system of coordinates.

In the case when the shape of a magnetic sample is not spherical, but, say, ellipsoidal, we must take into account the dependence of R on ϑ:

$$R(\vartheta) = \frac{b}{\sqrt{1+\epsilon b^2\sin^2\vartheta}} \approx b\left(1+\frac{1}{2}\epsilon b^2\sin^2\vartheta\right)^{-1} \approx b\left(1-\frac{1}{2}\epsilon b^2\sin^2\vartheta\right), \quad (14)$$

where b is the major semiaxis and ϵ is the eccentricity.

The parameters H_0 and B^2 of the distribution function $W(H)$ for a system of magnetic particles (ferro- or ferrimagnetic grains) with the dipole–dipole interaction can be calculated approximately using formulas (4):

$$H_0 \approx \frac{8}{15}\pi n^*\bar{m}\epsilon^2, \quad B^2 \approx \frac{4\pi}{r_0^3}n^*\bar{m}^2, \quad (15)$$

where $\bar{m} = \int mf(m)dm$ and r_0 is the size of a magnetic particle. In such a system, we have

$$\frac{H_0}{B} \approx 0.5\epsilon^2\sqrt{n^*r_0^3},$$

which means that $H_0/B \approx 0.2\sqrt{c} < 1$ even for $\epsilon \approx 1$ (it should be recalled that c is the volume concentration of the ferromagnet).

Thus, ferromagnetic ordering due to dipole–dipole interaction in an amorphous substance is ruled out. A similar result was obtained by Meilikhov [12] for particles located at the sites of a cubic lattice.

At the same time, a transition to the spin-glass state (macrospin glass) at temperature $T^* \sim \bar{m}B/k$ is possible if the relaxation time, which is determined primarily by the critical field and the volume of a ferromagnetic particle, is quite small. Since the system of single-domain particles makes it possible to interpret visually the long-term relaxation in macrospin glass, we consider the transition to the equilibrium state in greater detail.

An analysis of the magnetic properties of such a system requires primarily the knowledge of their distribution over the magnetic moments m and over the critical fields H_c of magnetization reversal. The field of magnetization reversal of a single-domain particle is determined by its shape (for strongly magnetic materials), crystallographic anisotropy, and anisotropy of stresses which inevitably emerge during the interaction of the particle with the nonmagnetic matrix. For a uniaxial particle, the magnetization reversal condition can be presented in the form

$$|\mathbf{H}+\mathbf{h}|(\sin^{2/3}\theta+\cos^{2/3}\theta)^{-3/2} > H_c,$$

where θ is the angle between field $\mathbf{H}+\mathbf{h}$ and the easy axis, $\mathbf{H}$ is the random field of interaction, and $\mathbf{h}$ is the applied field.

For fields h small as compared to H_c, these relations can be written in the following approximate form:

$$h > H_c - H, \text{ for } \cos(\mathbf{h},\mathbf{H}) > 0, \quad (16)$$

$$h > H_c + H, \text{ for } \cos(\mathbf{h},\mathbf{H}) < 0. \quad (17)$$

Accordingly, an individual hysteresis loop for each particle from the subsets (16) and (17) is characterized by the fields $a = H_c - H$, $b = H_c + H$, and $a = H_c + H$, $b = H_c - H$, respectively. For a known distribution of particles over the magnitude of the interaction field, $g(H)$, which can be obtained if we know $W(H)$, we can easily calculate the number density of particles on the Preisach–Neel (a, b) phase diagram [13]. The normalization condition for particles from subset (16) has the form

$$\int_0^\infty\int f(H_c)\gamma(H)dH_cdH = \frac{1}{2}.$$

Going over to new variables $a = H_c - H$ and $b = H_c + H$, we obtain

$$\int_0^\infty\int_{2H_c-b}^\infty f\left(\frac{a+b}{2}\right)\gamma\left(\frac{b-a}{2}\right)\Delta(a,b)dadb = \frac{1}{2},$$

where the Jacobian $\Delta(a, b) = 1/2$. Thus, the quantity

$$p(a,b) = \frac{1}{2}f\left(\frac{a+b}{2}\right)\gamma\left(\frac{b-a}{2}\right)$$

is the number density of points on the Preisach–Neel diagram, corresponding to particles with the critical fields a, b. Axis $a = b$ is a symmetry axis, and hence, we can confine the subsequent analysis to the region $a > b$ only.

6. LONG-TERM RELAXATION AND IRREVERSIBILITY

The estimation of the relaxation time for single-domain uniaxial particles with an elementary hysteresis cycle and critical fields a and b in an external magnetic field h ($h < a, b$) is based on the fact that the fluctuation probability is determined by the minimum work which must be done to rotate the magnetic moment through the angle required for a subsequent spontaneous irreversible U-process. If $h \ll a, b$, the relaxation time is defined by the approximate formula

$$\frac{1}{\tau_{a,b}} = f_0\left\{\exp\left[-\frac{m(a-h)}{2kT}\right]+\exp\left[-\frac{m(a+h)}{2kT}\right]\right\}, \quad (18)$$

where $f_0 \sim 10^{10}$–10^{12} is the frequency factor.

Equation (18) defines the relaxation time isolines on the Preisach–Neel (a, b) phase diagram. It can easily be seen that particles with larger and larger values of a and b are gradually involved in the process of establishment

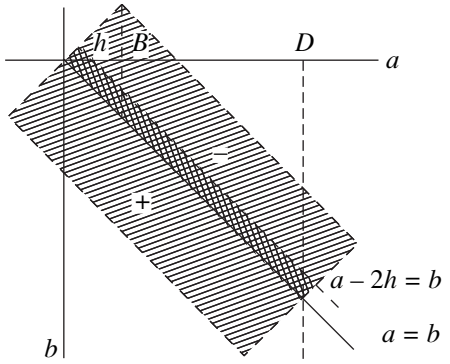

Fig. 1. Preisach–Neel (*a*, *b*) diagram.

of equilibrium. An order-of-magnitude increase in the relaxation time corresponds to the change in the value of *a* (*b*) by

$$\Delta a = \frac{\ln 10}{\alpha}, \quad \alpha = \frac{m}{2kT}.$$

Naturally, the spread in the particle distribution over critical fields and volumes leads to long-term relaxation, and equilibrium in the system may not be established in principle over a reasonable time interval. Indeed, if

$$\ln(f_0\tau) \sim \frac{ma}{2kT}$$

and $\tau \sim 1$, a doubling of *m* or *a* increases the relaxation time to $\tau \approx 10^{10}$ s for $f_0 \approx 10^{10}$ s^{-1}.

The possibility of independent establishment of equilibrium in a subsystem of particles with different values of *a* and *b* is of fundamental importance since in this case the magnetic susceptibility, the residual magnetization, and other characteristics of the system can be calculated quite easily in the case when the system as a whole is far from equilibrium. In order to simplify calculations, we assume that the function $p(a, b) = p$ is constant in the hatched region in Fig. 1, *B* coincides in order of magnitude with the maximum fields of interaction, and *D* are the maximum critical fields of particles in the system. The normalization condition in this case has the form

$$2pBD = N/2.$$

As the initial conditions, we consider the system in the so-called zero-field state which can be attained by applying an alternating field $\tilde{h}$ whose amplitude decreases from $\tilde{h}_{\max} > D$ to zero. In this case, all the particles which belong to the region $a > b$ on the phase diagram are in the state with a conditionally negative magnetization, while those from the region $a < b$ have a positive magnetization. The total magnetic moment is equal to zero. After the removal of the field, thermodynamic equilibrium will gradually set in, which corresponds to the emergence on the (*a*, *b*) diagram of positively magnetized particles whose critical fields correspond to region $a > b$ and negatively magnetized particles from region $a < b$, the total zero magnetic moment of the system being preserved. For particles with the given magnetic moment *m*, such a process will occupy, by the time *t*, the region

$$a, b \lesssim \frac{2kT}{m}\ln(f_0 t). \tag{19}$$

For particles with critical fields *a* and *b*, $a > b$, the extent of "mixing" of the signs "–" and "+" is determined by the transition probabilities:

$$\frac{n_+}{n_-} = \frac{e^{-\alpha a}}{e^{-\alpha b}}.$$

Obviously, in view of symmetry relative to the line $a = b$, such a mixing in the region $b > a$ leads to the compensation of the emerging magnetic moment.

The application of the external magnetic field *h* displaces the symmetry line to the region $a - h = b + h$; i.e., the compensation is violated. This result is quantitatively equivalent to the sign reversal in the magnetic moment of the particles represented in Fig. 1 by a point from the region with double hatching,

$$h < a < \frac{1}{\alpha}\ln(f_0 t), \quad a - 2h \le b \le a.$$

In our opinion, this circumstance is extremely important since only a small fraction of particles (h/B) take part in the process of magnetization in a low field *h*, namely, the particles whose fields of interaction are close to zero; consequently, the change in the orientation of the magnetic moment of such particles weakly affects the state of the entire system as a whole.

Thus, for low fields

$$h \ll \frac{1}{\alpha}\ln(f_0 t),$$

an additional magnetic moment

$$M_v = 4mph\left[\frac{1}{\alpha}\ln(f_0 t) - h\right] \approx \frac{4mph}{\alpha}\ln(f_0 t) \tag{20}$$

emerges due to thermal fluctuations. For a given value of α, saturation will be attained for

$$\frac{1}{\alpha}\ln(f_0 t) \approx D.$$

By the instant t^* of the first measurement, which is determined by the potentialities of the experimental setup, we have

$$M^* = M_0 + 4mph\left[\frac{1}{\alpha}\ln(f_0 t^*) - h\right], \qquad M_0 = mph^2. \tag{21}$$

The relatively high value of M_0 is possible only if there exist particles with a large magnetic moment (large value of α_0) such that

$$\frac{1}{\alpha_0}\ln(f_0 t^*) < h. \tag{22}$$

Since we are mainly interested in long-term relaxation, we will henceforth put $M_0 \ll M_v \equiv M$.

Let us now consider the temperature dependence of magnetic susceptibility $\partial M/\partial h = \chi$,

$$\chi = \frac{4mp}{\alpha}\ln(f_0 t) = \frac{kT}{4}\frac{N}{BD}\ln(f_0 t^*) = \frac{N\ln(f_0 t^*)kT}{4BD}, \tag{23}$$

provided that

$$\frac{1}{\alpha(T)}\ln(f_0 t^*) < D(T), \quad B(T) < h. \tag{24}$$

As the temperature increases further, the following two scenarios of susceptibility variation may take place.

1. At a certain temperature $T = T_B$, the value of $B(T_B)$ becomes equal to h; however, in this case,

$$\frac{1}{\alpha(T_B)}\ln(f_0 t^*) < D(T_B).$$

In this case, the susceptibility increases for $T > T_B$ since

$$\chi = \frac{N\ln(f_0 t^*)kT}{4hD(T)}, \quad D(T) \longrightarrow 0 \text{ for } T \longrightarrow T_C,$$

where T_C is the Curie point. At the temperature T_D for which

$$\frac{1}{\alpha T_D}\ln(f_0 t^*) = D(T_D),$$

the susceptibility attains its maximum value since all the particles are "involved" in the process:

$$\chi_{\max} = m(T_D)\frac{N}{h}.$$

A further decrease in χ upon a decrease in $T > T_D$ is associated with a decrease in m for $T \longrightarrow T_C$.

2. $T_D < T_B$. In this case, at $T = T_D$, we have

$$\chi = \frac{m(T_D)N}{2B(T_D)};$$

the value of χ remains constant up to $T = T_B (B(T_B) = h)$ and decreases further together with $m(T)$. The behavior of susceptibility as a function of temperature is presented schematically in Fig. 2. These formulas remain valid for cooling also if it occurs in zero field. If, however, the field has not been removed, the particles whose magnetic moments are blocked during cooling preserve their magnetic moment even after the "departure" of the corresponding points on the (a, b) diagram from the region $a - 2h \le b \le a$, thus increasing the susceptibility. Since the fraction of particles participating in the magnetization at a given temperature T is approximately equal to the ratio $h/B(T)$, averaging of this ratio upon cooling from T to T_0 in the field h gives

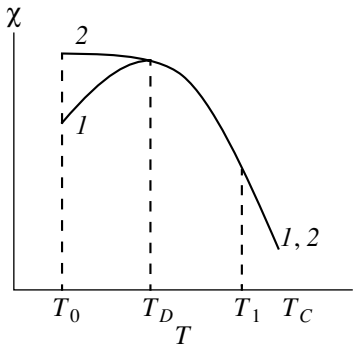

Fig. 2. Theoretical curve describing the temperature dependence of susceptibility: zero field cooling (*1*) and cooling in the field (*2*).

$$\left\langle \frac{h}{B(T)} \right\rangle = \frac{1}{T_1 - T_0}\int_{T_0}^{T_1}\frac{h}{B(T)}dT > \frac{h}{B(T_0)}.$$

In the simplest case, when

$$\frac{B(T)}{B(T_0)} = \sqrt{\frac{T_C - T}{T_C - T_0}},$$

$$\left\langle \frac{h}{B} \right\rangle = \frac{1}{T_1 - T_0}\int_{T_0}^{T_1}\frac{h}{B(T)}dT$$

$$= \frac{h}{T_1 - T_0}\int_{T_0}^{T_1}\frac{dT}{B(T_0)}\sqrt{\frac{T_C - T_0}{T_C - T}}$$

$$= \frac{2h}{B(T_0)}\frac{T_C - T_0}{T_1 - T_0}\left(1 - \sqrt{\frac{T_C - T_1}{T_C - T_0}}\right),$$

we have for T_1 close to T_C

$$\left\langle \frac{h}{B} \right\rangle \approx \frac{2h}{B(T_0)}.$$

Thus, as regards irreversibility and long-term relaxation, a system of magnetostatically interacting single-domain particles illustrates the properties inherent in spin glasses. Apparently, cluster spin glasses in which the interaction between the magnetic moments of inclusions occurs, say, through indirect RKKY exchange must also possess similar properties [14]. It should be

noted that a wide relaxation time spectrum in this case is ensured by the distribution of particles over critical fields and volumes, and the establishment of equilibrium in all the subsystems with equal relaxation times is assumed to be independent.

However, even for identical particles with critical fields H_c and relaxation times τ_0 such that

$$\frac{1}{\tau_0} = 2f_0 \exp\left(-\frac{mH_c}{2kT}\right),$$

the "inclusion" of interaction leads to the emergence of the relaxation time spectrum in the interval from $\tau_0/\cosh(mB/2kT)$ to τ_0, which readily follows from relation (18).

7. CONCLUSIONS

1. The distribution function for random interaction fields in an amorphous substance in the Ising model has the form of a Gauss function whose mathematical expectation H_0 and dispersion D are determined by the law governing the interaction between particles in accordance with expression (4). The ratio H_0/D determines the type of ordering (paramagnetism, spin glass, ferroglass).

2. The dipole–dipole interaction in an amorphous substance does not lead to a ferroglass- or ferromagnetic-type ordering.

3. Long-term relaxation and irreversibility of magnetization in weak fields in a system of interacting single-domain particles exhibit the spin-glass behavior even under the assumption concerning the independence of the transition of each subsystem with a definite relaxation time to the equilibrium state.

ACKNOWLEDGMENTS

The authors are grateful to L.L. Afremov for useful suggestions and remarks.

This work was supported financially by the Ministry of Education of the Russian Federation (grant no. 97-0-13.1-3).

REFERENCES

1. D. S. Fisher and D. A. Huse, Phys. Rev. B **38**, 373 (1988).
2. S. L. Ginzburg, *Irreversible Processes in Spin Glasses* (Nauka, Moscow, 1989), p. 149.
3. I. Ya. Korenblit and E. F. Shender, Usp. Fiz. Nauk **157**, 267 (1989) [Sov. Phys. Usp. **32**, 139 (1989)].
4. V. S. Dotsenko, Usp. Fiz. Nauk **160** (6), 1 (1993) [Phys. Usp. **36**, 455 (1993)].
5. T. Jonsson, K. Jonason, and P. Norbland, Phys. Rev. B **59**, 9402 (1999).
6. V. I. Belokon' and S. V. Semkin, Zh. Éksp. Teor. Fiz. **102**, 1254 (1992) [Sov. Phys. JETP **75**, 680 (1992)].
7. V. I. Belokon' and S. V. Semkin, Zh. Éksp. Teor. Fiz. **104**, 3784 (1993) [JETP **77**, 815 (1993)].
8. V. P. Shcherbakov and V. V. Shcherbakova, Izv. Akad. Nauk SSSR, Fiz. Zemli **9**, 101 (1975).
9. D. V. Berkov and S. V. Meshkov, Zh. Éksp. Teor. Fiz. **94** (11), 140 (1988) [Sov. Phys. JETP **67**, 2255 (1988)].
10. V. I. Belokon', Izv. Akad. Nauk SSSR, Fiz. Zemli **11**, 106 (1980).
11. G. S. Krinchik, *Physics of Magnetic Phenomena* (Mosk. Gos. Univ., Moscow, 1983), p. 302.
12. E. Z. Meĭlikhov, Zh. Éksp. Teor. Fiz. **117**, 1136 (2000) [JETP **90**, 987 (2000)].
13. V. I. Belokon', Izv. Akad. Nauk SSSR, Fiz. Zemli **21**, 123 (1985).
14. I. V. Zolotukhin and Yu. E. Kalinin, Usp. Fiz. Nauk **160** (9), 75 (1990) [Sov. Phys. Usp. **33**, 720 (1990)].

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 93, No. 1, 2001, pp. 143–152.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 1, 2001, pp. 164–174.

SOLIDS
Structure

Magnetic Resonance of Spin Clusters and Triplet Excitations in a Spin-Peierls Magnet with Impurities

V. N. Glazkov[a, *], R. M. Eremina[b], A. I. Smirnov[a], G. Dhalenne[c], and A. Revcolevschi[c]

[a]*P.L. Kapitza Institute for Physical Problems, Russian Academy of Sciences, ul. Kosygina 2, Moscow, 117334 Russia*
[b]*Zavoiskii Physicotechnical Institute, Kazan, 420029 Russia*
[c]*Laboratoire de Physico-Chimie de l'Etat Solide, Université Paris-Sud, Orsay Cedex, 91405 France*
**e-mail: glazkov@kapitza.ras.ru*

Received March 11, 2001

Abstract—The magnetic resonance spectrum of spin clusters formed in spin-Peierls magnets in the vicinity of impurity ions is investigated. The observed temperature dependences of the effective *g*-factor and the linewidth of the electron paramagnetic resonance (EPR) in crystals of $Cu_{1-x}Ni_xGeO_3$ are described in the model of the exchange narrowing of the two-component spectrum with one component ascribed to spin clusters and exhibiting an anomalous value of the *g*-factor and the other related to triplet excitations. An estimation of the size of the suppressed dimerization region around the impurity ion is obtained (this region includes about 30 copper ions). The dependence of the effective *g*-factor and the EPR linewidth on the impurity concentration at low temperatures indicates the interaction of clusters. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Crystals of quasi-one-dimensional magnet $CuGeO_3$ exhibit magnetic and crystallographic properties characteristic of spin-Peierls magnets [1–3]. The magnetic structure of this compound is based on one-dimensional chains of Cu^{2+} ions ($S = 1/2$) extended along the c axis of the orthorhombic crystal structure [4]. The value of the exchange integral along these chains is 10.4 meV [3].

Below the temperature of the spin-Peierls transition $T_{SP} = 14.5$ K, the dimerization of chains occurs: i.e., magnetic ions approach each other. The dimerization is accompanied by the alternation of the exchange integral, which in turn takes one of the two possible values $J_{1,2} = J(1 \pm \delta)$. An energy gap $E \sim \delta J$ opens between the ground singlet state and triplet excitations. Due to the presence of the gap in the energy spectrum, the magnetic susceptibility decreases and the pure crystal without defects becomes almost nonmagnetic at low temperatures. The lattice transformation due to the dimerization is correlated in space, and the dimers are located on a regular sublattice.

Interchain exchange interaction in $CuGeO_3$ is rather large (the value of the exchange integral along the two directions orthogonal to the chain is only by a factor of 10 or 100 less than the exchange integral along the chains [2, 3]). For this reason, in the absence of the spin-Peierls transition, antiferromagnetic ordering had to be observed. However, the spin-Peierls state is more advantageous and is preserved down to very low temperatures.

Copper germanate $CuGeO_3$ is the only spin-Peierls compound in which a controlled substitution of magnetic ions is possible. The introduction of impurities results in a local suppression of the dimerization in the vicinity of the defect. As a result, the temperature of the spin-Peierls transition decreases and a the long-range antiferromagnetic order is established at a sufficiently low temperatures [5–9].

The occurrence of the antiferromagnetic order and the suppression of the dimerization order is explained in [10, 11]. A cluster of antiferromagnetically correlated spins is formed around the impurity ion. In the chain of spins $S = 1/2$ with alternating exchange interaction, the antiferromagnetic correlations attenuate (see [12]), thus forming the wings of the cluster. As we recede from the defect, the mean value of the spin projection on the *z* axis decreases exponentially. Overlapping of the clusters' wings results in the expansion of the region of antiferromagnetic correlations and the establishment of the long-range ferromagnetic order.

Substitution of the part of the Cu ions by Ni has two significant differences compared to other dopants.

First, in the antiferromagnetically ordered phase, the easy axis of anisotropy is directed along the *a* axis, whereas for other substituting impurities the easy axis of anisotropy is aligned along *c* [5, 8]. Second, an anomalous temperature dependence of the *g*-factor is observed in the dimerized phase. As the temperature drops below the transition temperature T_{SP}, the value of the effective *g*-factor begins to decrease and achieves the value of 1.4 at low temperatures for $H \parallel c$ [8]. The anomalous value of the *g*-factor can be explained by the existence of the antisymmetric Dzyaloshinski–Moriya

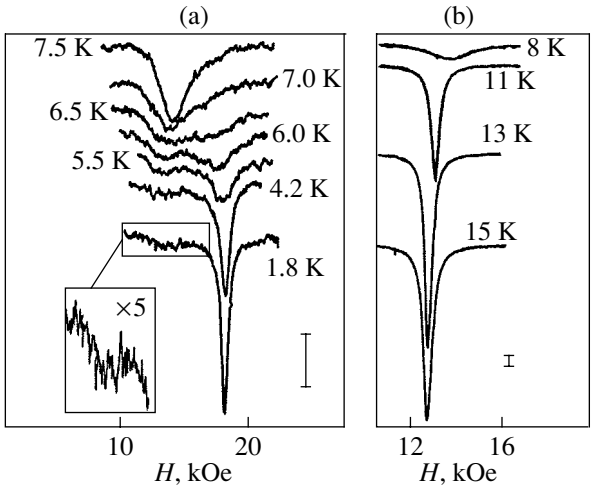

Fig. 1. The temperature evolution of the EPR line at x = 0.2%, **H** ∥ *c*, and f = 36 GHz. The vertical segments in figures *a* and *b* correspond to the same amplitude of the signal.

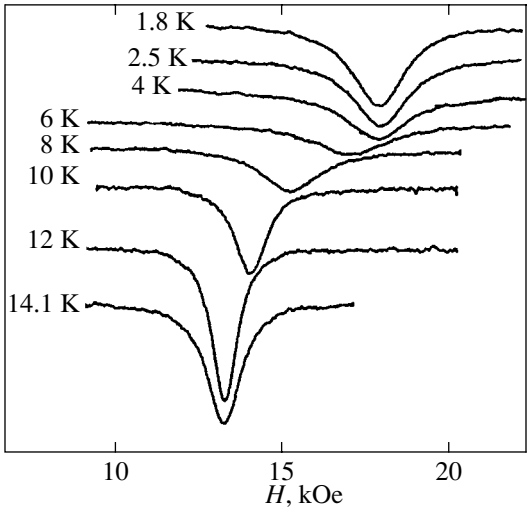

Fig. 2. The temperature evolution of the EPR line at x = 0.8%, **H** ∥ *c*, and f = 36 GHz.

exchange interaction in the vicinity of the defect. In a multispin system consisting of more than two spins, the existence of the Dzyaloshinski–Moriya interaction along with the symmetric exchange interaction results in a strong anisotropy of the effective *g*-factor and in the decrease of its value [13]. Calculations based on the six-spin model show that the existence of the antisymmetric exchange interaction with the value of the exchange integral of about 30% of the value of the intrachain exchange interaction is sufficient for the description of the deviation observed [8].

The present paper continues the study started in [8]. Its purpose is to investigate high-quality samples of $CuGeO_3$ doped with nickel including those with a low content of the impurity (x < 1%). The examination of samples with a small concentration of the impurity (when the average distance between the impurity ions exceeds the characteristic cluster size) makes it possible to observe the magnetic resonance of isolated clusters. A noticeable difference of the *g*-factor of clusters from the *g*-factor of excitations of the spin-Peierls matrix makes it possible to differentiate between their EPR signals. In turn, this fact opens the possibility to investigate the interaction of clusters with the environment and between themselves. The analysis of the experimental data allowed us to determine the characteristic size of the cluster that is formed around the impurity ion, namely, the size of the region where the dimerization is destroyed and that of the region in which the antisymmetric exchange interaction exists.

2. EXPERIMENTAL TECHNIQUE AND SAMPLES

For the experiment, high-quality samples of $Cu_{1-x}Ni_xGeO_3$ with an impurity concentration x = 0.2% and x = 0.8% were grown. In order to analyze the dependence of the *g*-factor on concentration, samples with higher concentrations of the impurity (x = 1.9% and 3.0%) were also used.

To control the quality of the samples, a monocrystal of the pure compound grown following the same technology was used. At the temperature of 4 K, the magnetic susceptibility of this sample determined by the integral intensity of the EPR signal was about 4% of the susceptibility at the transition temperature. This corresponds to the residual concentration of the magnetic defects per a copper ion equal to $x_0 \sim 0.05\%$.

The investigations were performed at a frequency of 36 GHz and the temperatures in the range 1.8–20 K with the help of an EPR spectrometer with a transmission type cavity. The magnetic resonance line was registered as the dependence of the intensity of the microwave power transmitted through the resonator on the magnetic field applied. In this case, the variation of the signal is proportional to the imaginary part of the magnetic susceptibility.

3. EXPERIMENTAL RESULTS

As the temperature drops below the spin-Peierls transition temperature (which is equal to 13.5 K for x = 0.2% and 12.0 K for x = 0.8%), the field of the resonance absorption starts to increase. The temperature of the spin-Peierls transition was determined by the beginning of the decrease of the integral intensity of the EPR signal. The increase of the resonance absorption field corresponds to the decrease of the *g*-factor. The variation of the EPR line with temperature is shown in Figs. 1 and 2.

The temperature dependences of the *g*-factor are presented in Figs. 3 and 4. At low temperatures (T < 4 K), the values of the *g*-factor remain constant and are equal to g_a = 1.75, g_b = 1.87, and g_c = 1.43 (for x = 0.2%).

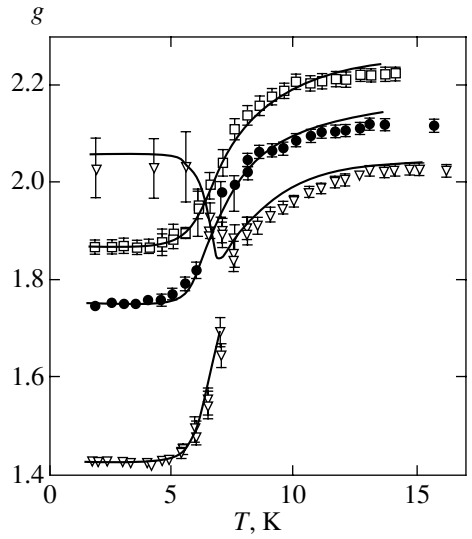

Fig. 3. Dependence of the effective g-factor on temperature for the sample with the impurity content $x = 0.2\%$: ●—**H** ∥ a, □—**H** ∥ b, and ▽—**H** ∥ c. Solid curves correspond to the theoretical calculation.

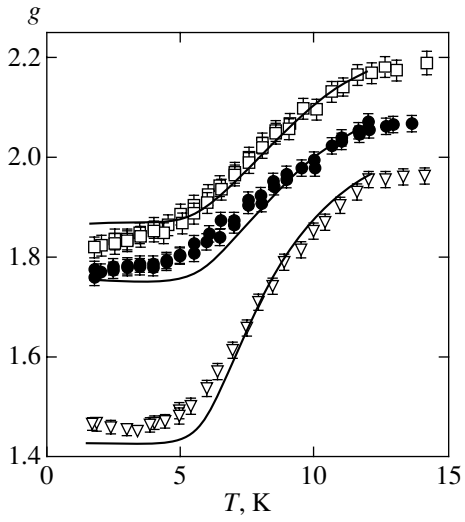

Fig. 4. Dependence of the effective g-factor on temperature for the sample with the impurity content $x = 0.8\%$: ●—**H** ∥ a, □—**H** ∥ b, and ▽—**H** ∥ c. Solid curves correspond to the theoretical calculation.

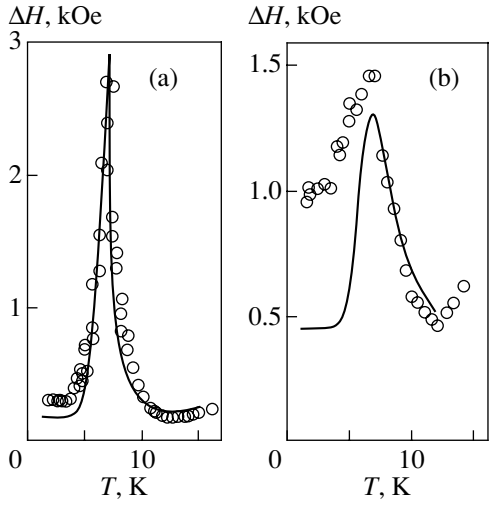

Fig. 5. Dependence of the half-width of the EPR line on temperature for $x = 0.2\%$ (a) and $x = 0.8\%$ (b) at **H** ∥ c and $f = 36$ GHz. Solid curves correspond to the theoretical calculation.

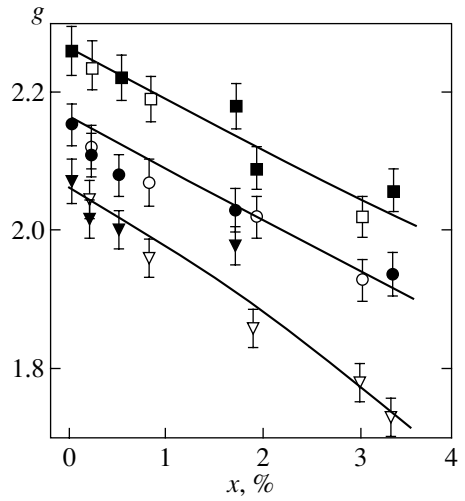

Fig. 6. Dependence of the effective g-factor on impurity concentration at $T = 15$ K: ○—**H** ∥ a, □—**H** ∥ b, and ▽—**H** ∥ c. Solid curves correspond to the theoretical calculation, and black symbols correspond to the data of the study [8].

For the sample with the impurity concentration $x = 0.2\%$, the magnetic resonance line splits into two components at the temperature $T' \approx 7$ K (Fig. 1). As the temperature decreases, one of those components continues to move to higher fields and its intensity increases. The second component remains in the field close to the EPR field above T_{SP}, but its intensity decreases and it almost vanishes as the temperature decreases further. The width of the magnetic resonance line has its maximum at the temperature close to the splitting temperature T' (Fig. 5). A similar splitting was observed at other orientations of the sample with respect to the field for $x = 0.2\%$; however, we were able to follow it down to very low temperatures only for **H** ∥ c (this is due to the fact that for this orientation there is the maximal difference of the g-factors of two spectral components, which makes it possible to distinguish the faint absorption line on the wing of the strong one).

For the sample with the impurity concentration $x = 0.8\%$, the magnetic resonance line consists of the single component at all temperatures; the maximum of the linewidth is observed in the vicinity of T' (Fig. 5).

In the paramagnetic phase, the value of the g-factor is also different from the value characteristic for the pure compound. The dependence of the g-factor value on the impurity concentration at $T > T_{SP}$ is shown in Fig. 6. As the impurity concentration increases, the

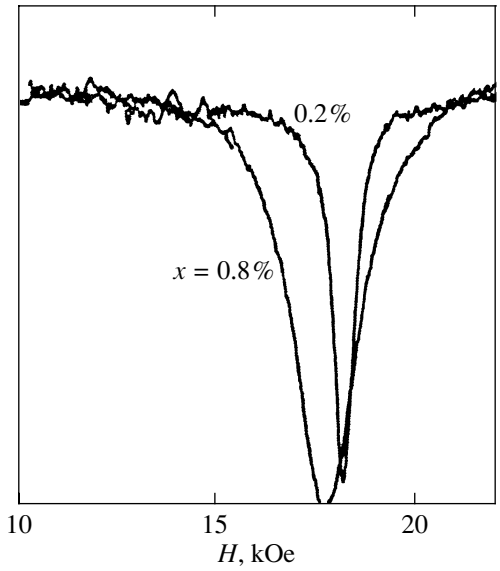

Fig. 7. Comparison of the EPR lines for $x = 0.2\%$ and $x = 0.8\%$ at $T = 1.8$ K, $\mathbf{H} \parallel c$, and $f = 36$ GHz.

value of the g-factor decreases for all orientations of the magnetic field.

For samples with the impurity concentration $x = 1.9\%$ and 3.0%, the long-range antiferromagnetic order is established, which manifests itself in the transition from the linear gapless EPR spectrum to a spectrum that is typical to antiferromagnets with orthorhombic symmetry. The Néel's temperatures are $T_N \approx 2.5$ K for $x = 1.9\%$ and $T_N \approx 3.5$ K for $x = 3.0\%$.

Comparison of the EPR lines at the minimal temperature (Fig. 7) shows that the field of resonance absorption and the width of the line of magnetic resonance are different for samples with different concentrations of impurity. Dependences of the linewidth and the g-factor value on impurity concentration are presented in Fig. 8. (For samples that exhibit the antiferromagnetic ordering, the data were taken at $T = T_N$.) For small x, the width of the EPR line linearly depends on the concentration. Dependences of the EPR linewidth and of the g-factor value on the orientation of the magnetic field at $T = 1.8$ K for the sample with $x = 0.8\%$ are shown in Fig. 9.

4. DISCUSSION

Before getting down to the quantitative analysis of the experimental data, we will present a qualitative description.

According to the concept developed in [10, 11], a cluster of exchange-correlated spins is formed around the impurity ion in the spin-Peierls matrix. Due to the existence of the antisymmetric exchange interaction in this cluster, the EPR of clusters is characterized by an unusually small value of the g-factor g_{cl} [8]. Clusters are surrounded by a dimerized spin-Peierls matrix. Triplet excitations of the dimerized matrix are characterized by the value of the g-factor of copper ions g_{Cu} close to 2. Due to the exchange interaction of clusters with excitations, an EPR line with an intermediate value of the g-factor is observed (the so-called exchange narrowing). At temperatures close to the spin-Peierls transition temperature, when the concentration of spin-Peierls excitations is large, an EPR line with the

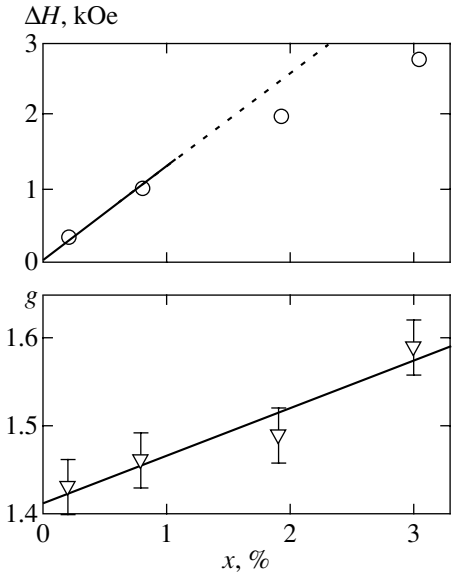

Fig. 8. Dependence of the half-width of the EPR line on impurity concentration for $\mathbf{H} \parallel c$: $T = T_N = 2.5$ K for $x = 1.9\%$, $T = T_N = 3.5$ K for $x = 3.0\%$, and $T = 1.8$ K for $x = 0.2\%$ and $x = 0.9\%$.

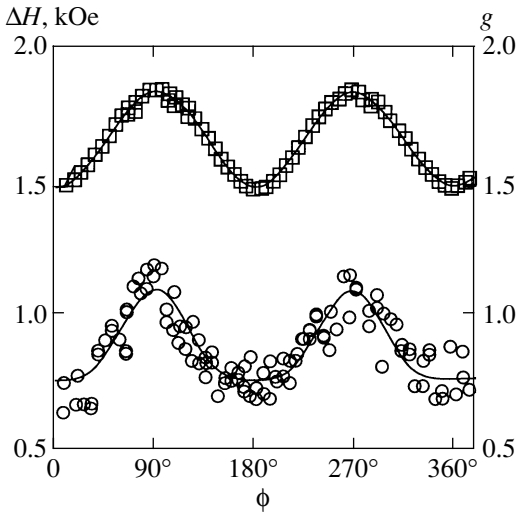

Fig. 9. Angular dependence of the width of the magnetic resonance line (○) and the effective g-factor (□) for the field applied in the plane bc. Solid curves correspond to formula (25), $x = 0.8\%$, $T = 1.8$ K, and $f = 36$ GHz; $\phi = 0$ corresponds to $\mathbf{H} \parallel c$.

g-factor close to the values characteristic to copper ions is observed. As the temperature decreases, the concentration of triplet excitations decreases due to the existence of an energy gap and the EPR lime shifts to the value characteristic of isolated clusters. As the temperature decreases further, the effectiveness of the interaction of clusters with excitations decreases and the EPR line splits into two components. A similar phenomenon was observed for the magnetic resonance of temperature-activated spins in radicals [14]. At last, at low temperatures, when triplet excitations are practically frozen out, the EPR line consists of two components—a strong one, characterized by the g-factor of clusters, and a faint one, which represents the residual triplet excitations and magnetic defects. This description corresponds to the observed evolution of the magnetic resonance line for the samples with the impurity concentration of 0.2%.

Similarly, one can explain the dependence of the g-factor on impurity concentration at temperatures greater than the transition temperature. In this case, one should consider the closest neighborhood of the impurity ion in which antisymmetric interaction exists as a cluster characterized by the anomalous value of the g-factor g_{cl}. The EPR of copper ion chains is characterized by the g-factor g_{Cu}. Due to the exchange interaction of the cluster with the surrounding copper matrix, the EPR line with an intermediate value of the g-factor is observed. The greater the number of clusters, the more the magnetic resonance line is shifted from g_{Cu} to g_{cl}. Thus, the value of the effective g-factor must decrease as the impurity concentration increases.

In this reasoning, we assumed that the g-factor is the same for all clusters. This is actually true if the interaction between clusters is negligible. In this case, the parameters of the EPR line would depend only on the interaction of clusters with triplet excitations. However, the fact that the resonance absorption fields for the samples with $x = 0.2\%$ and $x = 0.8\%$ are different, shows that even for these impurity concentrations the interaction between clusters must be taken into account. Clusters interact due to the fact that their wings overlap [11]; this makes it possible to obtain a coarse evaluation of the cluster size (assuming that clusters do not interact at $x = 0.2\%$ and that the interaction leads to the widening and shift of the line at $x = 0.8\%$):

$$L \sim 1/0.008 \sim 100. \tag{1}$$

This result is overestimated since the distance between the majority of clusters is less than the average one. Since antiferromagnetic correlations at cluster wings are destroyed by thermal fluctuations, the influence of the cluster interaction on the EPR line should decrease with the increase of temperature.

We will assume that an isolated cluster is characterized by the values of the g-factor observed for the sample with $x = 0.2\%$ at the minimal temperature ($g_{cl}^{(a)} = 1.75$, $g_{cl}^{(b)} = 1.87$, and $g_{cl}^{(c)} = 1.43$). The values of the g-factor for excitations correspond to the g-factor of copper ions in undistorted crystal environment, i.e., in pure $CuGeO_3$ ($g_{\mathrm{Cu}}^{(a)} = 2.15$, $g_{\mathrm{Cu}}^{(b)} = 2.26$, and $g_{\mathrm{Cu}}^{(c)} = 2.06$, and they are practically independent of temperature [15]).

In the subsequent analysis we use the following simplified model. We assume that in the close neighborhood of the impurity ion of size L_{dim} the dimerization is suppressed and triplet excitations of the spin-Peierls matrix do not reach this region. Antiferromagnetic correlations decay exponentially with distance from the defect. This attenuation is characterized by the magnetic correlation length of dimerized chains $\xi \sim v/\Delta$, where v is the speed of spin excitations and Δ is the energy gap, (see [12]). In addition, there exists the antisymmetric Dzyaloshinski–Moriya exchange interaction in a certain neighborhood of the impurity ion due to a local reduction of symmetry. The size of this region is $L_{DM} < L_{\mathrm{dim}}$. The values of L_{DM} and L_{dim} are measured in interatomic distances along the chains.

The analysis of the dependence of magnetic susceptibility on temperature for a similar model was conducted in [16, 17]. The advantage of the EPR method is in the fact that a noticeable difference in g-factors of clusters and excitations makes it possible to separate their contributions.

4.1. Interaction of Clusters with Excitations in the Molecular Field Approximation

First, we consider the case $T < T_{SP}$ when spin chains are dimerized. In the vicinity of the impurity ion, a cluster of exchange-coupled spins with the total spin $S = 1/2$ is formed. At a large distance from the defect, the spin-Peierls matrix remains unperturbed, and its magnetic properties are described by triplet excitations, which are separated by a gap from the ground state.

Propagation of antiferromagnetic correlations from the cluster into the dimerized matrix results in the appearance of an interaction between the cluster and excitations. Since this interaction appears due to the exchange interaction between spins, the average energy of the interaction can be written in the form

$$E_{\mathrm{int}} = \sum_{i = 1, \ldots, n} J_{\mathrm{eff}}(\langle \mathbf{S}_{cl} \rangle \cdot \langle \mathbf{S}_{\mathrm{Cu}}(i) \rangle). \tag{2}$$

Here J_{eff} is the effective exchange integral, $\langle \mathbf{S}_{cl} \rangle$ is the average total spin value of the cluster, and $\langle \mathbf{S}_{\mathrm{Cu}} \rangle$ is the average spin value on the copper ion located outside the cluster (this value is related to triplet excitations). The summation is performed over n effective neighbors of the cluster (since the major role is played by the interaction along spin chains, we assume that $n = 2$).

Following the molecular field theory, we obtain the following system of self-consistent equations for aver-

age magnetization of a cluster and a copper ion in the dimerized matrix:

$$\langle \mu_{cl} \rangle = \chi_{cl}^{(0)}\left(H + n\frac{J_{\text{eff}}}{g_{cl}g_{\text{Cu}}\mu_B^2}\langle \mu_{\text{Cu}} \rangle\right),$$
$$\langle \mu_{\text{Cu}} \rangle = \chi_{\text{Cu}}^{(0)}\left(H + n\frac{J_{\text{eff}}}{g_{cl}g_{\text{Cu}}\mu_B^2}\langle \mu_{\text{Cl}} \rangle\right). \quad (3)$$

Here $\chi_{cl,\,\text{Cu}}^{(0)}$ are susceptibilities per one cluster and per one copper ion in the absence of the interaction.

From Eqs. (3) one can derive the following equations for the magnetizations with regard for the interaction:

$$\chi_{cl} = \chi_{cl}^{(0)}\frac{1 + n\eta\chi_{\text{Cu}}^{(0)}}{1 - n\eta^2\chi_{cl}^{(0)}\chi_{\text{Cu}}^{(0)}},$$
$$\chi_{\text{Cu}} = \chi_{\text{Cu}}^{(0)}\frac{1 + \eta\chi_{cl}^{(0)}}{1 - n\eta^2\chi_{cl}^{(0)}\chi_{\text{Cu}}^{(0)}}, \quad (4)$$

where $\eta = J_{\text{eff}}/(g_{\text{Cu}}g_{cl}\mu_B^2)$.

The magnetic susceptibility of a single isolated cluster obeys the Curie law

$$\chi_{cl}^{(0)} = \frac{g_{cl}^2\mu_B^2 S(S+1)}{3kT}. \quad (5)$$

For the susceptibility due to triplet excitations, we will use the results obtained in [16, 17]. In those studies, an approximation of the magnetic susceptibility of pure $CuGeO_3$ crystals at temperatures below T_{SP} was obtained experimentally. This approximation of the molar susceptibility at $\mathbf{H} \parallel c$ has the form

$$F(t) = (a_0 + a_1 t + a_2 t^2)\exp\left(-\frac{A}{t}\right), \quad t = \frac{T}{T_{SP}}, \quad (6)$$

where $a_0 = 26.0 \times 10^{-3}$ cgs units/mol, $a_1 = -41.6 \times 10^{-3}$ emu/mol, $a_2 = 28.2 \times 10^{-3}$ emu/mol, and $A = 2.39$.

Then, we have for the magnetic susceptibility per copper ion in the dimerized matrix:

$$\chi_{\text{Cu}}^{(0)} = \left(\frac{g_{\text{Cu}}^{(i)}}{g_{\text{Cu}}^{(c)}}\right)^2\frac{F(T/T_{SP})}{N_A}. \quad (7)$$

Here $g_{\text{Cu}}^{(i)}$ is the g-factor of the copper ion in the corresponding direction.

If the impurity concentration is x, then the number of clusters is xN_A and the number of copper ions in the dimerized matrix is $(1 - xL_{\text{dim}})N_A$. Assuming that clusters do not interact, we obtain the following formulas for the total susceptibility of clusters and excitations:

$$\tilde{\chi}_{cl} = xN_A\chi_{cl},$$
$$\tilde{\chi}_{\text{Cu}} = (1 - xL_{\text{dim}})N_A\chi_{\text{Cu}}. \quad (8)$$

Equations (4)–(8) allow one to determine the contribution of clusters and triplet excitations to the susceptibility for all temperatures below the spin–Peierls transition temperature. We will use this result later.

The case $T > T_{SP}$ can be treated in a similar way. As it has already been mentioned above, in this case the neighborhood of an impurity ion in which the Dzyaloshinski–Moriya exchange interaction exists should be considered as a cluster; hence, the characteristic size in Eq. (8) is L_{DM}. Since the susceptibility of spin chains weakly depends on temperature above the transition temperature, we must set $T = T_{SP}$ in Eq. (7).

4.2. Dependence of the g-Factor on Temperature

As it has already been mentioned above, the evolution of the EPR line in the spectrum of a sample with a nickel concentration equal to 0.2% (Fig. 1) has the form typical of the exchange-narrowed two-component spectrum of the magnetic resonance with the frequency of exchange jumps dependent on temperature.

Following [14, 18], we assume that the influence of the exchange interaction on the magnetic resonance spectrum of the system can be considered as random transitions with the characteristic frequency ω_e between the states with different Zeeman's frequencies $\omega_2^{(0)} > \omega_1^{(0)}$.

The location of the center of gravity of the magnetic resonance spectrum is independent of ω_e and is determined by the formula

$$\bar{\omega} = \frac{\omega_1^{(0)}\tilde{\chi}_1 + \omega_2^{(0)}\tilde{\chi}_2}{\tilde{\chi}_1 + \tilde{\chi}_2}, \quad (9)$$

where $\tilde{\chi}_{1,2}$ are the susceptibilities of the corresponding states with regard for the interaction between them.

Analysis of these random transitions by statistical methods (see [18]) show that the frequencies of the spectral components and their widths are determined by the formulas

$$\omega_{1,2} = \bar{\omega} + \text{Im}(\lambda_{1,2}),$$
$$\Delta\omega_{1,2} = \text{Re}(\lambda_{1,2}), \quad (10)$$

where

$$\lambda_{1,2} = \frac{1}{2}\{-[\omega_e - i\delta] \pm \sqrt{\omega_e^2 - \Delta^2 - 2i\omega_e\delta}\}, \quad (11)$$

$$\Delta = \omega_2^{(0)} - \omega_1^{(0)}, \quad \delta = \omega_1^{(0)} + \omega_2^{(0)} - 2\bar{\omega}. \quad (12)$$

In the limit of $\omega_e \gg \Delta$, we have

$$\omega_1 = \bar{\omega} - \delta\frac{\delta^2 + \Delta^2}{4\omega_e^2}, \quad \Delta\omega_1 = \frac{\delta^2 - \Delta^2}{4\omega_e},$$
$$\omega_2 = \bar{\omega} + \delta, \quad \Delta\omega_2 = -\omega_e. \quad (13)$$

Thus, the EPR spectrum consists of a narrow line close to $\bar{\omega}$ and a wide background line.

In the absence of the interaction ($\omega_e = 0$), we have

$$\omega_{1,2} = \omega_{1,2}^{(0)}, \quad \Delta\omega_{1,2} = 0, \tag{14}$$

which corresponds to two narrow spectral components at the frequencies $\omega_1^{(0)}$ and $\omega_2^{(0)}$.

Qualitatively, this corresponds to the observed transition from the EPR line consisting of a single component to the two-component line. In this model, we neglect the intrinsic widths of lines in both states of the system.

Magnetic properties of the doped spin-Peierls system at $T > T_N$ correspond to free spins of clusters and triplet excitations of the dimerized matrix. The difference in g-factors of clusters and excitations lead to differences in Zeeman's frequencies.

The presence of an energy gap leads to a dependence of the concentration of triplet excitations on temperature. In this case, the frequency of exchange jumps ω_e also depends on temperature as

$$\omega_e(t) = \Omega_e \exp\left\{-\frac{E(t)/T_{SP}}{t}\right\}, \quad t = \frac{T}{T_{SP}}. \tag{15}$$

The dependence of the energy gap on temperature can be approximated as follows (see [7, 19]):

$$E(t) = E(0)(1-t)^a, \quad a \approx 0.1, \quad t = T/T_{SP}. \tag{16}$$

The magnitude of the energy gap at $T = 0$ K is related to the transition temperature by the equation (see [20])

$$E(0) = 1.76kT_{SP}. \tag{17}$$

Equations (9)–(12) and (15)–(17) make it possible to obtain temperature dependences of the resonance absorption frequencies (g-factors) and widths of spectrum components. To take into account the interaction of clusters with triplet excitations, we use the molecular field approximation (4)–(8).

The temperature dependences of the g-factor and the width of the magnetic resonance line are described with the help of three adjustable parameters—the size of the region of suppressed dimerization L_{dim}, the effective exchange integral value J_{eff}, and the preexponential coefficient of the exchange frequency Ω_e.

This model assumes that clusters do not directly interact. As has been mentioned above, the influence of the interaction of clusters decreases with increasing temperature. For this reason, when adjusting the parameters, we used the temperature dependence of the g-factor at $T > T' = 7$ K for all basic orientations of both samples and the temperature dependence of the g-factor below T' for the sample with the impurity content 0.2% for $\mathbf{H} \parallel c$.

Thus, the following values of the adjustable parameters were obtained:

$$L_{\text{dim}} = 32 \pm 2, \quad J_{\text{eff}} = -(13 \pm 1)\ \text{K}, \quad \Omega_e = (2.2 \pm 0.3) \times 10^{12}\ \text{s}^{-1}. \tag{18}$$

Note that $\hbar\Omega_e/k \sim 16$K, which is close to J_{eff}. This result could be expected since Ω_e and J_{eff} must be determined by the magnitude of the intrachain exchange integral.

The comparison of the theoretical and experimental results is illustrated in Figs. 3, 4. The theoretical dependences provide an accurate description of the experimental data for the sample with the impurity concentration 0.2%; however, for the sample with $x = 0.8\%$, there is a disagreement at low temperatures, which we attribute to the interaction of clusters.

The value of the suppressed dimerization region obtained here coincides with the result of the paper [17], which was obtained by the analysis of static susceptibilities.

We also can make a coarse evaluation of the impurity concentration at which the long-range spin-Peierls order must be completely destroyed: $x_c = 1/L_{\text{dim}} \sim 0.03$, which is in good agreement with the result obtained in [21].

4.3. Dependence of the Width of the Magnetic Resonance Line on Temperature

On the basis of the model described above, we can derive the dependence of the width of the EPR line on temperature. The comparison with experimental data is presented in Fig. 5. For convenience, the width of the line at the spin-Peierls transition point is added to the theoretical dependences. No additional adjustable parameters were used.

For the sample with the impurity concentration 0.2%, the agreement of the theory with the experiment is very good. The theory provides the correct location of the maximum of the linewidth and the correct value of at this point. The best agreement between the theory and the experiment is achieved for $\mathbf{H} \parallel c$. This could be expected, since in this case one of the basic assumption of our model is best satisfied, namely, that the intrinsic linewidth of the spectral components can be neglected as compared with the splitting between them.

For the sample with $x = 0.8\%$, a disagreement of the theory and the experiment is observed. The location of the maximum is determined rather well; however, the behavior of the linewidth at low temperatures is different from that predicted by the model. We attribute this fact to interaction between clusters.

4.4. Dependence of the g-Factor on Concentration above the Temperature of the Spin-Peierls Transition

The approach developed above can be also applied to the description of the dependence of the g-factor value on impurity concentration above T_{SP}. In this case, we consider as a cluster the neighborhood of the impurity ion of size L_{DM} in which the Dzyaloshinski–Moriya exchange interaction exists.

At temperatures close to T_{SP}, the condition $\omega_e \gg \Delta$ ($\omega_e \sim \Omega_e \sim 10^{12}$ s^{-1}, $\Delta \sim 10^{10}$ s^{-1}) holds. Hence, simplified Eqs. (13) can be used. Neglecting the terms of order Δ^2/ω_e^2, we obtain the following equation for the mean value of the g-factor (this equation is similar to (9)):

$$\bar{g} = \frac{g_{cl}\tilde{\chi}_{cl} + g_{\mathrm{Cu}}\tilde{\chi}_{\mathrm{Cu}}}{\tilde{\chi}_{cl} + \tilde{\chi}_{\mathrm{Cu}}}. \qquad (19)$$

As before (see Eqs. (4)–(8)), susceptibilities $\tilde{\chi}$ are determined in the molecular field approximation. We assume that in the absence of interaction, the cluster susceptibility is described by the Curie law (5), and the susceptibility of the copper ions surrounding the cluster is independent of temperature and equals the susceptibility at the point of the spin-Peierls transition (7).

We do not present the expression for the dependence of the g-factor value on concentration because it is too cumbersome. This expression includes two parameters: the effective exchange integral J_{eff} and L_{DM}. The parameter J_{eff} has already been determined earlier. This leaves us a single adjustable parameter to describe the dependence of the g-factor on impurity concentration for all orientations of the sample with respect to the magnetic field. As it has already been mentioned, the interval of the impurity concentration at which clusters can be considered uninteracting increases with temperature. Thus, at high temperatures our approach can be applied even in the case of large concentrations. Figure 6 presents data for samples with nickel concentrations up to 3.3% at the temperature of 15 K along with theoretical curves. The experimental dependences correspond to $L_{DM} = 18 \pm 2$.

4.5. Dependence of the EPR Linewidth on Concentration at Low Temperatures. Interaction of Clusters

The difference of the EPR lines of samples with the impurity concentration 0.2% and 0.8% (Fig. 7) indicates that clusters interact. The dependence of the linewidth on impurity concentration for small x is linear (Fig. 8). A linear dependence of the EPR linewidth on the concentration of magnetic centers was observed in experiments with diluted paramagnets (paramagnetic centers in a diamagnetic crystal) (see, e.g., [22]).

As a possible cause of the observed linewidth, one can suggest long-range dipole–dipole interactions or exchange interactions occurring due to overlapping of wings of nearby clusters.

In order to estimate the contribution of the dipole–dipole interaction to the linewidth, we notice that the dipole field magnitude is about 10 Oe at the distance of 10 Å from the magnetic moment μ_B. Thus, the observed width of the line (~1 kOe) cannot be explained by the existence of the dipole–dipole interaction between impurity ions.

Therefore, the linewidth must be determined by the antisymmetric or anisotropic exchange interaction of clusters. Reorientation of clusters due to thermal fluctuations leads to the appearance of a random effective field H_{eff}, which determines the linewidth.

Due the random distribution of the impurities, the number of closely spaced clusters (i.e., those spaced by a distance below average) constitutes a noticeable part of the total number of clusters. In the one-dimensional case, the probability of detecting an impurity ion at the distance of n interatomic distances from the given ion at the impurity concentration x is

$$p(n) = x(1-x)^n. \qquad (20)$$

Then, the probability that the distance between impurity ions is less than N is

$$P(n<N) = \sum_{n=0}^{N-1} p(n) = 1-(1-x)^N. \qquad (21)$$

In the limiting case $x \ll 1$, we obtain $P(n<N) \approx Nx$. Thus, if the impurity concentration is 1% (and the average distance between impurity ions in a chain is 100 interatomic distances), the part of the clusters that are closer than $L_{\mathrm{dim}} = 32$ to each other is about 30%.

On the basis of the observed values of the width of the magnetic resonance line, we can give a coarse evaluation of the magnitude of the random effective magnetic field generated by clusters.

Since antiferromagnetic correlations decrease exponentially when moving away from the defect into the dimerized matrix, we assume that the average value of the effective field depends on the distance L from the region of the destroyed dimerization according to the law

$$H_{\mathrm{eff}} = H_0 \exp(-L/\xi), \qquad (22)$$

where ξ is the magnetic correlation length and H_0 is the effective field on the boundary of the suppressed dimerization region.

Averaging over L with the help of distribution (20) and taking into account that x is small, we obtain the

following estimate for the width of the EPR line:

$$\Delta H \sim \sum_{L=0}^{\infty} H_{\text{eff}}(L)p(L) = \frac{xH_0}{1-(1-x)\exp(-1/\xi)} \approx \xi x H_0. \quad (23)$$

Hence, setting $\xi \approx 10$ (see [23]) and taking into account that at $x \sim 1\%$ the linewidth $\Delta H \sim 1$ kOe (see Fig. 8), we obtain for H_0 the estimate $H_0 \sim 10$ kOe. Such a magnitude of the effective field corresponds to energy of order of 1 K, which is about 1% of the intrachain exchange integral.

Additional information on the nature of the interaction that determines the width of the magnetic resonance line can be obtained with the help of the angular dependence of the linewidth. The dependences of the effective *g*-factor and the EPR linewidth on the orientation of the magnetic field in the plane *bc* of the crystal with $x = 0.8\%$ are presented in Fig. 9. The angular dependence of the *g*-factor is accurately approximated by the function

$$g_{\text{eff}}^2 = g_c^2 \cos^2\phi + g_b^2 \sin^2\phi, \quad (24)$$

where ϕ is the angle in the plane *bc* measured from the axis *c*. Thus, the anisotropy of the *g*-factor can be described in terms of the principal values of the *g*-tensor.

The contribution of the antisymmetric exchange interaction to the angular dependence of the linewidth is π-periodic, and the contribution of the anisotropic symmetric interaction is $\pi/2$-periodic [24]. In the case under consideration, both contributions are present. Figure 9 illustrates fitting of experimental data for the EPR linewidth by the function

$$A + B\cos(2\phi) + C\cos(4\phi). \quad (25)$$

However, one must take into account the fact that the anisotropy of the *g*-factor also affects the angular dependence of the linewidth, and this influence is periodic with the period equal to that of the angular dependence of the *g*-factor (24), i.e., π.

The magnitude of the parameter *D* of the anisotropic symmetric exchange is related to the isotropic exchange integral *J* by the equation (see [25])

$$D \sim \left(\frac{\Delta g}{g}\right)^2 J, \quad (26)$$

where $\Delta g = g - 2$. In $CuGeO_3$, $(\Delta g/g) \sim 0.1$, which yields an estimate of 1 K for *D*. Thus, it is possible that the observed magnitude of the EPR line is explained by the existence of the symmetric anisotropic exchange interaction.

5. CONCLUSIONS

When studying high-quality samples of the spin-Peierls magnet $CuGeO_3$ doped with nickel with a small impurity concentration $x < 1\%$, it was discovered that the *g*-factor decreases with temperature to unusually small values (down to 1.4). This fact is due to formation of the clusters of antiferromagnetically correlated spins with antisymmetric exchange interaction around impurity ions. Above the transition temperature, the value of the *g*-factor decreases as the impurity concentration increases.

The dependence of the *g*-factor on temperature and concentration can be explained in the framework of the model of exchange narrowing. An analysis of data allows one to evaluate the size of the region around an impurity in which the dimerization is suppressed ($L_{\text{dim}} \approx 30$ interatomic distances) and the size of the region in which the antisymmetric exchange interaction exists ($L_{DM} \approx 20$ interatomic distances).

Experimental data show that even at small impurity concentrations, the interaction between clusters plays an important role at low temperatures. The magnitude and the angular dependence of the width of the magnetic resonance line suggest the existence of an anisotropic exchange interaction in $CuGeO_3$.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research and the *Deutsche Forschungsgemeinschaft* (DFG), joint project no. 01-02-04007; INTAS, project no. 99-0155; the U.S. Civilian Research & Development Foundation (CRDF), project no. RP1-2097; and the BRHE Foundation, project no. REC007.

The authors are grateful to M.V. Eremin for his interest to the work and useful discussions. One of the authors (V. Glazkov) is also grateful to the Forshungcentrum Julih Gmbh for the continuous support of his research work.

REFERENCES

1. M. Hase, I. Terasaki, and K. Uchinokura, Phys. Rev. Lett. **70**, 3651 (1993).
2. L. P. Regnault, M. Ain, B. Hennion, *et al.*, Phys. Rev. B **53**, 5579 (1996).
3. M. Nishi, O. Fujita, and J. Akimitsu, Phys. Rev. B **50**, 6508 (1994).
4. M. Braden, G. Wilkendorf, J. Lorenzana, *et al.*, Phys. Rev. B **54**, 1105 (1996).
5. S. Coad, J.-G. Lussier, D. F. McMorrow, and D. McK. Paul, J. Phys.: Condens. Matter **8**, 6251 (1996).
6. T. Masuda, I. Tsukada, K. Uchinokura, *et al.*, Phys. Rev. B **61**, 4103 (2000).
7. S. Coad, O. Petrenko, D. McK. Paul, *et al.*, Physica B (Amsterdam) **239**, 350 (1997).
8. V. N. Glazkov, A. I. Smirnov, O. A. Petrenko, *et al.*, J. Phys.: Condens. Matter **10**, 7879 (1998).

9. S. Katano, O. Fujita, J. Akimitsu, *et al.*, Phys. Rev. B **57**, 10280 (1998).
10. H. Fukuyama, T. Tanimoto, and M. Saito, J. Phys. Soc. Jpn. **65**, 1182 (1996).
11. D. Khomskii, W. Geertsmaa, and M. Mostovoy, Czech. J. Phys. **46** (S6), 3229 (1996).
12. K. Okamoto, J. Phys. Soc. Jpn. **56**, 1627 (1987).
13. M. I. Belinskiĭ, B. S. Tsukerblat, and A. V. Ablov, Fiz. Tverd. Tela (Leningrad) **16**, 989 (1974) [Sov. Phys. Solid State **16**, 639 (1974)].
14. M. T. Jones and D. B. Chestnut, J. Chem. Phys. **38**, 1311 (1963).
15. M. Honda, T. Shibata, K. Kindo, *et al.*, J. Phys. Soc. Jpn. **65**, 691 (1996).
16. B. Grenier, J.-P. Renard, P. Veillet, *et al.*, Phys. Rev. B **57**, 3444 (1998).
17. B. Grenier, J.-P. Renard, P. Veillet, *et al.*, Phys. Rev. B **58**, 8202 (1998).
18. P. W. Anderson, J. Phys. Soc. Jpn. **9**, 316 (1954).
19. M. C. Martin, G. Shirane, Y. Fujii, *et al.*, Phys. Rev. B **53**, R14713 (1996).
20. E. Pytte, Phys. Rev. B **10**, 4637 (1974).
21. N. Koide, Y. Uchiyama, T. Hayashi, *et al.*, cond-mat/9805095 (1998).
22. S. A. Altshuler and B. M. Kozyrev, *Electron Paramagnetic Resonance in Compounds of Transition Elements* (Nauka, Moscow, 1972; Halsted, New York, 1975), Parag. 49.
23. K. M. Kojima, Y. Fudamoto, M. Larkin, *et al.*, Phys. Rev. Lett. **79**, 503 (1997).
24. M. Lohmann, H. A. Krug von Nidda, M. V. Eremin, *et al.*, Phys. Rev. Lett. **85**, 1742 (2000).
25. Yu. V. Yablokov, V. K. Voronkova, and L. V. Mosina, *Paramagnetic Resonance of Exchanges Clusters* (Nauka, Moscow, 1988), Chap. 1.

Translated by A. Klimontovich

Journal of Experimental and Theoretical Physics, Vol. 93, No. 1, 2001, pp. 153–160.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 1, 2001, pp. 175–182.

SOLIDS
Electronic Properties

Variational Tests of Current Order Parameters in the Hubbard Model

A. A. Ovchinnikov* and M. Ya. Ovchinnikova**

Semenov Institute of Chemical Physics, Russian Academy of Sciences, ul. Kosygina 4, Moscow, 117977 Russia
**e-mail: aovchin@mpipks-dresden.mpg.de*
***e-mail: movchin@center.chph.ras.ru*

Received December 6, 2000

Abstract—Variational tests are performed for current order parameters as probable sources of the pseudogap normal state of cuprates. The calculations are carried out based on the states with correlations of the valence bond type whose formation can induce in principle both the superconducting order of the *d* symmetry and current phases. It is shown for the *t*–*t*'–U Hubbard models with a large value of $U(\sim 8t)$ and the Hubbard splitting of the conduction band that (1) phases of alternating charge and longitudinal spin currents cannot be realized and (2) transverse spin currents are not compatible with the superconducting order and they could exist against the normal-state background only within a very narrow doping region near the optimal one. This region does not correspond to the region of existence of a pseudogap in cuprates, which refutes the above-mentioned hypothesis of the pseudogap origin. The requirements to the parameters of models for which the consideration of correlations of the valence bond type yields a reasonable phase curve. The existence of current phases in the *t*–*t*'–*U*–*V* Hubbard models with a strong interaction ($V > 0.25$t) of particles in neighboring sites is predicted when the *d*-superconductivity is completely suppressed. © *2001 MAIK "Nauka/Interperiodica".*

Recent measurements [1, 2] of tunneling spectra of cuprates in different magnetic fields at different temperatures made it possible to distinguish a pseudogap from a superconducting gap, i.e., demonstrated their different nature. This refutes the interpretation of the pseudogap as a precursor of the superconducting gap [3]. Several other hypotheses [4–11] have been proposed according to which the pseudogap Δ_{ps} and superconducting gap Δ_{sc} have a different origin. In all the theories [4–11], a total gap, which is manifested in photoemission experiments (ARPES), has the form

$$\Delta \sim \sqrt{\Delta_{ps}^2 + \Delta_{sc}^2}. \quad (1)$$

The main problem is the interpretation of the inverse dependence $\Delta_{ps}(\delta)$ on the doping δ, namely, an increase in Δ_{ps} from zero at almost optimal doping δ_{opt} to large values at $\delta \longrightarrow 0$ in the so-called insufficiently doped region $\delta < \delta_{opt}$.

The emergence of a pseudogap was explained in a number of papers [4–6] by the development of new hidden order parameters in the insufficiently doped region. Thus, the authors of papers [4] considered the charge-density waves with the vector $Q = (\pi, \pi)$ as such an order parameter. However, the nature of attraction in the *s*-channel that is capable of compensating for the local repulsion $U > 0$ remains unclear. The latter is responsible for antiferromagnetism of undoped systems, i.e., the spin-density waves, which are incompatible with charge-density waves. The authors of [5, 6] discussed the possible relation of the pseudogap to the development of the order parameters such as current states of the orbital antiferromagnetic (the *d*-density wave, DDW) or the states with analogous spin currents. The question of the dependence $\Delta_{ps}(\delta)$ remains open.

In papers [7–11] based on the *t*–*t*'–*J* or *t*–*t*'–*U* models, the appearance of the pseudogap was explained by a change in the topology of the Fermi surface at the optimal doping from a "small" to a large Fermi surface. In these models, the splitting into the upper and lower Hubbard subbands is retained within a rather large doping region. This splitting was obtained in [8] from variational calculations on the basis of the correlated state with correlations of the valence bond type. Some features of the dependence of the gap anisotropy $\Delta(\varphi)$ and of the ratio Δ^{max}/kT_c on doping (see review [12]) can be explained using this approach [13] because the increase in Δ_{ps} with decreasing δ naturally follows from the structure of the lower band at $t'/t > 0$. However, models involving the lower Hubbard band yield too high an asymmetry of tunnel spectra [13]. The latter are capable of probing the density of states of a system both below and above the chemical potential value.

Finally, slave-boson technique calculations [14] suggest that the pseudogap is related to the emergence of the short-range order phase for spin-density waves, i.e., to the initial stage of the formation of spin-density waves and a dielectric Hubbard gap. In this case, the required inverse dependence $\Delta_{ps}(\delta)$ can be naturally explained. However, the pseudogap anisotropy, which is similar to anisotropy of the superconducting gap of the *d* symmetry, remains unexplained.

The aim of this paper is to verify the hypothesis about hidden current order parameters as the cause of

the pseudogap's appearance. We verified this hypothesis by performing variational calculations with a correlated state with the formation of valence bonds (a band analog of the Anderson RVB states). It was shown earlier that the effective Hamiltonian derived in this approach and, in particular, the interaction of correlated hops, which appears upon formation of valence bonds, provides the attraction of holes in the *d*-channel and the corresponding *d*-superconductivity compatible with the antiferromagnetic order. This interaction also admits in principle the existence of states with current order parameters—various current phases. As shown below, this follows from the negativity of constants at the corresponding quadratic terms in the average energy. Therefore, using this effective Hamiltonian, one can study the phase diagram of such current states and their compatibility with the superconducting order.

The calculations were performed by the method proposed in [8] for the *t*–*t*'–*U*–*V* Hubbard model:

$$\begin{aligned} H &= H(U, t) + \Delta H(V, t'), \\ \Delta H(V, t') &= V\sum_{\langle nm\rangle} n_n n_m \\ &+ t'\sum_{\langle\langle nm\rangle\rangle}\sum_{\sigma}(c^{\dagger}_{n\sigma}c_{m\sigma} + \text{H.c.}). \end{aligned} \tag{2}$$

Here, $H(U, t)$ is the Hamiltonian of the classical Hubbard model with standard parameters U and t. The additional term ΔH includes the hopping interaction t' of next-to-neighboring sites and the interaction V of neighboring sites. The variational correlated state Ψ with correlations of the valence bond type is constructed [8] using the unitary transformation of the uncorrelated state Φ:

$$\begin{aligned} \Psi &= \hat{W}(\alpha)\Phi, \quad \hat{W}(\alpha) = \exp\left[\alpha\sum_{\langle nm\rangle} Z_{nm}\right], \\ Z_{nm} &= \frac{1}{2}\sum_{\sigma}(c^{\dagger}_{n\sigma}c_{m\sigma} - \text{H.c.})(n_{n,\sigma} - n_{m,-\sigma}). \end{aligned} \tag{3}$$

The choice of the unitary operator $W(\alpha)$ with the variational parameter α was substantiated in [8]. In [8], the effective Hamiltonian $\tilde{H}$ was also derived, which acts in the basis of uncorrelated states $\{\Phi\}$:

$$\begin{aligned} H_{\text{eff}}(\alpha) &= W^{\dagger}(\alpha)HW(\alpha) \\ &\approx H + \alpha[H, Z] + \frac{\alpha^2}{2}[[H, Z], Z]. \end{aligned} \tag{4}$$

As the uncorrelated function $\{\Phi\}$, a function with a doubled magnetic cell was used for testing the possible system ordering: antiferromagnetic spin order, superconducting order of the *d*-symmetry, and current order parameters, which were not studied earlier (charge- or spin-density waves of the *d*-symmetry). The average energy

$$\langle H\rangle_{\Psi} = \langle H_{\text{eff}}\rangle_{\Phi} = \bar{H}(y_{\nu})$$

of the correlated state (3) is calculated as an explicit function of a set of one-electron averages $y_{\nu} = \langle \mathbf{y}_{\nu}\rangle_{\Phi}$ over the uncorrelated state Φ (below, the subscript $\langle\ldots\rangle_{\Phi}$ at averages is omitted). As a result, the function $\bar{H}(y_{\nu})$ for the effective Hamiltonian (2) proves to be dependent on quantities $y_{\nu} = \{r_l - d_{l0}, d_l, w_l, J_l\}_{\nu}$:

$$\begin{aligned} r_l &= \frac{1}{2N}\sum_{n,\sigma}\langle c^{\dagger}_{n\sigma}c_{n+l,\sigma}\rangle, \\ d_l &= \frac{1}{2N}\sum_{n,\sigma}(-1)^n\frac{\sigma}{|\sigma|}\langle c^{\dagger}_{n\sigma}c_{n+l,\sigma}\rangle, \\ w_l &= \frac{1}{2N}\sum_{n}\text{sgn}(l_x^2 - l_y^2)\langle c^{\dagger}_{n\uparrow}c^{\dagger}_{n+l\downarrow} + \text{H.c.}\rangle, \\ J_{\mu} &= \frac{i}{8N}\sum_{n}\sum_{s,s'}(-1)^n(\sigma_{\mu})_{s,s'} \\ &\times\langle[c^{\dagger}_{ns}c_{n+\bar{x},s'} - c^{\dagger}_{ns}c_{n+\bar{y},s'}] - \text{H.c.}\rangle. \end{aligned} \tag{5}$$

Here, r_l are the density components, d_l are similar components of the alternating spin, w_l are anomalous averages of the *d*-symmetry, and J_{μ} are current order parameters. Because of the symmetry, the quantities r_l, d_l, and w_l depend only on the modulus $l = |\mathbf{l}|$, but not on the direction of the vector $\mathbf{l} = (l_x, l_y)$. In equations (5), σ_{μ} ($\mu = 0, 1, 2, 3$) are Pauli matrices and $\bar{x}$, $\bar{y}$ are the unit vectors along the *x* and *y* axes, respectively.

Our earlier calculations [8] were related to the states without current components for models with parameters $U/t = 8$, $V/t = 0$–0.1, $t'/t = 0.05$–0.1. These models yield solutions with the antiferromagnetic splitting of the band within a rather broad region of doping and predict the superconductivity of the *d*-symmetry within the same region, which is compatible with the antiferromagnetic order. The value of δ_{opt} depended on t'/t and was $\delta_{\text{opt}} = 0.18$–$0.22$ for the above values of t'/t. It was also verified that a consideration of only the first harmonic $l = 1$ among all anomalous averages w_l describes the phase curve sufficiently accurately. For this reason, we will retain in calculations of the phase curves only the first harmonic both in anomalous averages and in current order parameters. In this approximation, the average energy per lattice center is

$$\frac{1}{N}\bar{H}(y_{\nu}) = \bar{H}^N(r_l, d_l) + k_w w_1^2 + k_J\sum_{\mu} J_{\mu}^2, \tag{6}$$

where $\bar{H}^N$ is related to the normal phase. The current components J_{μ} ($\mu = 0, 1, 2, 3$) are related to alternating charge currents ($\mu = 0$) or spin currents of different

projections of the spin. We are dealing with circular currents on elementary plaquettes of the square lattice. The quantities $J_{\mu=1,2}$ are spin currents with the spin polarization that is transverse with respect to the quantization axis z of alternating spins in d_0.

The expressions for quantities w_1 and J_μ in terms of operators in the k-space have the form

$$w_1 = \frac{1}{2N}\sum_k^G \langle c_{k\uparrow}^\dagger c_{-k\downarrow}^\dagger + \text{H.c.}\rangle \varphi^d(k), \tag{7}$$

$$J_\mu = \frac{i}{4N}\sum_k^G \sum_{s,s'} (\sigma_\mu)_{s,s'} \langle c_{ks}^\dagger c_{\tilde{k}s'} - \text{H.c.}\rangle \varphi^d(k), \tag{8}$$

$$\tilde{k} = k + (\pi,\pi), \quad \varphi^d(k) = \frac{1}{2}(\cos k_x - \cos k_y).$$

Functions $\varphi^d(k)$ in (7) and (8) have the same angular dependence of the $d_{x^2-y^2}$-symmetry. The fact that the observed pseudogap has a close anisotropy gave grounds to the authors of [5, 6] to assume the relation between the pseudogap and current order parameters. The summation over k is performed over the entire Brillouin zone G of the initial lattice.

Consider the interactions in the effective Hamiltonian (4) that can in principle induce both the d-superconductivity and current order parameters J_μ. The interaction of correlated hops of the type

$$V_{chi} = -\frac{1}{2}\alpha U \sum_{\langle nm\rangle,\sigma} (c_{n\sigma}^\dagger c_{m\sigma} + \text{H.c.}) \times (n_{n,-\sigma} + n_{m,-\sigma} - 2n_{n,-\sigma}n_{m,-\sigma}) \tag{9}$$

appears already in the first-order term over the variational parameter α in the effective Hamiltonian (4). The average of (9) over the states with order parameters (5) per lattice center is

$$\frac{1}{N}\langle V_{chi}\rangle = -8\alpha U r_1[r_0(1-r_0) + d_0^2 + r_1^2] + \kappa^{(1)}\left(w_1^2 + \sum_\mu J_\mu^2\right), \quad \kappa^{(1)} = -8\alpha U r_1. \tag{10}$$

The required decrease in the normal-state energy upon formation of valence bonds, i.e., the negative sign of the first term in (10), results in the sign of α at which constants $\kappa^{(1)}$ at terms w_1^2 and J_μ^2 in (10) are negative. This means that current phases can exist in principle. For each $\langle nm\rangle$ coupling, the interaction (9) contains the contribution $V_{nm} \sim \alpha U c_{n\sigma}^\dagger c_{n\sigma} n_{n,-\sigma} n_{m,-\sigma}$ that provides the attraction of hole in the d-channel. This term was omitted in the truncated interaction of correlated hops of the form $V_{nm} \sim c_{n\sigma}^\dagger c_{m\sigma}(n_{n,-\sigma} + n_{m,-\sigma})$. This interaction was applied by Hirsh [15] to the additional holes of an oxygen system at a low concentration. It can provide attraction only in the s-channel. However, the single-band Hubbard model (2) is relevant to the hybridized p–d band with a high, of the order of unity, total concentration of holes, and in this case the s-superconductivity is suppressed by strong repulsion U. Meanwhile, the term of type $\sim c_{n\sigma}^\dagger c_{m\sigma} n_{n,-\sigma} n_{m,-\sigma}$, which is significant at a high concentration, can provide attraction in the d-channel.

The total constants κ_w and κ_J of superconducting and current order parameters in (6),

$$\kappa_w = \kappa_w^{(1)} + \kappa_w^{(2)} + 4V, \quad \kappa_J = \kappa_J^{(1)} + \kappa_J^{(2)} - 4V, \tag{11}$$

include the first- and second-order contributions over α and the Coulomb-like interaction between neighbors. We calculated all the contributions to κ_w in our earlier studies of superconductivity [8]. For $U/t = 8$ and $V = 0$, the constant κ_w takes the values $\kappa_w(V=0) = -(1.2\text{–}1.5)t$ in the region $\delta \le 0.3$. This constant provides the d-superconductivity caused by the formation of valence bonds. The introduction of the interaction $V > 0$ suppresses the d-superconductivity by decreasing $T_c^{\max}$, and for $V \ge 0.3t$, when the constant κ_w becomes positive, the formation of pairs impossible. However, the same interaction V increases the modulus of the negative constant κ_J. The first-order contributions over α to κ_w and κ_J are identical. The second-order contribution $\kappa_J^{(2)}$ was estimated similarly as $\kappa_J^{(2)} \sim \kappa_w^{(2)}$. This estimate is adequate under conditions when the parameter V was varied within broad limits. Thus, the calculations were performed for models with constant satisfying of the following relations:

$$\kappa_J = \kappa_w - 8V, \quad \kappa_w = \frac{1}{2}\frac{\partial^2 \bar{H}}{\partial w_1^2}. \tag{12}$$

The existence of current order parameters and their compatibility with the superconducting order were studied based on the mean-field solutions for the problem with the effective Hamiltonian (4).

The procedure of minimizing $\bar{H}'$ over Φ is standard. The self-consistent uncorrelated state $\Phi = \prod_{k\lambda}\langle\chi_{k\lambda}^\dagger\rangle$ is determined by the occupation of the lower one-electron states $\chi_{k\lambda}^\dagger$ of the linearized Hamiltonian:

$$H_{\text{lin}} = \sum_\nu \frac{\partial \bar{H}}{\partial y_\nu}(\hat{y}_\nu - y_\nu) + \bar{H}(y_{\nu'}) = \sum_\sigma \sum_k^{G/4} \hat{h}_{k\sigma} + \text{const}. \tag{13}$$

Here, the operators $\hat{y}_\nu$ correspond to averages $\hat{y}_\nu$ in definitions (5) [8]. In the general case, when all the

order parameters (5) are nonzero, the vector k in the sum runs through one-fourth of the Brillouin zone G of the initial lattice, for example, the region $\{|k_x + k_y| \le \pi, k_x > 0\}$. The corresponding linear operator $\hat{h}_k$ for each k in (13) has the form

$$\hat{h}_k = h_{ij}(k) b^\dagger_{ik} b_{jk} \tag{14}$$

in the basis of the following operators of the number representation:

$$b^\dagger_{ik} = \{c^\dagger_{k\uparrow}, c^\dagger_{\tilde{k}\uparrow}, c^\dagger_{k\downarrow}, c^\dagger_{\tilde{k}\downarrow}, c_{-k\uparrow}, c_{-\tilde{k}\downarrow}, c_{-\tilde{k}\downarrow}\}_i, \tag{15}$$
$$i = 1, \ldots, 8, \quad \tilde{k} = k + (\pi, \pi).$$

The matrix $h_{ij}(k)$ is defined by expression (20) in the Appendix. The diagonalization of h_{ij} determines, according to (24), one-electron energies $E_{k\lambda}$ and eigenfunctions $\chi^\dagger_{k\lambda}$ of the linearized Hamiltonian. The latter, in turn, determine the required order parameters $\{y_\nu\}$ according to expressions (25) in Appendix.

Before studying the compatibility of current and superconducting orders, we performed a search for phase curves $T_J(\delta)$ and $T_c(\delta)$. Recall that the normal phase for solutions under consideration corresponds to the spin antiferromagnetic disorder and Hubbard splitting of the band. The energies of the corresponding solutions with the two-dimensional antiferromagnetic order are substantially lower than the energies of paramagnetic states. This does not mean that there exists a real long-range order in the system but indicates that the length of antiferromagnetic correlations is rather large. Although the mean-field approximation for H_{eff} cannot give the length of antiferromagnetic correlations, we assume nevertheless that this approximation can be used to study the short-range interaction effects produced by correlations of the valence bond type. Note also that the superconductivity region over temperature and doping proves to be reasonable only for antiferromagnetic solutions because the density of states in the lower subband increases. For paramagnetic solutions, the value of T_c and the doping width of the superconductivity region are too small.

Temperatures $T_J(\delta)$ and $T_c(\delta)$ of the appearance of the current phase with polarization μ or the d-superconductivity involving the lowest normal state with the antiferromagnetic order can be determined from equations, which represent linear expansions of the corresponding equations (25) for J_μ and w_1 for $J_\mu \longrightarrow 0$ and $w_1 \longrightarrow 0$. As a result, we obtain the following equations for $T_J(\delta)$ and $T_c(\delta)$

$$1 - \frac{1}{2N}\kappa_J \sum_k^{G/2} [g_\mu(k)\varphi^d_k]^2 \frac{1}{kT} \times f(E_1)[1 - f(E_1)] = 0, \tag{16}$$

$$1 - \frac{1}{4N}\kappa_w \sum_k^{G/2} [g_w(k)\varphi^d_k]^2 \frac{1}{|E_1|} \times [1 - 2f(|E_1|)] = 0, \tag{17}$$

$$E_1 = \bar{\xi}_k - \mu - D_k, \quad D_k = \sqrt{\delta\xi_k^2 + \Delta_k^2},$$
$$\bar{\xi}_k = \frac{1}{2}(\xi_k + \xi_{\tilde{k}}),$$
$$\delta\xi_k = \frac{1}{2}(\xi_k - \xi_{\tilde{k}}), \quad g_{\mu=1,2}(k) = g_w(k) = -\frac{\Delta_k}{D_k}, \tag{18}$$
$$g_{\mu=0,3}(k) = -\frac{\delta\xi_k}{D_k}.$$

Here, f is the Fermi function, $E_1 = E_1(k)$ is the energy of the lower Hubbard band measured relative to the chemical potential, and functions ξ_k and Δ_k are defined by expressions (21) in the Appendix. The vector k in the sum runs the values within the magnetic Brillouin zone (half the Brillouin zone G of the initial lattice). The function $g_{(k)}$ reflects the renormalization of pairing interactions, i.e., of matrix elements for the current operators J_μ or w_1 over the states of the lower Hubbard band. The constant κ_w in (17) was calculated as in [8], taking into account the contributions from anomalous averages to $\bar{H}$ from all terms of the effective Hamiltonian (2). The first-order contributions over the variational parameter α to constants κ_J and κ_w are identical. Taking this into account, we estimated κ_J from expression (12).

The calculations showed that the alternating charge and longitudinal current order parameters $J_{\mu=0,3}$ cannot be realized in models with the antiferromagnetic splitting of the band. The reason is that for $\mu = 0.3$, the function g_μ, which characterizes matrix elements of J_μ over the states of the lower band, vanishes in regions $k \sim (\pi, 0)$ which are responsible for the Van Hove singularity in the density of states. Similar functions for transverse current polarizations with $\mu = 1, 2$ in (16) and for anomalous averages of the d-symmetry tend to 1 for $k \longrightarrow (\pi, 0)$. In the absence of the real long-range antiferromagnetic order, we deal only with polarizations with respect to the local orientation of the alternating spin. The difference between the properties of transverse and longitudinal current polarizations has the same nature as that between longitudinal and transverse spin susceptibilities in models with the antiferromagnetic splitting of the band [16, 17].

Figure 1 shows phase curves $T_J(\delta)$ for transverse polarizations and $T_c(\delta)$ for models with U/t = 8 and two variants of parameters t' and V. The parameter t' directly affects the position of the Van Hove singularity and, hence, the value of δ_{opt}. This parameter was chosen to obtain reasonable values of $\delta_{\text{opt}} \sim 0.2$–$0.24$. The introduction of the interaction $V > 0$ suppresses the super-

conducting order but enhances the current order according to its contributions (12) to constants κ_w and κ_J, i.e., it reduces $T_c^{\max}$ and increases $T_J^{\max}$. We are dealing with the temperatures of emergence of one or two order parameters, superconducting or current ones, against the background of the normal state without these orders. The phase curve of the current phase restricts only a very narrow doping region near the optimal value. Its width is substantially narrower than that of the superconductivity region.

The total self-consistent calculations of superconducting and current order parameters show their incompatibility. The iteration converges either to the solution with the superconducting order without currents ($w_l \neq 0$, $J = 0$) or to the current state without anomalous averages ($J \neq 0$, $w_l = 0$). The latter exists only in a narrow doping region around the optimal value and only for the ratio of constants $\kappa_J/\kappa_w > 1.45$. In particular, for systems with $V = 0$ ($\kappa_J \sim \kappa_w$), no solutions with the current order exist: the superconducting order suppresses the current order over the entire doping region. As V increases, the solutions with $J \neq 0$ appear in a narrow doping region near the optimal value, but with anomalous averages suppressed in this region. Figure 2 shows the order parameters w_l and $J_{\mu=2}$ obtained in self-consistent calculation as functions of doping for the model with parameters $U/t = 8$, $t'/t = 0.05$, $V/t = 0.1$. However, such a behavior is inconsistent with the experiment because no suppression of the superconductivity is observed in cuprates in the middle of the region of its existence (at $\delta \sim \delta_{opt}$).

It is interesting to note that for the t–t'–U–V Hubbard models with strongly interacting ($V > 0.25t$) particles in neighboring sites, the d-superconductivity is completely suppressed; however, stable current phases appear with the transverse polarization of spin currents. Similarly to curves $T_c(\delta)$ at $V = 0$, the corresponding phase curves $T_\mu(\delta)$ have a characteristic maximum at $\delta = \delta_{opt}$ at which the chemical potential coincides with the Van Hove singularity in the density of states.

For the model with the Hubbard band splitting ($U/t \sim 8$), the possibility of superconducting correlation pairing in itself and the value of T_c depend on the parameter V. The values $V/t \leq 0.1$ used are lower than estimates $V/t \sim 0.2$–0.3, which follow from the cluster derivation of the single-band Hubbard model [18]. Another parameter $t'/t \leq 0.1$, which determines δ_{opt}, was also chosen smaller than the values $t'/t \sim 0.2$–0.3, which are commonly used in strong-coupling models with the nonsplit band [19]. For this reason, is was interesting to calculate phase curves $T_c(\delta)$ for systems with large $t'/t \sim 0.3$. In such systems, the optimum doping at which the chemical potential coincides with the Van Hove singularity proves to be quite high and lies outside the region of ferromagnetic spin ordering, where the lower-energy mean-field solutions for H_{eff} are paramagnetic. Figure 3 shows the dependences of the superconducting transition temperature $T_c(\delta)$ and the component $d_0(\delta)$ of the alternating spin on doping for models with large values of t'. They feature large values of δ_{opt} and a drastic decrease in T_c at the moment of disappearance of the antiferromagnetic band splitting (vanishing of d_0). The latter is caused by a decrease in the density of states at the Fermi boundary. However, current phases in systems with large t' are absent. It is reasonable that the mean-field method for H_{eff} only roughly describes antiferromagnetic correlations. This method cannot describe antiferromagnetic correlations with a

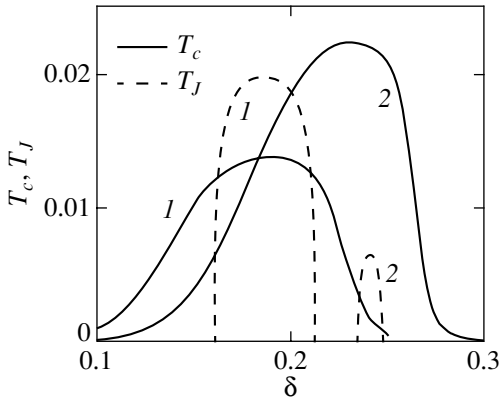

Fig. 1. Phase curves $T_c(\delta)$ and $T_J(\delta)$ corresponding to the transition of the normal-phase solutions to the superconducting state or current state with the transverse polarization for the Hubbard model with $U = 8$. Curves *1* correspond to the model with parameters $t' = 0.05$ and $V = 0.1$; curves *2*, to the model with $t' = 0.1$ and $V = 0$. All the quantities T_c, T_J, U, t', and V are expressed in units of t.

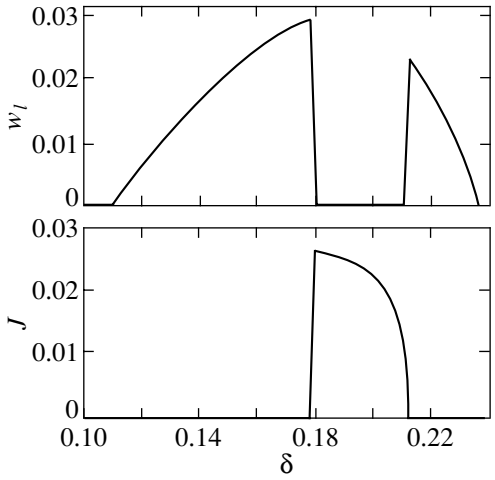

Fig. 2. Dependences of the anomalous and current order parameters on doping in the total self-consistent mean-field solution for the effective Hamiltonian (4) taking into account correlations of the valence bond type. The model parameters: $U = 8$, $t' = 0.05$, $V = 0.1$ (in units of t). In Fig. 1, these parameters correspond to phase curves *1*.

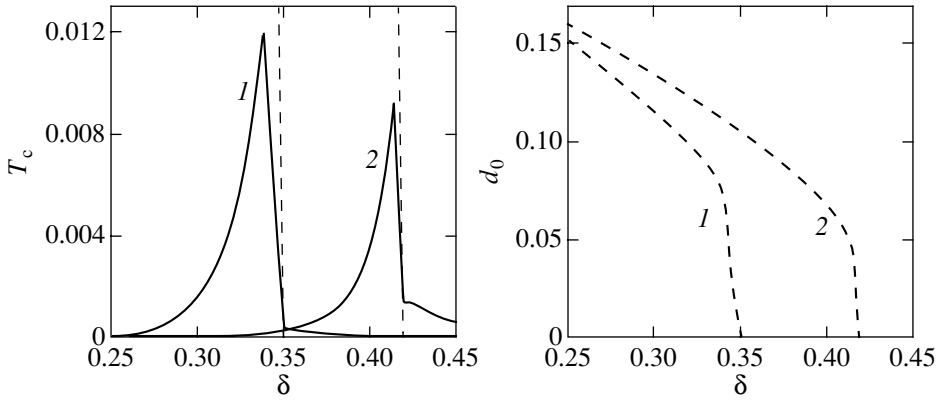

Fig. 3. Phase curves $T_c(\delta)$ and the alternating spin $d_0(\delta)$ for models with $U = 8$, $V = 0$, and $t' = 0.3$ or 0.4 (curves *1* and *2*, respectively). All the quantities and T_c are expressed in units of t.

finite radius, except for antiferromagnetic correlations of neighboring sites caused by the formation of valence bonds. It seems likely that this causes a drastic increase in d_0 and a decrease in T_c when the paramagnetic solution transfers to the antiferromagnetic one. Nevertheless, the results presented in Fig. 3 show that one can expect the inclusion of local antiferromagnetic correlations into the region of optimal doping in models with the large parameter t'. In such systems, the states with the local antiferromagnetic order or without it are realized in the regions of insufficient doping or excess doping, respectively, i.e., the initial stage of formation of the Hubbard band splitting can take place at optimal doping. This is also indicated by calculations of the Hubbard model by the slave-boson method [14], which predict the appearance of a phase with the local spin order in insufficiently doped systems ($\delta < \delta_{opt}$). However, the mean-field treatment of H_{eff} cannot provide adequate accuracy for the quantitative description of the evolution of the antiferromagnetic state to the paramagnetic one.

Thus, the narrow doping region in which solutions with the current phase exist never coincides with the region of the pseudogap behavior of cuprates. The behavior of $T_J(\delta)$ and the dependence of the Van Hove singularity splitting $\Delta_J(\delta)$ on doping related to the current order parameter drastically differ from the behavior of the corresponding quantities $T^*(\delta)$ and $\Delta_{ps}(\delta)$ for the pseudogap. The incompatibility of the superconducting and current orders predicted by the calculations also contradicts the existence of a gap and a pseudogap in the insufficiently doped region. Therefore, the hypothesis about the current phase as a reason for the pseudogap behavior of cuprates should be ruled out if the models with large U/t and the Hubbard splitting of the conduction band are adequate for the description of doped cuprates.

ACKNOWLEDGMENTS

The authors thank V.Ya. Krivnov for stimulating discussion of the problem. This work was supported by the Russian Foundation for Basic Research, grant nos. 00-03-32981 and 00-15-97334.

APPENDIX

The eight-order matrices $h_{ij}(k)$ determining the linearized Hamiltonian (14) in the basis (15) have the form

$$h_{ij}(k) = h_{ij}^N + \frac{1}{2}\frac{\partial \overline{H}}{\partial w_1}\theta_{ij}^w \varphi^d(k) + \frac{1}{2}\sum_\mu \frac{\partial \overline{H}}{\partial J_\mu}\theta_{ij}^\mu \varphi^d(k), \tag{19}$$

$$h_{ij}^N = \begin{pmatrix} a+b & & & \\ & a-b & & \\ & & -a+b & \\ & & & -a-b \end{pmatrix}, \tag{20}$$

$$a = \begin{pmatrix} \xi_k & 0 \\ 0 & \xi_{\tilde{k}} \end{pmatrix}, \quad b = \begin{pmatrix} 0 & \Delta_k \\ \Delta_k & 0 \end{pmatrix}.$$

Matrices θ^ν are defined below by expressions (26)–(28). Empty places in h^N and θ^ν are zero matrices. Functions ξ_k and Δ_k have the form

$$\xi_k = \frac{1}{2}\sum_l \frac{\partial \overline{H}}{\partial r_l}\varphi_l^s(k) - \mu, \tag{21}$$

$$\Delta_k = \frac{1}{2}\sum_l \frac{\partial \overline{H}}{\partial d_l}\varphi_l^s(k), \quad \tilde{k} = k + (\pi, \pi),$$

$$\varphi_l^s(k) = (-1)^l \varphi_l^s(\tilde{k}) = \frac{1}{2}[\cos k_x l_x \cos k_y l_y + \cos k_y l_x \cos k_x l_y], \quad (22)$$

$$\varphi_1^d(k) = \frac{1}{2}(\cos k_x - \cos k_y). \quad (23)$$

The vector $\mathbf{l} = (\mathbf{l}_x, \mathbf{l}_y)$ in harmonics (22) is one of the lattice vectors with the given modulus ($l = |\mathbf{l}| = 0, 1, \sqrt{2}, 2, \sqrt{5}, 3$ for r_l or $l = 0, \sqrt{2}, 2$ for d_l). The diagonalization of $h_{ij}(k)$,

$$h_{ij}(k)U_{j\lambda} = U_{i\lambda}E_{k\lambda}, \quad (24)$$

determines the one-electron energies $E_{k\lambda}$ and the matrix $U_{i\lambda}$ of eigenvectors in the basis (15). The latter allow one to calculate the required order parameters

$$y_\nu = \frac{1}{2N}\sum_k^{G/4} \frac{\partial \overline{H}}{\partial y_\nu} \theta_{ij}^\nu \varphi_\nu(k) U_{i\lambda}^* U_{j\lambda} f(E_{k\lambda}). \quad (25)$$

Matrices $\theta^\nu = \theta^w$ or $\theta^\nu = \theta^\mu$ for $y_\nu = w_1, J_\mu$ in (25) are

$$\theta^w = \begin{pmatrix} & & & \sigma_3 \\ & & -\sigma_3 & \\ & -\sigma_3 & & \\ \sigma_3 & & & \end{pmatrix},$$

$$\theta^{\mu=0,3} = \begin{pmatrix} \sigma_2 & & & \\ & \zeta_\mu\sigma_2 & & \\ & & \zeta_\mu\sigma_2 & \\ & & & \sigma_2 \end{pmatrix}, \quad (26)$$

$$\theta^{\mu=1} = -\begin{pmatrix} & \sigma_2 & & \\ \sigma_2 & & & \\ & & -\sigma_2 & \\ & & \sigma_2 & \end{pmatrix},$$

$$\theta^{\mu=2} = \begin{pmatrix} & \sigma_2 & & \\ -\sigma_2 & & & \\ & & -\sigma_2 & \\ & & \sigma_2 & \end{pmatrix}. \quad (27)$$

Here, $\zeta_\mu = (-1)^\mu$ and σ_ν are the Pauli matrices.

Similar matrices θ^d and θ^r for calculations of d_l and r_l are

$$\theta^d = \begin{pmatrix} \sigma_1 & & & \\ & -\sigma_1 & & \\ & & -\sigma_1 & \\ & & & \sigma_1 \end{pmatrix},$$

$$\theta^r = \begin{pmatrix} \sigma_\gamma & & & \\ & \sigma_\gamma & & \\ & & -\sigma_\gamma & \\ & & & -\sigma_\gamma \end{pmatrix}, \quad (28)$$

where $\sigma_\gamma = \sigma_0$ or $\sigma_\gamma = \sigma_3$ for calculations of r_l with "even" ($l = 0, \sqrt{2}, 2$) or "odd" ($l = 1, \sqrt{5}, 3$). The function $\varphi_\nu(k)$ in (25) used for the calculation of $y_\nu = d_l, r_l$ is equal to the corresponding l-harmonic $\varphi_l^s(k)$ of the generalized s-symmetry defined by expression (22), or the function $\varphi_\nu(k) = \varphi_1^d(k)$ is equal to the analogous d-harmonic (23) for the case $y_\nu = w_1$ or J_μ.

REFERENCES

1. V. M. Krasnov, A. E. Kovalev, A. Yurgens, and D. Winkler, cond-mat/0006479 (2000); Phys. Rev. Lett. **86**, 2657 (2001).
2. Yu. I. Latyshev, V. N. Pavlenko, S. J. Kim, *et al.*, cond-mat/0005116 (2000).
3. V. J. Emery and S. A. Kivelson, Nature **374**, 4347 (1995).
4. R. S. Markiewicz, Phys. Rev. B **56**, 9091 (1997); R. S. Markiewicz, C. Kusko, and V. Kidambi, Phys. Rev. B **60**, 627 (1999); R. S. Markiewicz and C. Kusko, Phys. Rev. Lett. **84**, 5674 (2000).
5. S. Chakravarty, R. B. Laughlin, D. K. Morr, and Ch. Nayak, cond-mat/0005443 (2000).
6. T. Dahm, D. Manske, and L. Tewordt, Phys. Rev. B **56**, R11419 (1997).
7. N. M. Plakida, V. S. Oudovenko, R. Horsch, and A. J. Liechtenstein, Phys. Rev. B **55**, 11997 (1997).
8. A. A. Ovchinnikov and M. Ya. Ovchinnikova, Phys. Lett. A **249**, 531 (1998); A. A. Ovchinnikov, M. Ya. Ovchinnikova, and E. A. Plekhanov, Pis'ma Zh. Éksp. Teor. Fiz. **67**, 350 (1998) [JETP Lett. **67**, 369 (1998)]; Zh. Éksp. Teor. Fiz. **114**, 985 (1998) [JETP **87**, 534 (1998)]; Zh. Éksp. Teor. Fiz. **115**, 649 (1999) [JETP **88**, 356 (1999)].
9. R. O. Kuzian, R. Hayn, A. F. Barabanov, and L. A. Maksimov, Phys. Rev. B **58**, 6194 (1998).
10. F. Onufrieva, P. Pfeuty, and M. Kisilev, Phys. Rev. Lett. **82**, 2370 (1999).

11. P. Lou and H. Wu, cond-mat/0003229 (2000).

12. T. Timusk and B. Statt, Rep. Prog. Phys. **62**, 61 (1999).

13. A. A. Ovchinnikov and M. Ya. Ovchinnikova, cond-mat/9912096 (1999); A. A. Ovchinnikov and M. Ya. Ovchinnikova, Zh. Éksp. Teor. Fiz. **118**, 1434 (2000) [JETP **91**, 1242 (2000)].

14. B. Kyung, cond-mat/0003492 (2000).

15. J. E. Hirsch, Phys. Rev. B **67**, 10741 (1992); J. E. Hirsch, cond-mat/0007115 (2000); cond-mat/0007328 (2000).

16. A. V. Chubukov and K. A. Musaelyan, Phys. Rev. B **51**, 12605 (1995).

17. A. A. Ovchinnikov and M. Ya. Ovchinnikova, Zh. Éksp. Teor. Fiz. **116**, 1058 (1999) [JETP **89**, 564 (1999)].

18. H. B. Schuttler and A. J. Fedro, Phys. Rev. B **45**, 7588 (1992).

19. R. J. Radke and M. R. Norman, Phys. Rev. B **50**, 9554 (1994).

Translated by M. Sapozhnikov

Journal of Experimental and Theoretical Physics, Vol. 93, No. 1, 2001, pp. 161–167.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 1, 2001, pp. 183–190.

SOLIDS
Electronic Properties

Phase Transitions in $La_{1-x}Ca_xMnO_{3-x/2}$ Manganites¶

I. O. Troyanchuk[a, *], S. V. Trukhanov[a], H. Szymczak[b, **], J. Przewoznik[c], and K. Bärner[d]

[a]*Institute of Solids and Semiconductor Physics, National Academy of Sciences of Belarus, Minsk, 220072 Belarus*
**e-mail: troyan@ifttp.bas-net.by*
[b]*Institute of Physics, Polish Academy of Sciences, 02-668, Warsaw, Poland*
***e-mail: szymh@ifpan.edu.pl*
[c]*University of Mining and Metallurgy, 30-059, Krakow, Poland*
[d]*IY Physikalishes Institut, D37073, Göttingen, Germany*

Received December 8, 2000

Abstract—The crystal structure parameters and magnetic and electrical properties of $La_{1-x}Ca_xMnO_{3-x/2}$ reduced manganites with $0 \le x \le 0.5$ are established. These investigations contribute to the understanding of magnetic interactions in manganites without Mn^{4+} ions. It is found that these manganites show a long-range antiferromagnetic order up to $x = 0.09$ and transform into spin glasses at $0.09 < x \le 0.35$. The compositions in the range $0.35 < x \le 0.5$ show a strong increase in the spontaneous magnetization and critical point associated with the appearance of spontaneous magnetization and can therefore be viewed as inhomogenious ferromagnets. The magnetic and crystal structure peculiarities of $La_{0.5}Ca_{0.5}MnO_{2.75}$ are established by the neutron diffraction method. The strongly reduced samples show a large magnetoresistance below the point where the spontaneous magnetization develops. The magnetic phase diagram of $La_{1-x}Ca_xMnO_{3-x/2}$ is established by magnetization measurements. The magnetic behavior is interpreted assuming that the Mn^{3+}–O–Mn^{3+} magnetic interaction is anisotropic (positive-negative) in the orbitally ordered phase and isotropic (positive) in the orbitally disordered phase. Introduction of the oxygen vacancies changes the magnetic interaction sign from positive to negative, thereby leading to a spin glass state in strongly reduced compounds. The results obtained reveal unusual features of strongly reduced manganites such as a large ferromagnetic component, a high magnetic ordering temperature, and a large magnetoresistance despite the absence of Mn^{3+}–Mn^{4+} pairs. In order to explain these results, the oxygen vacancies are supposed to be ordered. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The hole-doped $La^{3+}_{1-x}A^{2+}_x(Mn^{3+}_{1-x}Mn^{4+}_x)O_3$ perovskites, where A represents divalent alkaline-earth elements such as Ca, Sr, and Ba, have recently attracted much attention because of their unusual magnetic and transport properties [1, 2]. The parent compound $La^{3+}Mn^{3+}O_3$ is an antiferromagnetic insulator with the *A*-type magnetic order [3]. This compound contains Mn^{3+} ions with the $t^3_{2g}e^1_g$ ($S = 2$) electron configuration, surrounded by the oxygen octahedron. It is accepted that the t^3_{2g} and e^1_g electrons are localized and their total magnetic moment is determined only by the total spin because their orbital magnetic moment is frozen. It is known that Mn^{3+} is a Jahn–Teller ion with the higher energy e_g and the smaller energy t_{2g} electron levels in the octahedral oxygen coordination. An orbital ordering is observed and the interaction of the magnetic moments of manganese is anisotropic as a result of the Jahn–Teller effect. The antiferromagnetic spin axis is directed almost along the *b* axis and the ferromagnetic moment along the *c* axis [4].

$La_{1-x}A_xMn^{3+}_{1-x}Mn^{4+}_xO_3$ exhibits a ferromagnetic ground state and a metal–insulator transition occurring near the Curie point at $x > 0.15$. In this case, Mn^{4+} ions are created from Mn^{3+}. The Mn^{4+} ion contains t^3_{2g} electrons only ($S = 3/2$). The interplay between the electrical transport and ferromagnetism in these systems is traditionally interpreted within the framework of double exchange interaction [5–7], where the magnetic coupling between Mn^{3+} and Mn^{4+} ions is considered to result from the motion of the e_g electron between two partially filled *d* shells with the strong Hund coupling on site.

However, the double exchange alone is not sufficient to explain all the properties of these compounds [8]. It has been shown that $A^{3+}_2Mn^{4+}_2O_7$ pyrochlores also exhibit a colossal magnetoresistance despite the absence of mixed manganese valence [9]. Recently, many compounds ($(La(Nd)_{1-x}Pb_xMnO_{3-x}F_x$, $La_{0.66}Ba_{0.34}Mn_{1-x}Nb_xO_3$, etc.) have been found and they show the ferromagnetic behavior and large magnetoresistance in spite of the absence of Mn^{4+} [10]. Goodenough [11] adduced arguments for the ferromagnetism to be due not only to the double exchange but

¶This article was submitted by the authors in English.

also to the specific character of superexchange interactions in Mn^{3+}–O–Mn^{3+}, Mn^{3+}–O–Mn^{4+}, and Mn^{4+}–O–Mn^{4+} ion systems [11]. In the superexchange model, the ferromagnetic fraction of the exchange is determined by a virtual electron transfer from the half-filled e_g orbitals of the Mn^{3+} ion to the empty ones. Many recent experimental results have given evidence to the existence of a phase separation in manganites [12]. There is no general agreement concerning exchange interaction mechanisms in the hole-doped manganites. Despite numerous theoretical and experimental studies, the true nature of the colossal magnetoresistance in perovskites is still a matter of discussion.

Because Mn ions play a key role in electrical and magnetic properties of the manganites by providing charge carriers, magnetic moments, and local Jahn–Teller distortions, it is certainly worth investigating the properties of manganites containing only Mn^{3+} ions. The Mn^{3+}/Mn^{4+} ratio can be changed by various methods: (i) the substitution of Ln^{3+} ions by A^{2+} (Ca, Sr, Ba, Pb, and Cd) ions in the A sublattice of ABO_3 perovskite; (ii) the substitution of Mn ions by different transition elements (Cr, Fe, Co, etc.) in the B sublattice of perovskite, and (iii) the removal of oxygen ions, which produces a reduction process where Mn^{4+} ions are converted into Mn^{3+} and their coordination number decreases.

In this work, the third method is chosen to change the Mn^{3+}/Mn^{4+} ratio because there are not enough data on the results of its application. Moreover, very intriguing magnetic and magnetoresistance properties have been revealed for $La_{0.5}Ca_{0.5}MnO_{3-\gamma}$ [13] and $Ln_{0.5}Ba_{0.5}MnO_{3-\gamma}$ (Ln = Pr, Nd) [14, 15] systems in our previous investigations. It was shown that these compounds can exhibit both a large magnetoresistance and a ferromagnetic component. It is notable that the oxygen content in manganites can vary from 2.5 [16] to 3.27 [17]. A much smaller degree of the oxygen nonstoichiometry was found for $LaVO_{3+\gamma}$ ($0 \le \gamma \le 0.05$) and $LaTiO_{3+\gamma}$ ($0 \le \gamma \le 0.08$), while $LaCrO_3$ and $LaFeO_3$ do not perfectly reveal the deviation from the stoichiometry. This behavior may result from a much larger reduction in the ionic radius from Mn^{3+} to Mn^{4+} than it is found for other transition metal perovskites [18].

In this paper, we study the oxygen reduction effect on the magnetization and resistivity of Ca-doped manganites $La_{1-x}Ca_xMnO_{3-x/2}$ for $x \le 0.5$. These compounds have already been known at the very early stage of the experimental [3, 4, 19, 20] and theoretical [21, 22] studies of transition-metal oxides. The phase diagram of $La_{1-x}Ca_xMnO_3$ compounds with $0 \le x \le 1$ can be found elsewhere [23–26]. However, the phase diagram for $La_{1-x}Ca_xMnO_{3-x/2}$ series has not been proposed earlier. Our study shows that both magnetic and electric properties are very sensitive to the oxygen content variation.

2. EXPERIMENT

Polycrystalline $La_{1-x}Ca_xMnO_{3+\gamma}$ samples with $-0.03 \le \gamma \le 0.1$, x = 0, 0.05, 0.09, 0.12, 0.15, 0.18, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50 were fabricated using conventional ceramic technology. La_2O_3, $CaCO_3$, and MnO_2 were mixed, compacted, and prefired at 1000°C for 2 h in air. The pellets were then again ground and synthesized at 1550°C during 2 h in air, which was followed by quenching the samples with a low concentration of Ca ions from 900°C and slowly cooling the samples with a high concentration of Ca ions at a rate of 80°C/h in order to obtain the stoichiometric oxygen content. The X-ray powder diffraction data were recorded at room temperature with the DRON-3 diffractometer in CoK_α radiation. According to X-ray measurements, all the as-prepared samples were single-phase perovskites with the orthorhombic symmetry of the unit cell. The thermogravimetric study revealed the as-prepared samples in the range $0 \le x \le 0.18$ to have an oxygen content slightly above the stoichiometric value ($\gamma \le 0.1$). The excess of oxygen decreases gradually as the calcium content increases. The samples with x = 0.25, 0.30, 0.35 are stoichiometric ($\gamma \approx 0$). The samples with x = 0.40, 0.45, 0.50 have an oxygen content slightly less than the stoichiometric value ($\gamma \ge -0.03$). In order to prepare the stoichiometric samples, the compositions in the range $0 \le x \le 0.18$ were annealed in small evacuated silica tubes at 700°C during 30 h using metallic tantalum as an oxygen getter. The amount of Ta was calculated assuming that the final products are Ta_2O_5 and stoichiometric compositions $La_{1-x}Ca_xMnO_3$ according to the relation

$$La_{1-x}Ca_xMnO_{3+\gamma} + \frac{2\gamma}{5}Ta \longrightarrow La_{1-x}Ca_xMnO_3 + \frac{\gamma}{5}Ta_2O_5. \quad (1)$$

In contrast, the samples with x = 0.40, 0.45, and 0.50 were annealed in air at 900°C during 48 h. These reactions can be described by

$$La_{1-x}Ca_xMnO_{3+\gamma} + \frac{|\gamma|}{2}O_2 \longrightarrow La_{1-x}Ca_xMnO_3. \quad (2)$$

Polycrystalline $La_{1-x}Ca_xMnO_{3-x/2}$ samples with x = 0, 0.05, 0.09, 0.12, 0.15, 0.18, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50 were obtained by the topotactic reduction method in the above-mentioned manner. The reduction of the samples was performed at 900°C during 2 h in small evacuated silica tubes in presence of metallic Ta. The final oxygen content was calculated from the change in the weight of samples during the reduction. The relative error in oxygen content measurements did not exceed 1%. Therefore, the chemical formula for the reduced samples can be written as $La_{1-x}Ca_xMnO_{3-x/2\pm0.02}$. The reoxidation process leads to an increase in the weight corresponding to the loss of the weight during the reduction. According to X-ray measurements, almost all the sam-

ples from the $La_{1-x}Ca_xMnO_{3-\gamma}$ series (with $\gamma = 0$, $x/2$) were single-phase perovskites with the O^1-orthorhombic ($x \le 0.09$) or O-orthorhombic ($0.12 \le x \le 0.50$) unit cells similar to the as-prepared ones.

Magnetic and electric measurements have been performed for compositions corresponding to both $La_{1-x}Ca_x(Mn^{3+}_{1-x}Mn^{4+}_x)O_3$ and $La_{1-x}Ca_xMn^{3+}O_{3-x/2}$ series. For the magnetic measurements, an OI-3001 vibrating-sample magnetometer was used in the temperature range from 4 to 300 K. Resistivity measurements were performed using the standard four-probe method with ultrasonically deposited indium contacts. The dc-resistivity data were collected for well-sintered samples in the form of bars with $10 \times 2 \times 2$ mm dimensions over the temperature range from 77 to 350 K. The magnetoresistance MR was calculated using the formula

$$\mathrm{MR} = \frac{\rho(H) - \rho(H = 0)}{\rho(H = 0)} \times 100\%, \qquad (3)$$

where $\rho(H)$ is the resistivity in the magnetic field of 9 kOe and $\rho(H = 0)$ is the resistivity without the magnetic field.

Neutron diffraction measurements for the $La_{0.5}Ca_{0.5}MnO_{2.75}$ sample were performed in the Berlin Neutron Scattering Center using an E2 Flat Cone diffractometer with the wavelength of neutrons λ = 1.79635 Å.

3. RESULTS AND DISCUSSION

The crystal structure parameters for both $La_{1-x}Ca_xMnO_3$ and $La_{1-x}Ca_xMnO_{3-x/2}$ series are displayed in Fig. 1. Both stoichiometric and reduced series exhibit orthorhombic distortions in the entire range of the calcium concentration; however, O^1-orthorhombic distortions ($c/\sqrt{2} < a \le b$) transform into O-orthorhombic ones ($a < c/\sqrt{2} < b$) at $x \sim 0.1$. According to Goodenough, the O^1 distortions are caused by the orbital ordering, which is a result of the cooperative static Jahn–Teller distortions of Mn^{3+} in $LaMnO_3$. As a dopant ion concentration increases, the removal of cooperative Jahn–Teller distortions is observed [25]. The reduced compounds have only Mn^{3+} ions; however, oxygen vacancies destabilize the parent orbital ordering, and therefore, the Jahn–Teller distortions. For the samples with $x < 0.1$, oxygen vacancies are insufficient in order to remove the cooperative Jahn–Teller distortions and the unit cell of these samples has the O^1-orthorhombic symmetry. For both series, the volume of the unit cell decreases gradually as the calcium content increases. This is explained by the decrease in the size effect contribution to the crystal structure distortions. However, this process is much less pronounced for $La_{1-x}Ca_xMnO_{3-x/2}$ series. The appearance of vacancies leads to a decrease in the average oxidative state of manganese. It is well known that the ionic radius of Mn^{3+} is larger than that of Mn^{4+}. The effective ionic radii of Mn^{3+} and Mn^{4+} in the octahedral oxygen coordination are 0.645 and 0.530 Å, respectively [27]. The vacancies must therefore reduce the unit cell volume, whereas the transformation of Mn^{4+} into Mn^{3+} must give rise to it. Our data indicate that the latter process dominates.

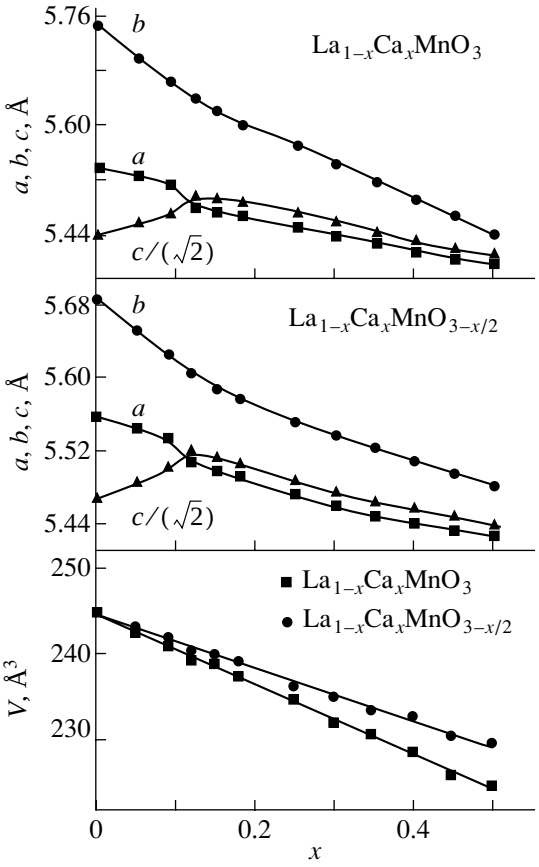

Fig. 1. The crystal structure parameters for $La_{1-x}Ca_xMnO_3$ stoichiometric (top panel) and $La_{1-x}Ca_xMnO_{3-x/2}$ reduced series (medium panel). The bottom panel displays the unit cell volume as a function of Ca concentration.

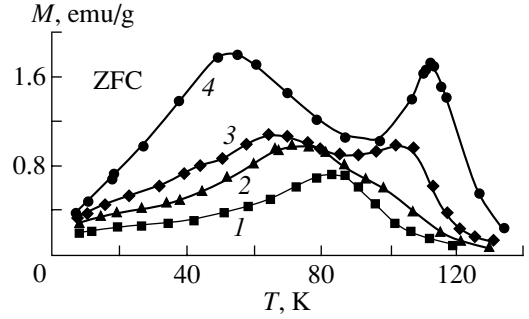

Fig. 2. ZFC magnetization versus temperature for samples with x = 0.09 (curve *1*), 0.12 (*2*), 0.15 (*3*), and 0.18 (*4*).

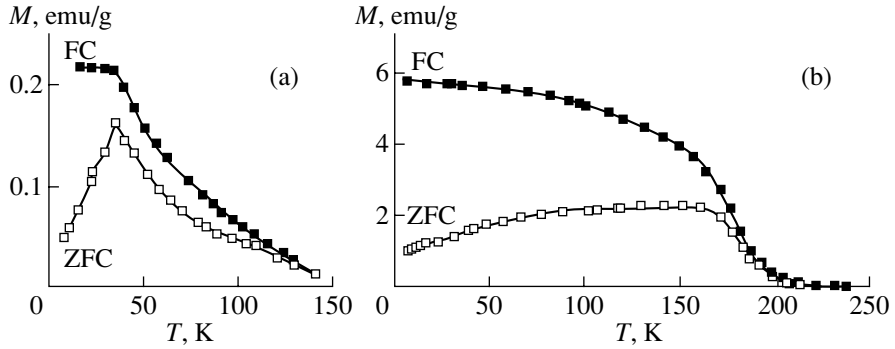

Fig. 3. ZFC and FC magnetizations versus temperature for (a) $x = 0.30$, (b) $x = 0.50$, $H = 100$ Oe.

The zero-field-cooled (ZFC) and field-cooled (FC) magnetizations in the field of 100 Oe for $La_{1-x}Ca_xMnO_{3-x/2}$ series are presented in Figs. 2 and 3. The samples in the ranges $0 \le x \le 0.12$ and $0.30 \le x \le 0.50$ show one ZFC-magnetization peak. The samples with $x = 0.15$, 0.18, 0.25 have two ZFC-magnetization peaks. The second magnetization peak at higher temperatures can be ascribed to a large magnetic anisotropy of these samples. The temperature corresponding to the largest magnetization on the $M(T)$ curve gradually decreases as the Ca concentration increases. We adopt two methods of estimating the critical temperature T_{cr} at which the magnetic transition occurs: (i) the onset of magnetic transition; T_{cr} is defined as the temperature point where ZFC and FC magnetizations become different in the field of 100 Oe; (ii) the completion of magnetic transition; T_{cr} is defined as the temperature point where the ZFC magnetization reaches its maximum value on the *M(T)* curve measured in the field of 100 Oe. In the range $0 \le x \le 0.09$, these two temperatures are close to each other, which indicates a well-defined transition to the magnetically ordered state. The samples in the interval $0.12 \le x \le 0.18$ show an entirely different magnetic behavior. The ZFC magnetization for the sample with $x = 0.15$ demonstrates two peaks on the $M(T)$ curve, which implies a complex character of the magnetic ordering in this composition. The transition to the paramagnetic state remains narrow for all the compositions with $x \le 0.18$. For the $x = 0.30$ sample (Fig. 3), ZFC and FC magnetizations start to differ around 140 K, however a ZFC-magnetization peak is observed at a sufficiently low temperature 35 K. The ZFC-magnetization peak shifts towards high temperatures up to 160 K as the Ca concentration reaches $x = 0.50$.

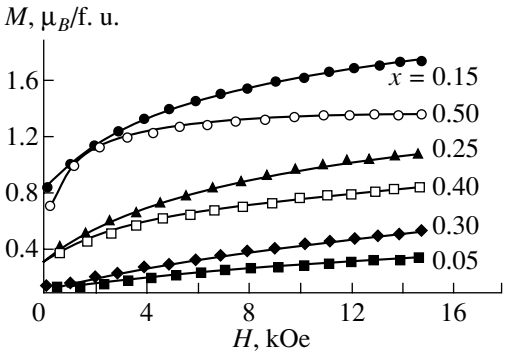

Fig. 4. The magnetization versus magnetic field curves for $La_{1-x}Ca_xMnO_{3-x/2}$ samples with $x = 0.05$, 0.15, 0.25, 0.30, 0.40, 0.50 measured at 6 K.

In Fig. 4, the magnetization is shown as a function of field at temperature of 6 K. It is difficult to estimate the spontaneous magnetization because for a majority of the reduced samples, the magnetization is not saturated in the field up to 1.6 T. The large slope in the $M(H)$ curve could be attributed to magnetic inhomogeneity or large magnetic anisotropy of these samples. The spontaneous magnetization first increases with increasing Ca content up to the $x = 0.15$ composition and then decreases. The largest spontaneous magnetization $M_s = 1.35\mu_B$ per formula unit (μ_B/f.u.) corresponds to the $x = 0.15$ composition; however, even in this case there is no pure ferromagnetic ordering because the expected value for the parallel ordering of all Mn^{3+} magnetic moments is around $4\mu_B$/f.u. The minimum spontaneous magnetization $M_s = 0.19\mu_B$/f.u. occurs for the $x = 0.30$ composition, where spin-glass properties are most pronounced. The surprise is that M_s rises again starting from the $x = 0.40$ composition and reaches $1.26\mu_B$/f.u. for the $x = 0.50$ composition.

For the $La_{0.5}Ca_{0.5}MnO_{2.75}$ sample, two neutron diffraction patterns were collected at 250 K (in the paramagnetic state) and 1.6 K to check the character of its magnetic ground state. The patterns were Rietveld refined with the two-phase model. In the refinements, the pattern profile was simulated by a split pseudo-Voigt function and the background was fitted to a seventh-degree polynomial function. The $La_{0.5}Ca_{0.5}MnO_{2.75}$ phase shows a clear asymmetric line broadening indicating the existence of large microstrains in this phase. The problem with the appropriate modeling of the (*hkl*)-dependent shape of the lines is the main source of the difference between the measured pattern and fitted

Structural parameters obtained in the Rietveld refinement of the NPD pattern with $\lambda = 1.79635$ Å; *Pbnm* space group

$La_{0.5}Ca_{0.5}MnO_{2.75}$ phase at 250 K*				
Atom	*x*	*y*	*z*	B_{iso}, Å^2***
La/Ca (4*c*)	0.013(8)	0.005(10)	0.25	0.75(36)
Mn (4*b*)	0	0.5	0	0.97(39)
O1 (4*c*)	0.049(16)	0.485(16)	0.25	3.55(39)
O2 (8*d*)	0.734(11)	0.263(13)	0.030(6)	3.55(39)
$La_{0.5}Ca_{0.5}MnO_{2.75}$ phase at 1.6 K**				
Atom	*x*	*y*	*z*	B_{iso}, Å^2
La/Ca (4*c*)	0.995(13)	0.008(8)	0.25	0.57(35)
Mn (4*b*)	0	0.5	0	0.78(37)
O1 (4*c*)	0.054(13)	0.489(13)	0.25	3.12(30)
O2 (8*d*)	0.733(9)	0.267(9)	0.030(6)	3.12(30)

* Cell parameters are $a = 5.428(4)$ Å, $b = 5.414(5)$ Å, $c = 7.675(6)$ Å; the total number of reflections is 165; the conventional Rietveld *R* factors are $R_P = 5.72\%$, $R_{WP} = 7.85\%$, $R_{exp} = 3.72\%$; the Bragg R_B is 7.73%; $\chi^2 = 4.77$.

** Cell parameters are $a = 5.421(5)$ Å, $b = 5.406(5)$ Å, $c = 7.664(7)$ Å; the total number of reflections is 165; the conventional Rietveld *R* factors are $R_P = 6.04\%$, $R_{WP} = 8.07\%$, $R_{exp} = 3.75\%$; the Bragg R_B is 6.73%; $\chi^2 = 4.62$; the magnetic *R*-factor is 5.60%.

*** B_{iso} is an isotropic temperature factor

curve. The $La_{0.5}Ca_{0.5}MnO_{2.75}$ phase was fitted assuming orthorhombic *Pbnm*.

The most important structural parameters and agreement factors of the refinement are listed in the table. The refined low-temperature value of the magnetic moment on the manganese atom is $\mu = 0.81(45)\mu_B$. The refined value of the Mn moment is not sensitive to the Mn spin direction.

According to electric resistivity measurements, all the reduced samples are semiconductors. The resistivity markedly increases as the temperature decreases (Fig. 5). There is no metal-insulator transition even for the $x = 0.15$ sample, which shows the largest ferromagnetic component in the entire series. The magnetoresistance gradually increases below the point where the magnetic order develops. However, there is no magnetoresistance peak observed in the mixed-valence ferromagnetic manganites.

Summarizing our magnetization data, we constructed a hypothetic magnetic phase diagram of $La_{1-x}Ca_xMnO_{3-x/2}$ series (Fig. 6), i.e., a dependence of the magnetic transition temperature on the calcium concentration. The phase diagram of $La_{1-x}Ca_xMnO_3$ compounds (with $0 \le x \le 1$) can be found elsewhere [23–26]. The $x = 0$ and $x = 1$ members of the $La_{1-x}Ca_xMnO_3$ system (namely, $LaMnO_3$ and $CaMnO_3$) are antiferromagnetic insulators at low temperatures, with the A- and G-type of magnetic ordering, respectively [3]. The G-type ordering is the antiferromagnetic ordering on the nearest-neighbor magnetic sites. Adding Ca ions destroys the antiferromagnetic order. The ferromagnetic behavior starts to manifest itself at $x \approx 0.1$, and the compositions with *x* up to 0.3 have both antiferromagnetic and ferromagnetic characteristics. The composition with $x = 0.3$ is clearly ferromagnetic, while the compositions with $x > 0.5$ are antiferromagnetic. In agreement with extensive NMR data [12, 28], the concentration phase transitions go through the mixed two-phase state at the dopant ion values $x < 0.1$ and $x > 0.5$ [29].

The diagram of $La_{1-x}Ca_xMnO_{3-x/2}$ series (Fig. 6) consists of six regions. The concentration boundaries are traced through the critical points. The compositions in the range $0 \le x \le 0.09$ demonstrate a magnetization peak, whereas the spontaneous magnetization gradually increases. We assume that the ground state is antiferromagnetic in this range, probably of the A-type similar to that observed for the parent $LaMnO_3$ composition [3]. The ferromagnetic component could be attributed to the noncollinear magnetic structure or ferromagnetic clusters. It is well known that noncollinear magnetic structure must result from the double exchange between Mn^{3+}/Mn^{4+} ions [6]. However, the $La_{1-x}Ca_xMnO_{3-x/2}$ system does not contain Mn^{3+}/Mn^{4+} pairs. Therefore, the magnetic behavior results from superexchange interactions between Mn^{3+}–O–Mn^{3+}. In this analysis, we assumed that the ferromagnetic component develops because of ferromagnetic clusters associated with the domains where the static Jahn–Teller distortions are removed. According to Goodenough's considerations [11], the Mn^{3+}–O–Mn^{3+} superexchange magnetic interaction is ferromagnetic and no static Jahn–Teller correlations occur. The removal of the static Jahn–Teller distortions is provided by a small amount of the oxygen vacancies that weakly affect the anion coordination of the majority of Mn^{3+} ions. Taking all this into account, it seems reasonable to assume that in the range $0 \le x \le 0.09$, our compounds are ferromag-

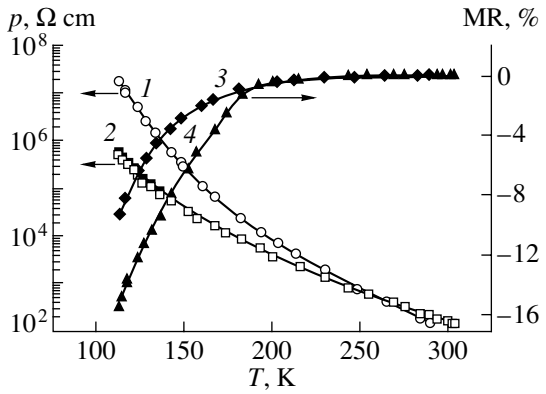

Fig. 5. The resistivity and magnetoresistance versus temperature for samples with $x = 0.15$ (curves *1* and *3*) and $x = 0.50$ (curves *2* and *4*).

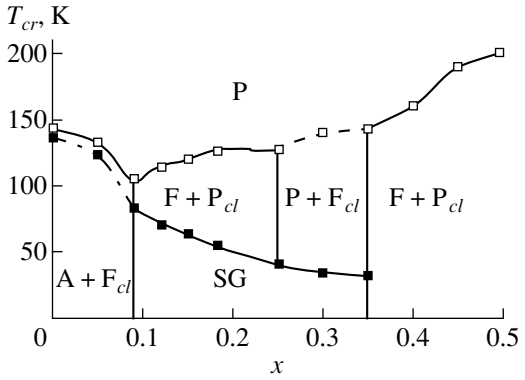

Fig. 6. Magnetic phase diagram of $La_{1-x}Ca_xMnO_{3-x/2}$. A denotes the antiferromagnet with the magnetic structure of the A-type, F_{cl} are ferromagnetic clusters, $F + P_{cl}$ is the inhomogeneous ferromagnet, SG is the spin glass, and P is the paramagnet. The solid line indicates critical temperatures. The dash-and-dot line traced through the full squares is not significant. The dashed line is traced through $x = 0.30$ because it is a singular point.

netic clusters in the antiferromagnetic medium. The open squares in Fig. 6 denote the magnetic transition onset temperature. The full squares demonstrate the ZFC-magnetization peak temperature. In the range $0 \le x \le 0.09$, these two temperatures are sufficiently close, and the dash-and-dot line traced through the filled symbols is not significant.

We note that the appearance of the oxygen vacancies stabilizes the other local orbital state in the two nearest Mn^{3+} ions rather than the state in the parent matrix. This process gradually destroys the long-range orbital ordering inherent to $LaMnO_3$, thereby leading to a collapse of the long-range orbital order at $0.09 \le x \le 0.35$. However, the pure ferromagnetic ground state does not develop in this region. Magnetization data (Figs. 2 and 3) indicate that the ferromagnetic component is strongly destabilized, which is likely due to a competition between antiferromagnetic and ferromagnetic exchange interactions. We think that the nature of the antiferromagnetic interactions in this region is different from that for the parent $LaMnO_3$. It is well known that $LnMnO_3$ (Ln = Y, Er, Ho, Lu, Sc) with the hexagonal structure are antiferromagnets with T_N around 80 K [30]. In these compounds, the Mn^{3+} ions are located in the 5-fold coordination. For example, magnetic properties of the $La_{0.85}Ca_{0.15}MnO_{2.92}$ composition can be understood assuming that Mn^{3+} (fivefold coordination)–O–Mn^{3+} (fivefold coordination) and Mn^{3+} (fivefold coordination)–O–Mn^{3+} (sixfold coordination) exchange interactions are antiferromagnetic, whereas the Mn^{3+} (sixfold coordination)–O–Mn^{3+} (6-fold coordination) exchange interactions are ferromagnetic. For compositions in the range $0.09 < x \le 0.35$, the long-range ferromagnetic ordering is not realized, which is likely due to the increasing oxygen vacancy number above the critical concentration. In this region, the transition to the magnetic ordering state goes through two stages. This can be understood from the ZFC and FC curves. The ZFC magnetization value first starts to increase, reaches its maximum, and then rapidly decreases. The FC magnetization behaves differently. It does not undergo a fall at low temperatures. This magnetization fall may be explained by the magnetic interaction energy between ferromagnetic clusters being insufficient to align their magnetic moments after zero-field cooling. After the field cooling, however, the magnetic moment directions remain the same. This magnetization behavior is typical of the spin glasses where magnetic clusters of magnetic moments are gradually blocked with decreasing temperature. The dashed line is traced through $x = 0.3$ because it is a singular point where the spontaneous magnetization is minimal and the temperature coincides with that of the ZFC magnetization peak. It is possible that the magnetic behavior of the compounds in the vicinity of this point has sharper modifications than those presented in this diagram.

The developing ferromagnetic component in the compositions with the calcium concentration $0.35 < x \le 0.50$ may be the result of a short-range ordering of oxygen vacancies. Apparently, the oxygen vacancies tend to order such that the ferromagnetic part of the exchange interaction between Mn^{3+} placed in the 6-fold coordination becomes more intense than the antiferromagnetic one. In this region, the samples appear to be inhomogeneous ferromagnets. In support of this hypothesis, let us recall several research results. Recently, it was found from high-resolution electron microscopy and selected-area electron diffraction measurements that the stoichiometric $La_{0.5}Ca_{0.5}MnO_3$ and reduced $La_{0.5}Ca_{0.5}MnO_{2.75}$ compounds differ by the domain sizes. The reduced samples have a much smaller domain size and a larger number of domains than the stoichiometric ones. It is therefore reasonable to assume the oxygen vacancies to be accommodated in the domain walls [16].

It is interesting to note that the compositions with $x = 0.15$ and $x = 0.50$ show a relatively large magnetoresistance despite the absence of the mixed valence in manganese. We suppose that the electric conductivity of reduced samples has an impurity nature. There are impurity levels associated with a small number of Mn^{2+} or Mn^{4+} ions. These impurity states are located near the structure defects, such as the oxygen vacancies, and are probably characterized by very wide energy spectra.

The magnetoresistance may result from a strong decrease of the energy gap between shallow impurity levels and a wide conduction band induced by the external magnetic field applied to the magnetically ordered state.

4. CONCLUSIONS

We can summarize our results as follows.

1. The sign of the Mn^{3+}–O–Mn^{3+} superexchange magnetic interaction depends on the orbital orientation in the orbitally ordered phases of manganites. The magnetic structure can be deduced from the orbital ordering and Goodenough–Kanamori rules [11].

2. Oxygen vacancies remove the orbital ordering in the manganites. This process is in some aspects analogous to the Mn^{4+} doping of the $LaMnO_3$ parent compound.

3. In the orbitally disordered phase, the sign of the Mn^{3+}–O–Mn^{3+} superexchange interaction depends on the oxygen neighborhood. When both Mn^{3+} ions are in a six-fold oxygen surrounding, the Mn^{3+}(VI)–O–Mn^{3+}(VI) magnetic interaction is ferromagnetic. In the case where both Mn^{3+} ions or one of them are in a five-fold oxygen surrounding, the Mn^{3+}(V)–O–Mn^{3+}(V) and Mn^{3+}(VI)–O–Mn^{3+}(V) interactions are antiferromagnetic.

4. The $La_{1-x}Ca_xMnO_{3-x/2}$ system with $x \geq 0.35$ is decomposed into clusters with different chemical compositions. The clusters with a high calcium content are ferromagnetic, whereas those with a low calcium content are antiferromagnetic.

5. The manganites can exhibit a large magnetoresistance despite the absence of Mn^{3+}–Mn^{4+} pairs. These data support the superexchange picture of magnetic interactions in manganites.

This work was supported in part by the Belarus Foundation for Basic Research (grant no. F99R-038), the German Academic Exchange Service (DAAD) and a NATO linkage grant (PST.CLG 975703).

REFERENCES

1. S. Jin, T. H. Tiefel, M. McCormack, *et al.*, Science **264**, 13 (1994).
2. S. Yunoki, J. Hu, A. L. Malvezzi, *et al.*, Phys. Rev. Lett. **80**, 845 (1998).
3. E. O. Wollan and W. C. Koehler, Phys. Rev. **100**, 545 (1955).
4. G. Matsumoto, J. Phys. Soc. Jpn. **29**, 606 (1970).
5. C. Zener, Phys. Rev. **82**, 403 (1951).
6. P.-G. De Gennes, Phys. Rev. **118**, 141 (1960).
7. P. W. Anderson and H. Hasagawa, Phys. Rev. **100**, 675 (1955).
8. A. J. Millis, P. B. Littlewood, and B. I. Shraiman, Phys. Rev. Lett. **74**, 5144 (1995).
9. M. A. Subramanian, B. H. Tobi, A. P. Ramírez, *et al.*, Science **273**, 81 (1996).
10. I. O. Troyanchuk, D. D. Khalyavin, E. F. Shapovalova, *et al.*, Phys. Rev. B **58**, 2422 (1998).
11. J. B. Goodenough, A. Wold, R. J. Arnott, and N. Menyuk, Phys. Rev. **124**, 373 (1961).
12. G. Allodi, R. De Renzi, F. Licci, and M. W. Piepper, Phys. Rev. Lett. **81**, 4736 (1998).
13. I. O. Troyanchuk, D. D. Khalyavin, S. V. Trukhanov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **70**, 583 (1999) [JETP Lett. **70**, 590 (1999)].
14. I. O. Troyanchuk, S. V. Trukhanov, H. Szymczak, *et al.*, J. Phys.: Condens. Matter **12**, L155 (2000).
15. I. O. Troyanchuk, D. D. Khalyavin, S. V. Trukhanov, *et al.*, J. Phys.: Condens. Matter **11**, 8707 (1999).
16. J. M. Gonzalez-Calbet, E. Herrero, N. Rangavittal, *et al.*, J. Solid State Chem. **148**, 158 (1999).
17. I. Maurin, P. Barboux, Y. Lassailly, *et al.*, J. Magn. Magn. Mater. **211**, 139 (2000).
18. B. C. Tofield and W. R. Scott, J. Solid State Chem. **100**, 183 (1974).
19. G. Matsumoto, J. Phys. Soc. Jpn. **29**, 615 (1970).
20. C. W. Searle and S. T. Wang, Can. J. Phys. **47**, 2023 (1969).
21. K. Kubo, J. Phys. Soc. Jpn. **33**, 21 (1972).
22. K. Kubo, J. Phys. Soc. Jpn. **33**, 929 (1972).
23. P. Shiffer, A. P. Ramírez, W. Bao, *et al.*, Phys. Rev. Lett. **75**, 3336 (1995).
24. A. Urushibara, Y. Moritomo, T. Arima, *et al.*, Phys. Rev. B **51**, 14103 (1995).
25. I. O. Troyanchuk, Zh. Éksp. Teor. Fiz. **102**, 251 (1992) [Sov. Phys. JETP **75**, 132 (1992)].
26. T. I. Arbuzova, I. B. Smolyak, S. V. Naumov, *et al.*, Zh. Éksp. Teor. Fiz. **116**, 1664 (1999) [JETP **89**, 899 (1999)].
27. R. D. Shannon, Acta Crystallogr., Sect. A: Cryst**.** Phys., Diffr. Theor. Gen. Crystallogr. **A32**, 751 (1976).
28. G. Papavassiliou, M. Fardis, M. Belesi, *et al.*, Phys. Rev. B **59**, 6390 (1999).
29. C. N. R. Rao, R. Manesh, A. K. Raychaudhuri, *et al.*, J. Phys. Chem. Solids **59**, 487 (1998).
30. W. C. Koehler, H. L. Yakel, E. O. Wollan, *et al.*, Phys. Lett. **9**, 93 (1964).

Journal of Experimental and Theoretical Physics, Vol. 93, No. 1, 2001, pp. 168–179.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 1, 2001, pp. 191–204.

SOLIDS
Electronic Properties

Quantum States and Fermi Surfaces in Metals with an fcc Lattice in an Ultrastrong Magnetic Field

V. Ya. Demikhovskii* and D. V. Khomitsky

Nizhni Novgorod State University, pr. Gagarina 23, Nizhni Novgorod, 603600 Russia
**e-mail: demi@phys.unn.runnet.ru*

Received December 8, 2000

Abstract—The expression for the electron wave function for a 3D crystal in a constant magnetic field is obtained in the strong coupling approximation. A 3D Harper-type equation describing the electron spectrum in magnetic 3D subbands is derived. The Fermi surfaces for monovalent noble metals are constructed for various orientations and magnitudes of magnetic fields corresponding to a rational number p/q of the magnetic flux quanta; radical changes in the topology of the Fermi surfaces in a strong magnetic field are observed. As a result, considerable changes in the physical properties of crystals in a strong magnetic field can be expected. In particular, a metal–semiconductor transition occurs for all even values of q, while metallic properties are preserved for odd values of q. The total energy of electrons as a function of the magnetic field is also calculated and shows a minimum for $p/q = 1/2$. The type of thermodynamic oscillations in an ultrastrong magnetic field is discussed. The effects considered by the authors may be observed in fields with a strength of several tens of megagausses.

1. INTRODUCTION

The problem of Bloch electrons in a strong magnetic field for which the magnetic flux through a unit cell is comparable with a flux quantum (the attainment of such a field is discussed, for example, in [1]) has been considered for the last several decades. In the pioneering works by Harper [2], Zilberman [3], Azbel [4], and Hofstadter [5], basic properties of the quantum states of a Bloch 2D electron in a transverse magnetic field were established. For example, it was proved that in the case of a strong periodic potential, a complex energy spectrum is formed, which strongly depends on the number of magnetic flux quanta through a unit cell and is determined by the ratio $\Phi/\Phi_0 = p/q$, where Φ is the magnetic flux, $\Phi_0 = 2\pi\hbar c/|e|$ is the magnetic flux quantum, and p and q are coprime integers. The corresponding energy spectrum is known as the Hofstadter "butterfly" [5]. In the case of a weak periodic potential, the Landau levels split into magnetic subbands whose number is equal to p [6, 7].

During the last decade, one-electron states in semiconducting 2D superlattices in a transverse magnetic field were studied intensely. The interest in these objects is due to the physical effects typical of electrons with a Hofstadter butterfly-type spectrum, which may be observed in principle in such structures owing to considerable advances in nanotechnology. This problem was considered in many theoretical [6–10] and experimental [11, 12] works. For example, the effects of magnetic breakdown in quasi-two-dimensional organic conductors were studied in [9, 10] by using the strong coupling method, which is employed for deriving the Harper equation, and experimental data confirming the observation of the energy band structure of electrons in a magnetic field, which is typical of Hofstadter butterfly-type spectra, were presented in [12].

The problem of quantum states in 3D crystals in a strong magnetic field was also investigated. Peter *et al.* [13] proved that a simple broadening of magnetic bands occurs in a simple cubic lattice in a field oriented along (0, 0, 1) upon a transition from the 2D to 3D spectrum. Fractional quantization of the Hall conductivity was detected in [14] in a 3D anisotropic lattice in a tilted field when the vector **H** lies in the *yz* plane. The total energy of electrons and the density of states were calculated by Hasegawa [15] for a simple cubic lattice for certain field orientations ((0, 0, 1), (0, 1, 1), and (1, 1, 1)); in some cases, a decrease smaller than an increase in the total electron energy in a magnetic field was observed. These investigations were carried out for an arbitrary orientation of the field in a simple cubic lattice by Kunszt and Zee [16], who also calculated the total electron energy and the density of states. They also observed that the total energy decrease again in a magnetic field. The constant-energy surfaces in the magnetic Brillouin zone for a simple cubic lattice in a magnetic field with orientations (0, 0, 1) and (1, 1, 0) were constructed for the first time in [17]. These surfaces make it possible to predict the kinetic and thermodynamic properties of a metal (the type of magnetic susceptibility, magnetic breakdown, conductivity oscillations, and so on) for magnetic flux values $p/q = 1/2$ and 1/3. It was also proved that in the vicinity of prime rational values of p/q (e.g., for $p/q = 1/2 + 1/q'$, where $q' \gg 1$), a system of narrow magnetic subbands (of the type of Landau levels) is formed,

which can be calculated using the Onsager–Lifshits quantization rule [18] on a new Fermi surface.

In the present work, the electron states of more realistic 3D crystals with an fcc structure in an ultrastrong magnetic field are considered. Vector **H** may be directed along any translational axis in the crystal. In a weak magnetic field, discrete Landau levels (magnetic tubes in the 3D problem) are formed in a metal, whose positions are determined by the semiclassical Onsager–Lifshits quantization rule, while in a strong magnetic field, magnetic energy subbands (whose number is q) are formed. Each magnetic subband corresponds to the range of definition of quasimomentum, which is known as the Brillouin magnetic band. The kinetic and thermodynamic properties of the electron gas in this case are determined by the spectrum and by the shape of the Fermi surface in the magnetic subbands. We studied the states of a Bloch electron in crystals with an fcc lattice (Cu, Ag, Au, etc.) placed in a magnetic field oriented along (0, 0, 1) and (1, 1, 0).

In Section 2, the choice of the unit cell and the Brillouin zone in a magnetic field is considered. In Subsection 3.1, the strong coupling approximation is used for constructing the wave function satisfying the generalized Bloch–Peierls conditions [19]. In Subsections 3.2 and 3.3, the energy spectrum is determined and the Fermi surfaces are constructed in the vicinity of prime values of the number of magnetic flux quanta (1/2 and 1/3). It is shown that the physical properties of metals with an fcc lattice may change radically in a magnetic field. For instance, in the $\mathbf{H} \parallel (0, 0, 1)$ orientation for even values of the denominator of p/q, when the energy bands contact one another, the metal is converted into an insulator (semimetal). For odd values of q, metallic properties are preserved, but the number of electrons in the upper partially filled band is of the order of N/q, where N is the total number of free electrons in the metal. The quantum states and the Fermi surface for the field orientation along (1, 1, 0) for $p/q = 1/2$ are investigated in Section 4. The formation of magnetic subbands for energies corresponding to open orbits is investigated for a magnetic flux close to $p/q = 1/2$. In Section 5, peculiarities of the de Haas–Van Alphen effect in a strong magnetic field are considered. It is shown that in the interval $0 \le p/q \le 1$, several series of magnetic susceptibility oscillations must be observed in the region of fields corresponding to values of p/q with small denominators q. The effect of Landau level broadening on the type of magnetic oscillations is also considered. The total energy of the electron gas for $\mathbf{H} \parallel (0, 0, 1)$ is also calculated at characteristic points $p/q = 0$, 1/4, 1/3, 1/2, 2/3, 3/4, and 1. It is shown that the minimum is attained for $p/q = 1/2$. The latter circumstance indicates that in an ultrastrong magnetic field (with the given orientation), a monovalent metal with an fcc lattice becomes paramagnetic.

Obviously, the experimental observation of the effects considered by us here requires crystals with a large lattice period and ultrastrong magnetic fields. The record value of the magnetic field (H = 28 MG) was attained recently at the Russian Federal Nuclear Center (VNIIF) (Sarov). If the crystal lattice period is 3–4 Å, the magnetic flux in such fields is $p/q \sim 0.1$, and we can speak only on the Landau level broadening. In lattices with a period of 5–6 Å, $p/q \sim 0.2$–0.3, and the effects considered below can be observed experimentally.

2. MAGNETIC TRANSLATIONS AND MAGNETIC UNIT CELL

Let us first determine the law of transformation of wave eigenfunctions in a magnetic field and the shape of the magnetic Brillouin zone. This will be done mainly following [20] as well as [21]; however, it will be convenient to use the Landau gauge. Let the magnetic field **H** be directed along a certain translation $\mathbf{a}_3$. We choose the Cartesian system of coordinates (x_1, x_2, x_3) so that one of the axes (we denote it by x_3) is directed along the magnetic field, while the other two axes lie in the plane perpendicular to **H**. The law of transformation of the wave function upon translations is determined by the vector potential gauge; we choose the vector potential in the form $\mathbf{A} = (0, Hx_1, 0)$.

Let $\psi(\mathbf{r})$ be a certain eigenfunction of the Hamiltonian $\hat{H}(\mathbf{r})$. Under the translation $\mathbf{r} \longrightarrow \mathbf{r} + \mathbf{a}$ ($\mathbf{a}$ is a lattice period), this function is transformed into $\psi(\mathbf{r} + \mathbf{a})$, but it is now an eigenfunction of the Hamiltonian $\hat{H}(\mathbf{r} + \mathbf{a})$ which does not coincide with $\hat{H}(\mathbf{r})$ since the vector potential has been transformed as follows:

$$\mathbf{A}(\mathbf{r}) \longrightarrow \mathbf{A}(\mathbf{r} + \mathbf{a}) = \mathbf{A}(\mathbf{r}) + Ha_1\mathbf{n}_2.$$

Here, $\mathbf{a}$ has components (a_1, a_2, a_3) in the Cartesian system of coordinates (x_1, x_2, x_3). In order to obtain the eigenfunction of $\hat{H}(\mathbf{r})$, we must return to the initial Hamiltonian, which can be carried out through the gauge transformation

$$\mathbf{A} \longrightarrow \mathbf{A} + \nabla f, \quad f = -Ha_1x_2,$$

under which the wave function acquires an additional phase factor:

$$\psi \longrightarrow \psi \exp(ief/\hbar c).$$

Denoting the result of all the operations by $\hat{T}_\mathbf{a}\psi(\mathbf{r})$, we define the magnetic translation operator

$$\hat{T}_\mathbf{a}\psi(\mathbf{r}) = \psi(\mathbf{r} + \mathbf{a})\exp(iha_1x_2), \qquad (1)$$

where $\mathbf{h} = |e|\mathbf{H}/\hbar c$. It is well known [21] that the operator $\hat{T}_\mathbf{a}$ introduced here can be referred to as the translation operator in a magnetic field.

It follows from definition (1) that after two successive translations by $\mathbf{a}$ and $\mathbf{a}'$, we have

$$\hat{T}_{\mathbf{a}}\hat{T}_{\mathbf{a}'} = \hat{T}_{\mathbf{a}+\mathbf{a}'}\omega(\mathbf{a}, \mathbf{a}'),$$
$$\omega(\mathbf{a}, \mathbf{a}') = \exp(-iha_1'a_2), \quad (2)$$

and, hence, the operators $\hat{T}_{\mathbf{a}}$ and $\hat{T}_{\mathbf{a}'}$ generally do not commute,

$$\hat{T}_{\mathbf{a}}\hat{T}_{\mathbf{a}'} = \hat{T}_{\mathbf{a}'}\hat{T}_{\mathbf{a}}\exp(-2iha_1'a_2),$$

and ensure not an ordinary, but a projected representation of the translational group. The basis of these representations are the wave functions of the stationary states of a Bloch electron in a magnetic field. Obviously, this representation is not projected, but ordinary if the condition $\omega(\mathbf{a}, \mathbf{a}') = 1$ is satisfied in (2). Accordingly, the translations for which this condition is satisfied form the group of magnetic translations ensuring the Bloch law of transformation of the wave function. We will distinguish between these translations using the subscript m: $\mathbf{a}_m = (a_{1m}, a_{2m}, a_{3m})$. It follows from relations (2) that translations are magnetic if the following condition is satisfied:

$$ha_{1m}a_{2m}' = 2\pi\frac{Ha_{1m}a_{2m}'}{\Phi_0} = 2\pi p, \quad p = 1, 2, 3, \ldots; \quad (3)$$

i.e., the magnetic flux through the area $a_{1m}a_{2m}'$ is equal to an integral number of flux quanta $\Phi_0 = 2\pi\hbar c/|e|$. In formula (3), a_{1m}, a_{2m}' are the components of the two vectors $\mathbf{a}_m$ and $\mathbf{a}_m'$ carrying out magnetic translations along the axes x_1 and x_2. In a crystal with a unit cell constructed on noncoplanar shortest translations $\mathbf{a}_1$ and $\mathbf{a}_2$ as well as $\mathbf{a}_3 \parallel \mathbf{H}$, condition (3) will be satisfied in a magnetic field satisfying the condition

$$\frac{Ha_{1x_1}a_{2x_2}}{\Phi_0} = \frac{p}{q}, \quad (4)$$

where a_{1x_1} and a_{2x_2} are the components of the basis vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ along x_1 and x_2, respectively, and q is an integer. In this case, in accordance with relations (3) and (4), vectors $\mathbf{a}_m$ and $\mathbf{a}_m'$ must be chosen so that $a_{1m}a_{2m}' = qa_{1x_1}a_{2x_2}$. Consequently, condition (3) defines not the entire magnetic translation vector, but only its components along the axes x_1 and x_2, lying in the plane perpendicular to $\mathbf{H}$. Consequently, a magnetic unit cell in the 3D space of the crystal can be chosen in various ways. For its optimal choice, we can use the principles similar to those employed for deriving the Bravais lattices: if condition (3) is satisfied, the magnetic cell symmetry must correspond to the symmetry of the entire lattice; the number of right angles and equal sides must be maximum, and the volume of the magnetic unit cell must be minimal.

Since we are studying the quantum states of a Bloch electron in an fcc lattice, we define a magnetic unit cell and the magnetic Brillouin zone for this type of lattices. We direct the Cartesian axes of coordinates (x, y, z) along the sides of a cube. In such a geometry, the vectors forming the unit cell have the coordinates ($a/2$, 0, $a/2$), (0, $a/2$, $a/2$), ($a/2$. $a/2$, 0). For the sake of definiteness, we consider two field orientations: $\mathbf{H} \parallel (0, 0, 1)$ and $\mathbf{H} \parallel (1, 1, 0)$.

(a) Let us assume that the magnetic field is directed along the axis $\mathbf{a}_3 = a(0, 0, 1)$ and the vector potential has the form $\mathbf{A} = (0, Hx, 0)$. It follows from relation (4) that for the given field orientation, the fraction p/q is the number of magnetic flux quanta through the area $a^2/4$. It is convenient to choose the vectors defining magnetic translations for even and odd values of q in different ways. Namely, for an even q, the magnetic unit cell is constructed as a right parallelepiped on vectors $\mathbf{a}_{1m} = (qa/2, 0, 0)$, $\mathbf{a}_{2m} = (0, qa/2, 0)$, $\mathbf{a}_{3m} = (0, 0, a)$. Consequently, the magnetic Brillouin zone is also a right parallelepiped with a square base, constructed on the reciprocal lattice vectors $\mathbf{b}_{1m} = (2/qa, 0, 0)$, $\mathbf{b}_{2m} = (0, 2/qa, 0)$, $\mathbf{b}_{3m} = (0, 0, 2/a)$ due to the presence of atoms at the centers of the faces of the fcc lattice. For odd q, a magnetic cell is an oblique prism whose basis vectors have the coordinates ($qa/2$, 0, $a/2$), (0. $qa/2$, $a/2$), ($qa/2$, $qa/2$, 0) in view of the presence of atoms at the centers of the faces of the fcc lattice. It should be noted that on these basis vectors, one can also construct a Wigner–Seitz cell possessing all the symmetry elements of a magnetic lattice. The Brillouin zone for $q = 3$ and for the given orientation of the magnetic field will be constructed in Subsection 3.3.

(b) In the case when $\mathbf{H} \parallel \mathbf{a}_3$, where $\mathbf{a}_3 = (a/2, a/2, 0)$, it is convenient to choose a new Cartesian system of coordinates, in which axis $x_3 \parallel \mathbf{H}$, and the new coordinates (x_1, x_2, x_3) are connected with old ones through the following relations:

$$x_1 = \frac{y - x}{\sqrt{2}}, \quad x_2 = z, \quad x_3 = \frac{y + x}{\sqrt{2}}.$$

Once again, we choose the vector potential in the Landau gauge: $\mathbf{A} = (-Hx_2, 0, 0)$. It follows from relation (4) that in this case the fraction p/q is the number of magnetic flux quanta through a rectangle with sides $a/2\sqrt{2}$ and $a/2$, lying in the plane $(x_1x_2) \perp \mathbf{H}$. In analogy with case (a), it is convenient to construct a magnetic 3D cell for even and odd values of q in different ways. Namely, for an even q, it may be chosen in the form of a right parallelepiped constructed on the vectors having the following components in the (xyz) system: ($a/2$, $a/2$, 0), ($a/2$, $-a/2$, 0), (0, 0, $qa/2$). The reciprocal lattice in this case is a rectangle with the basis vectors $\mathbf{b}_{1m} = (2/a, 2/a, 0)$, $\mathbf{b}_{2m} = (2/a, -2/a, 0)$, $\mathbf{b}_{3m} = (0, 0, 2/qa)$. If the value of q is odd, the magnetic cell is an oblique prism constructed on the basis vectors ($a/2$, $a/2$, 0), ($a/2$, $-a/2$, 0), ($a/2$, 0, $qa/2$).

3. QUANTUM STATES AND FERMI SURFACE FOR THE FIELD ORIENTATION ALONG (0, 0, 1)

3.1. Derivation of the Harper Equation

The wave function, which is an eigenfunction of the magnetic translation operator [17], for an fcc lattice in the strong coupling approximation will be written in the form

$$\psi_{\mathbf{k}}(\mathbf{r}) = \sum_{n,m,l} g_n(\mathbf{k})\exp(i\mathbf{k}\mathbf{a}_{\mathbf{n}}) \times \exp\left(-2\pi i\frac{p}{q}\frac{(y-ma/2)}{a/2}n\right) \times \psi_0\left(x-\frac{na}{2}, y-\frac{ma}{2}, z-\frac{la}{2}\right), \quad (5)$$

where the wave function $\psi_0(\mathbf{r}-\mathbf{a}_{\mathbf{n}})$ describing the *s*-state of an electron in an isolated atom in a magnetic field satisfies the Schrödinger equation with the atomic potential $U(\mathbf{r}-\mathbf{a}_{\mathbf{n}})$. Summation in expression (5) is carried out over all sites of the fcc lattice, integral indices (n, m, l) defining the coordinates of the atoms in the crystal lattice. The introduction of the phase factor

$$\exp\left(-2\pi i\frac{p}{q}\frac{(y-ma/2)n}{a/2}\right)$$

into relation (5) ensures the correct law of transformation of the wave function under translations, which was formulated in Section 2. Coefficients g_n describe the distribution over the sites of a magnetic cell.

Let us now derive the system of equations for coefficients g_n. For this purpose, we substitute function (5) into the Schrödinger equation with the Hamiltonian

$$\hat{H} = \frac{1}{2m}\left(\mathbf{p}-\frac{e}{c}\mathbf{A}\right)^2 + V(\mathbf{r}),$$

where $V(\mathbf{r})$ is the crystal potential. Proceeding in compliance with the method of strong coupling, we calculate the overlap integrals between adjacent sites, thus obtaining the system of difference equations for coefficients $g_n(\mathbf{k})$. The overlap integrals between adjacent sites in the *xy* plane in the presence of a magnetic field has the form

$$A = \int \exp\left(\pm 2\pi i\frac{p}{q}\frac{(y-(m\pm 1/2)a/2\pm(n\pm 1/2)a/2)}{a/2}\right) \times \psi_0(\mathbf{r}-\mathbf{a}_{\mathbf{n}}^{xy})(V(\mathbf{r})-U(\mathbf{r}-\mathbf{a}_{\mathbf{n}}))\psi_0(\mathbf{r}-\mathbf{a}_{\mathbf{n}})d\tau, \quad (6)$$

where $\mathbf{a}_{\mathbf{n}}^{xy} = ((n\pm 1)a/2, (m\pm 1)a/2, la/2)$. Obviously, in integral (6), the narrow region of maximum overlap of the wave functions, which is located at points $x = (n\pm 1/2)a/2$, $y = (m\pm 1/2)a/2$, plays a significant role. Substituting this value of *y* into the phase factor in the integrand, we obtain

$$A = \exp\left(\pm 2\pi i\frac{p}{q}\left(n\pm\frac{1}{2}\right)\right)\alpha,$$

where α is the value of the overlap integral in zero magnetic field. In should also be borne in mind that the integrand in formula (6) contains atomic wave functions in the presence of a magnetic field. Consequently, the constant in the overlap integral must also differ from its value for $\mathbf{H} = 0$. We can expect that the change in the form of the atomic function is significant in the magnetic field for which the cyclotron frequency in a plane perpendicular to $\mathbf{H}$ is much larger than the characteristic atomic frequency. In the magnetic field $H \approx 30$ MG, $\omega_c \approx 5\times 10^{14}$ s^{-1}; i.e., energy $\hbar\omega_c$ is much smaller than the separation between atomic levels. Consequently, it can be expected that the form of function ψ_0 in such fields and, hence, the overlap integrals will not differ significantly from the value obtained above. It can also be verified easily that the Zeeman splitting in magnetic fields $H \approx 10^7$ G is also much smaller than the characteristic atomic energy. This means that the positions of the centers of energy bands will change insignificantly as compared to the case when $\mathbf{H} = 0$.

The overlap integral in the *xz* plane has the form

$$B = \int \exp\left(\pm 2\pi i\frac{p}{q}\frac{(y-ma/2)}{a/2}\right)\psi_0(\mathbf{r}-\mathbf{a}_{\mathbf{n}}^{xz}) \times (V(\mathbf{r})-V(\mathbf{r}-\mathbf{a}_{\mathbf{n}}))\psi_0(\mathbf{r}-\mathbf{a}_{\mathbf{n}})d\tau, \quad (7)$$

where $\mathbf{a}_{\mathbf{n}}^{xz} = ((n\pm 1)a/2, ma/2, (l\pm 1)a/2)$. In analogy with the previous case, we put $y = ma/2$ and find that $B = \alpha$. Finally, the overlap integral between the atoms in the *yz* plane can be presented in the form

$$C = \exp\left(\pm 2\pi i\frac{p}{q}n\right)\alpha.$$

The calculations for the second group of the nearest neighbors located at points $(a, 0, 0)$, $(0, a, 0)$, $(0, 0, a)$, etc. can be carried out in a similar way [17]. The overlap integral between the wave functions for this group of neighbors in zero magnetic field will be denoted by *r*. It is well known that the ratio $r/\alpha \ll 1$ for most metals (Cu, Ag, Au, Al) [22].

The final form of the equation for coefficients g_n in the given orientation of the magnetic field is

$$2\alpha\left[g_{n-1}\exp\left(-i\frac{k_x a}{2}\right) \times\left(\cos\left(\frac{k_y a}{2}+2\pi\frac{p}{q}\left[n-\frac{1}{2}\right]\right)+\cos\frac{k_z a}{2}\right) + g_{n+1}\exp\left(i\frac{k_x a}{2}\right)\right.$$

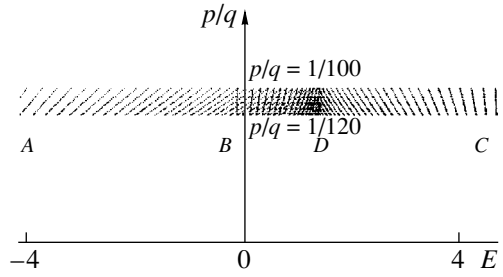

Fig. 1. Magnetic subbands (Landau levels) in an fcc lattice in a weak magnetic field **H** || (0, 0, 1) for $k_z = 3\pi/4a$.

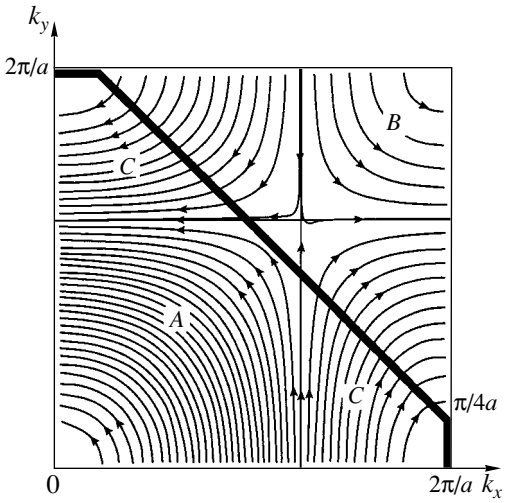

Fig. 2. Cross sections of the family of constant-energy surfaces defined by relation (9) by the plane $k_z = 3\pi/4a$ for $r = 0$. The bold line indicates the boundary of the first Brillouin zone and arrows indicate the direction of rotation in a magnetic field.

$$\times\left(\cos\left(\frac{k_y a}{2} + 2\pi\frac{p}{q}\left[n+\frac{1}{2}\right]\right) + \cos\frac{k_z a}{2}\right) \qquad (8)$$

$$+ 2\cos\frac{k_z a}{2}\cos\left(\frac{k_y a}{2} + 2\pi\frac{p}{q}n\right)g_n\Big]$$

$$+ r\Big[g_{n-2}\exp(-ik_x a) + g_{n+2}\exp(ik_x a)$$

$$+ 2\left(\cos\left(k_y a + 4\pi\frac{p}{q}n\right) + \cos k_z a\right)g_n\Big] = -\varepsilon g_n.$$

It should be noted that in contrast to the standard Harper equation [2], the difference equation (8) displays the modulation of the diagonal as well as nearly diagonal terms. Besides, system (8) is five-diagonal when the interaction with the next group of nearest neighbors ($r \neq 0$) is taken into account. It should also be noted that in contrast to the standard Harper equation, system (8) is not symmetric relative to the substitution $\varepsilon \longrightarrow -\varepsilon$. The quasimomentum components (k_x, k_y, k_z) in the present case vary in the magnetic Brillouin zone described in Section 2 (item (a)).

Note that the wave function (5) with coefficients g_n satisfying system (8) describes an electron density distribution which possesses translational invariance only relative to magnetic translations and which is nonuniform within a magnetic cell. However, the electron density distribution in real crystals can be nonuniform only within a single unit cell of the crystal due to the Coulomb interaction. In order to make the density of states homogeneous within a magnetic cell, we must take a linear combination of functions (5) using the degeneracy of Eq. (8) in the quantum number k_y. Such functions are obviously the regular functions of the zeroth approximation in the Coulomb interaction. The corresponding calculations are completely similar to those made in [17] for a simple cubic lattice.

In zero magnetic field ($p/q = 0$), system (8) leads to the standard dispersion relation of an fcc lattice for **H** = 0 [22, 23]:

$$\varepsilon(\mathbf{k}) = -4\alpha\Big(\cos\frac{k_x a}{2}\cos\frac{k_y a}{2} + \cos\frac{k_x a}{2}\cos\frac{k_z a}{2} + \cos\frac{k_y a}{2}\cos\frac{k_z a}{2}\Big) - 2r(\cos k_x a + \cos k_y a + \cos k_z a). \qquad (9)$$

Let us first consider the energy spectrum defined by system (8) for $r = 0$ in the range of weak magnetic fields ($p/q \ll 1$) and compare the result with the Onsager–Lifshits semiclassical approximation [18]. For this purpose, we determine the spectrum of system (8) for a fixed k_z parallel to **H**. Figure 1 shows a fragment of the spectrum defined by Eq.(8) for $k_z = 3\pi/4a$. It should be noted at the very outset that in a weak magnetic field, the energy is virtually independent of k_x and k_y; i.e., the spectrum consists of narrow bands. In other words, the Landau levels are practically not broadened in this case. At the same time, Fig. 2 shows the sections of the family of constant-energy surfaces defined by relation (9) by the same plane $k_z = 3\pi/4a$. It is these sections that give the semiclassical trajectories and discrete Onsager–Lifshits spectrum for $k_z = 3\pi/4$. Figure 1 shows several sequences of energy levels (narrow magnetic subbands) which may correspond to different semiclassical trajectories. The energy spectrum in regions *A*, *B*, and *C* in Fig. 1 corresponds to trajectories *A*, *B*, and *C* in Fig. 2. The trajectories in Fig. 2 are of the electron type in regions *A* and *B* and of the hole type in region *C*. In Fig. 1, this corresponds to different slopes of the levels (as functions of

the magnetic flux) in regions *A*, *B*, and *C*. The region of condensation of levels in Fig.1 (region *D*) corresponds to the separatrix in Fig. 2. It can be verified that the separations between the levels in Fig. 1 completely correspond to the values obtained in the case of the Onsager–Lifshits quantization.

3.2. Dispersion Relation, Magnetic Brillouin Zone, and the Fermi Surface for p/q = 1/2

For $p/q = 1/2$, the spectrum can be determined from the condition of vanishing of the second-order determinant of system (8). In this case, the energy in the upper and lower subbands is defined as

$$\varepsilon_{1,2} = \pm 4\alpha\left[\sin^2\frac{k_x a}{2}\sin^2\frac{k_y a}{2} + \cos^2\frac{k_z a}{2}\left(\cos^2\frac{k_x a}{2} + \cos^2\frac{k_y a}{2}\right)\right]^{1/2} - 2r(\cos k_x a + \cos k_y a + \cos k_z a). \quad (10)$$

Consequently, $\varepsilon_1 \in [-2r, 4\alpha\sqrt{2} + 6r]$ and $\varepsilon_2 \in [-4\alpha\sqrt{2} - 6r, 2r]$; i.e., in region $[-2r, 2r]$, band overlapping takes place due to the inclusion of the second group of nearest neighbors in the lattice. For $r = 0$, the bands just touch each other at the point $\varepsilon = 0$. Thus, the second term in the dispersion relations (9) and (10) becomes significant in the presence of root-type singularities in the region $\varepsilon = 0$. For the given magnetic field orientation, a similar situation takes place for all even values of the denominator of the fraction p/q. It follows from the above analysis (see Section 2) that for $p/q = 1/2$ and for the given orientation of the magnetic field, the Brillouin zone is a cube constructed on vectors $(2\pi/a, 0, 0)$, $(0, 2\pi/a, 0)$, $(0, 0, 2\pi/a)$. Function (10) is periodic in this region, as expected. It should be noted that the volume of the magnetic Brillouin zone is equal to $1/q^2 = 1/4$ of the volume of the initial Brillouin zone for $\mathbf{H} = 0$, all the states in the magnetic band being doubly degenerate in k_y.

The Fermi level and the Fermi surface in a magnetic field can be determined from the condition that electrons in a monovalent metal (taking into account the spin) occupy a volume equal to half the volume of the initial Brillouin zone. This can be done by calculating, using relation (10), the volume occupied in the **k** space by electrons whose energy does not exceed the given value. The results of such a calculation for $r/\alpha = 1/8$, $\mathbf{H} \parallel (0, 0, 1)$, and $p/q = 1/2$ are presented in Fig. 3. These results show that the Fermi level lies in the region of band overlapping. The Fermi surface consists of two types of cavities: the electron type (cavities *1*) in the upper magnetic subband and the hole type (cavities *2*) in the lower subband. Since the region of subband overlapping is small (in view of the condition $r \ll \alpha$), the volumes of the cavities occupied by electrons and holes constitute a small part of the Brillouin zone. Thus, for $p/q = 1/2$, a monovalent metal with an fcc lattice is transformed into a semimetal or an insulator (for $r = 0$). It follows from relation (10) and Fig. 3 that at point *t* with the coordinates $k_x = \pi/2a$, $k_y = 0$, $k_z = -\pi/a$ and at equivalent points, self-intersection of electron and hole cavities takes place. In this case, a magnetic breakdown can obviously take place.

3.3. Fermi Surface for p/q = 1/3

If the magnetic flux p/q through area $a^2/4$ is equal to 1/3, the spectrum is determined from the condition of vanishing for the third-order determinant of system (8). In this case, we can put $r = 0$ in Eq. (8) since the inclusion of the second term in relation (9) is required only in the regions of small overlapping of subbands. At the same time, as will be proved later, the Fermi level in the case under investigation lies far from the band edge. Three roots of the cubic equation correspond to three magnetic subbands defined by

$$\varepsilon_1 = 2s\cos\left(\frac{1}{3}\arccos\frac{t}{s^3}\right), \quad (11)$$

$$\varepsilon_2 = -2s\cos\left(\frac{\pi}{3} + \frac{1}{3}\arccos\frac{t}{s^3}\right), \quad (12)$$

$$\varepsilon_3 = -2s\cos\left(\frac{\pi}{3} - \frac{1}{3}\arccos\frac{t}{s^3}\right), \quad (13)$$

where the energy is measured in the units of α, $s = \sqrt{6 + 4\cos k_z a}$, and

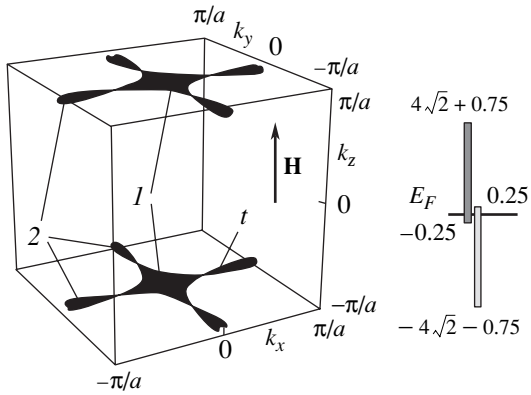

Fig. 3. Electron (*1*) and hole (*2*) Fermi surfaces for a monovalent metal for $p/q = 1/2$ and $\mathbf{H} \parallel (0, 0, 1)$. Self-intersection takes place at point *t* and at equivalent points. The diagram of overlapping magnetic subbands and the position of the Fermi level $E_F = 0$ are shown on the right. The energy is measured in the units of α.

$$t = -\cos\frac{3(k_y - k_z)a}{2} - \cos\frac{3(k_y + k_z)a}{2}$$

$$-12(1 + \cos k_z a) - 2\cos\frac{3k_x a}{2}\left(\cos\frac{3k_z a}{2} - \cos\frac{3k_y a}{2}\right).$$

The mutual arrangement of the magnetic subbands is shown on the right side in Fig. 4.

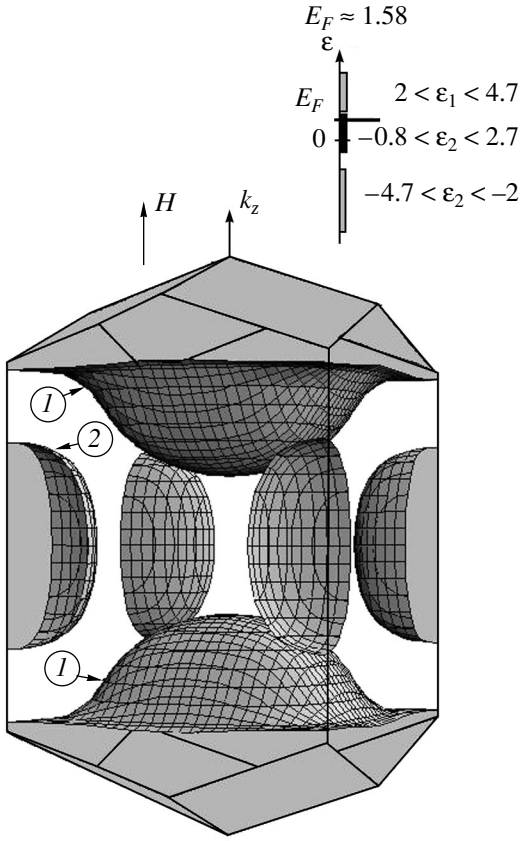

Fig. 4. Two types of the electron Fermi surfaces for a monovalent metal for $p/q = 1/3$ and $\mathbf{H} \parallel (0, 0, 1)$: "hills" and valleys" (*1*) and ellipsoidal-type surfaces (*2*). The diagram of magnetic subbands and the position of the Fermi level are shown at the top.

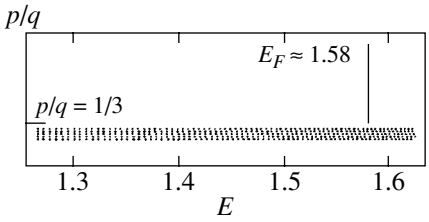

Fig. 5. A fragment of the spectrum in an fcc lattice in a magnetic field $\mathbf{H} \parallel (0, 0, 1)$ in the range of p/q close to 1.3 for $k_z = 0$ and $k_x = k_y = 0$. The position of the Fermi level is indicated.

In accordance with what was said in Section 2, the magnetic Brillouin band is a truncated parallelepiped constructed on the vectors $\frac{2\pi}{3a}(1, 1, 3)$, $\frac{2\pi}{3a}(1, -1, -3)$, $\frac{2\pi}{3a}(-1, -1, 3)$ and enclosed in the parallelepiped $-2\pi/3a \le k_x \le 2\pi/3a$, $-2\pi/3a \le k_y \le 2\pi/3a$, $-2\pi/a \le k_z \le 2\pi/a$ (see Fig. 4). Its volume constitutes $1/q^2 = 1/9$ part of the volume of the first Brillouin zone for $\mathbf{H} = 0$, and all the states are triply degenerate in k_y. The functions defined by relations (11)–(13) are periodic in this Brillouin zone.

The Fermi level can also be determined in the same way as for $p/q = 1/2$. It lies in the middle band defined by Eq. (12), for $E_F \approx 1.58$; the corresponding Fermi surface is depicted in Fig. 4. It should be noted that it consists of two cavities of the electron type: cavities *1*, which can be called "hills" (below) and "valleys" (above), and ellipsoidal cavities *2*. Electrons fill the regions located between the upper and lower boundaries of the Brillouin zone and cavities *1* as well as between the lateral faces of the Brillouin zone and the cavities of type *2*. In contrast to even values of the denominator of fraction p/q, a considerable part of the Brillouin zone is occupied in this case: the regions depicted in Fig. 4 contain 1/3 of the total number of electrons. Thus, a monovalent metal with an fcc lattice for $p/q = 1/3$ remains a typical metal.

Let us also consider the form of the spectrum for magnetic flux values close to 1/3:

$$\frac{p}{q} = \frac{1}{3} + \frac{1}{q'}, \quad q' \gg 1. \tag{14}$$

It was proved above that in weak fields ($p/q \ll 1$), the quantization law determined by system (8) is identical to semiclassical Onsager–Lifshits quantization rules. In the range of p/q defined by relation (14), both methods of calculation also lead to the same result. For example, Fig. 5 shows a fragment of the spectrum for $k_z = 0$ for the band containing the Fermi level. It should be noted that for $q' \gg 1$, magnetic subbands are very narrow and appear as a system of nearly discrete levels. It can be seen from Fig. 5 that the extreme cross section corresponds to a sequence of equidistant levels in the region of E_F. The separation between these levels is in good agreement with the value that can be obtained by applying the Onsager–Lifshits quantization rules to spectrum (12) for $k_z = 0$.

4. QUANTUM STATES AND FERMI SURFACE FOR THE FIELD ORIENTATION ALONG (1, 1, 0)

Let us now consider the case when the field $\mathbf{H} \parallel \mathbf{a}_3$, where $\boldsymbol{a}_3 = (a/2, a/2, 0)$. In analogy with Subsection 3.1, we can write the wave function in the strong-coupling approximation in form (5) with a new phase factor:

$$\psi_{\mathbf{k}}(\mathbf{r}) = \sum_{n,m,l} g_n(\mathbf{k})\exp(i\mathbf{k}\mathbf{a}_{\mathbf{n}}) \times \exp\left(2\pi i\frac{p}{q}\frac{(x_1 - ma/2\sqrt{2})}{a/2\sqrt{2}}n\right) \times \psi_0\left(x_1 - \frac{ma}{2\sqrt{2}}, x_2 - \frac{na}{2}, x_3 - \frac{la}{2\sqrt{2}}\right), \quad (15)$$

where the integral indices (n, m, l) define the coordinates of atoms in the crystal lattice. The evaluation of the overlap integrals in the planes x_1x_3, x_1x_2, and x_2x_3 leads to the following system of difference equations for coefficients g_n:

$$4\alpha\left[\cos\frac{k_3a}{2\sqrt{2}}\left(g_{n-1}\exp\left(-\frac{ik_2a}{2}\right)\times\cos\left(\frac{k_1a}{2\sqrt{2}} + 2\pi\frac{p}{q}\left[n - \frac{1}{2}\right]\right)\right.\right.$$
$$\left. + g_{n+1}\exp\left(\frac{ik_2a}{2}\right)\cos\left(\frac{k_1a}{2\sqrt{2}} + 2\pi\frac{p}{q}\left[n + \frac{1}{2}\right]\right)\right)$$
$$\left. + \frac{1}{2}\left(\cos\frac{k_3a}{\sqrt{2}} + \cos\left(\frac{k_1a}{\sqrt{2}} + 4\pi\frac{p}{q}n\right)\right)g_n\right]$$
$$+ r\left[g_{n-2}\exp(-ik_2a) + g_{n+2}\exp(ik_2a)\right.$$
$$\left. + 4\cos\frac{k_3a}{\sqrt{2}}\cos\left(\frac{k_1a}{\sqrt{2}} + 4\pi\frac{p}{q}n\right)g_n\right] = -\varepsilon g_n. \quad (16)$$

The form of this system is the same as for system (8). The quasimomentum components (k_1, k_2, k_3) run through all values in the magnetic Brillouin zone defined in Section 2 (item (b)). It should be noted that in the given orientation, as well as for $\mathbf{H} \parallel (0, 0, 1)$, system (16) remains a five-diagonal system, while for $r = 0$ it is a three-diagonal system, which is ensured by the optimal choice of the vector potential gauge. The following interesting features of the solutions of system (16) are worth noting. It can easily be seen that for even values of q, the solutions are doubly degenerate: the energy ε_n corresponds to functions $\{g_n\}$ and $\{(-1)^n g_{n+q/2}\}$. Besides, as in the standard Harper equation, system (16) for $k_1 = k_2 = 0$ is invariant to the substitution $n \longrightarrow q - n$. This symmetry property for the wave functions is reflected in Figs. 7b and 8b presenting the eigenfunctions of system (16). The symmetry $k_3 \longrightarrow -k_3$ also takes place as in zero magnetic field.

For $p/q = 1/2$, system (16) leads to the spectrum in two magnetic subbands:

$$\varepsilon_{1,2} = \pm 8\alpha\left|\sin\frac{k_1a}{2\sqrt{2}}\sin\frac{k_2a}{2}\cos\frac{k_3a}{2\sqrt{2}}\right| - 2\alpha\left(\cos\frac{k_1a}{\sqrt{2}} + \cos\frac{k_3a}{\sqrt{2}}\right). \quad (17)$$

Here we put $r = 0$ since the Fermi level (see below) lies far from the edges of the bands in the region of their strong overlapping. The energy defined by relation (17) lies in the intervals $\varepsilon_1 \in [-4\alpha, (4\sqrt{2} + 2)\alpha]$ and $\varepsilon_2 \in [-(4\sqrt{2} + 2)\alpha, 4\alpha]$. In other words, we

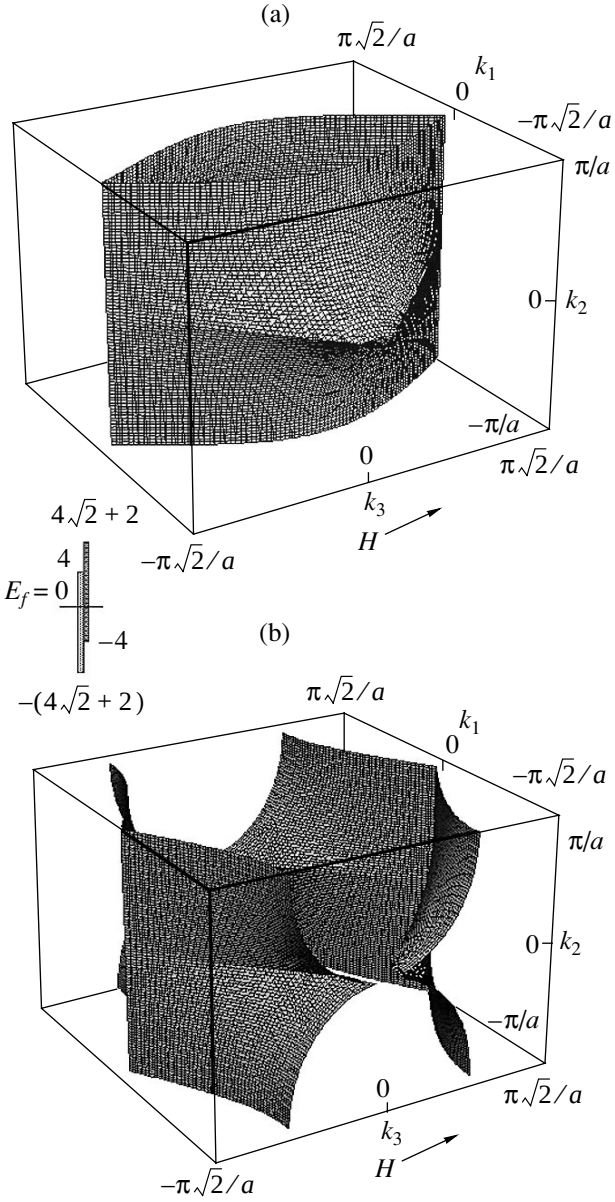

Fig. 6. Electron (a) and hole (b) Fermi surfaces corresponding to the upper and lower magnetic subbands (17) for a monovalent metal for $\mathbf{H} \parallel (1, 1, 0)$ and $p/q = 1/2$. The diagram of magnetic subbands and the position of the Fermi level are shown on the left side of (a).

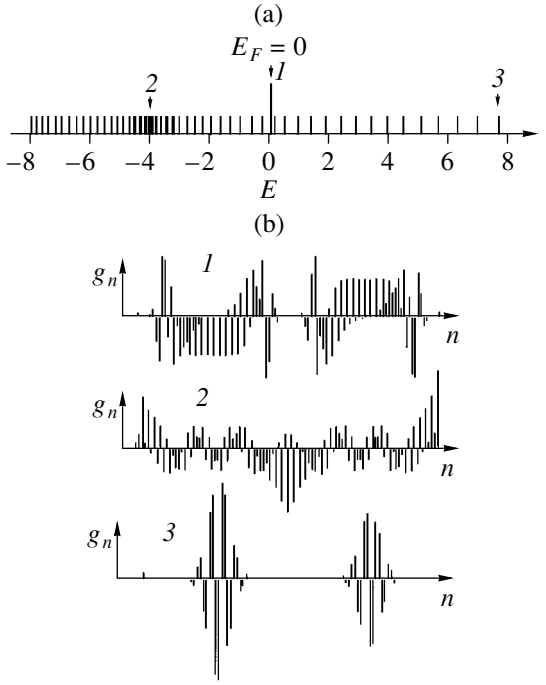

Fig. 7. The spectrum (a) and the wave functions (b) for **H** || (1, 1, 0), $p/q = 51/100$, and cross section $k_3 = 0$. In (a), the following notation is used: Fermi level (*1*), levels in the region of open trajectories near the bottom of the upper magnetic subband (17) (*2*), and levels in the region of closed trajectories near the top of the upper magnetic subband (17) (*3*). In (b), the eigenfunctions corresponding to the energy levels singled out in (a) for the values $k_1 = k_2 = 0$.

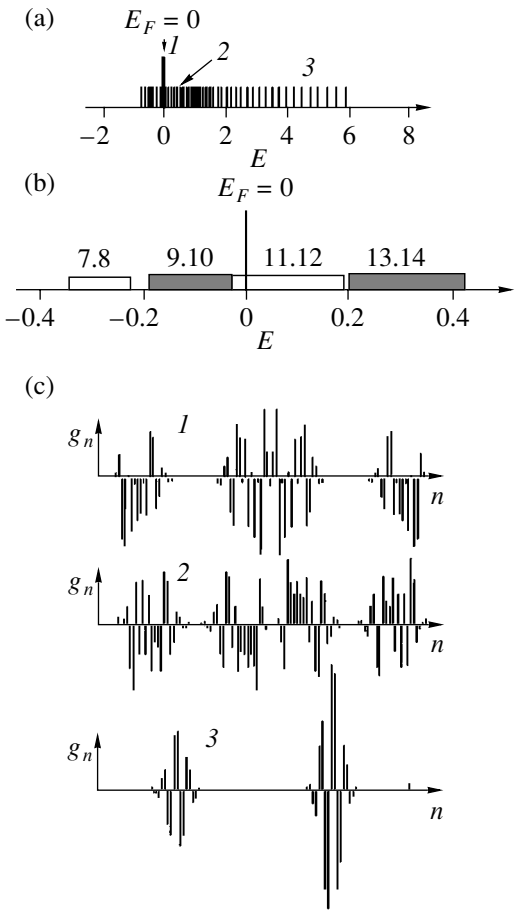

Fig. 8. The spectrum (a) and the wave functions (c) for **H** || (1, 1, 0), $p/q = 51/100$, and cross section $k_3 = 4\pi\sqrt{2}/5a$. In (a), the following notation is used: Fermi level (*1*), levels in the region of open trajectories (*2*), and levels in the region of closed trajectories from the upper magnetic subband (17) (*3*). Magnetic subbands in the region of open orbits corresponding to region *1* are shown in (b) (with their numbers) on a magnified scale for $p/q = 21/40 = 1/2 + 1/40$. In (c), the eigenfunctions corresponding to the energy levels singled out in (a) are shown for $k_1 = k_2 = 0$.

observe a strong overlapping of magnetic subbands even when only the first group of nearest neighbors is taken into account (see the diagram of the magnetic subbands on the right side in Fig. 6a). We can state that in contrast of the 2D case, in which magnetic subbands either touch one another or are separated by forbidden gaps, the bands in the 3D problem with the energy dependence on the three quasimomentum components overlap as a rule. For this reason, the contact of energy bands for even values of q, which was considered in Subsection 3.2, is rather an exception from the general rule.

Figure 6 shows the Fermi surfaces for a monovalent metal for $p/q = 1/2$: the electron-type Fermi surface in the upper magnetic subband is presented in Fig. 6a and the hole-type Fermi surfaces in the lower subband are presented in Fig. 6b. Since the overlap region for two magnetic subbands for the given orientation of **H** is not small, the Fermi surfaces in Fig. 6 occupy a considerable part of the Brillouin zone. It follows hence that a monovalent metal with an fcc lattice for the given field orientation and for $p/q = 1/2$ remains a typical metal in contrast to the case when **H** || (0, 0, 1). Relations (17) and the results presented in Fig. 6 show that the electron and hole cavities touch one another in the planes where the expression in the modulus in (17) vanishes.

Let us now consider the spectrum of system (16) for magnetic flux values differing insignificantly from 1/2: $p/q = 1/2 + 1/q'$, where $q' \gg 1$. In this case, we must pay attention to the structure of magnetic subbands in the energy range corresponding to open orbits in the semiclassical approximation. It is convenient to construct the energy spectrum for a fixed value of k_3. Such a spectrum is depicted in Fig. 7a for $p/q = 51/100$ and for $k_1 = k_2 = 0$ (section $k_3 = 0$) and in Fig. 8a (section $k_3 = 4\pi\sqrt{2}/5a$). Figures 7b and 8c show several eigenfunctions for the levels representing a certain spectral region: functions *1* correspond to the Fermi level, *2* to open trajectories and *3* to close orbits. A comparison of Figs. 7 and 8 shows that as the section $k_3 = \text{const}$ approaches the boundary of the Brillouin zone, the fraction of energy values in the continuous spectrum increases in accordance with the dispersion relation (17) and Fig. 6, where $p/q = 1/2$.

In the quantum-mechanical problem, the structures of magnetic subbands in the region of open and closed orbits differ significantly. This can be clearly seen in Fig. 8a: closed orbits correspond to exponentially narrow energy bands (Landau levels in the semiclassical approximation), whose separations are determined by the cyclotron frequency, while open orbits correspond to relatively wide bands separated by narrow forbidden gaps. We established that at the boundary between the energies corresponding to open and closed orbits, the width of the magnetic subband increases sharply, while the separation between subbands decreases. Such a behavior can be explained as follows: open orbits correspond to wave functions which are strongly delocalized in the region of a magnetic cell, while closed orbits correspond to localized functions (see Figs. 7b and 8c). The width of a magnetic subband is obviously determined by the tunneling probability (overlap integral) between neighboring cells. For this reason, the width of a magnetic subband for delocalized states is larger than for localized states. The structure of magnetic subbands in the region of the Fermi level is depicted in Fig. 8b. The Fermi level lies in the region of the eleventh and twelfth bands, which are in contact with the ninth and tenth bands. It should be noted that magnetic subbands are arranged in pairs in view of the double degeneracy of the solutions of system (16) described above for even q.

5. TOTAL ENERGY OF ELECTRONS AND MAGNETIC SUSCEPTIBILITY

The dispersion relation and Fermi surfaces determined above for an fcc lattice in an ultrastrong magnetic field make it possible to calculate the total energy of conduction electrons and to predict qualitatively the type of magnetic susceptibility oscillations of the electron gas. The following two typical features of oscillatory effects in ultrastrong magnetic fields are worth noting.

First, a change in the magnetic field in the interval of p/q from 0 to 1 must give rise to several series of typical oscillations of the magnetic susceptibility. These series lie in the vicinity of prime rational values of p/q with small values of q. In particular, in the vicinity of $p/q = 1/3$, the oscillations are associated with the passage of narrow magnetic subbands (see Fig. 5) through the Fermi level, the period of oscillations being determined by the area of the extreme cross section $k_z = 0$ of the type (2) surface in Fig. 4. It was mentioned above that for the magnetic field value corresponding to $p/q = 1/2$, the Fermi surface consists of the small electron and hole cavities depicted in Fig. 3. The extreme cross sections are the cross sections $k_z = \pm\pi/a$ whose area is much smaller than, for example, the area of the "paunch" (the cross section of the Fermi surface of a monovalent metal by a plane passing through the origin). Since the distance between two cavities on the Fermi surface is small, we can expect a considerable effect of the magnetic breakdown on the form and frequency of the oscillations. A similar situation in quasi-two-dimensional organic conductors was discussed in [9, 10].

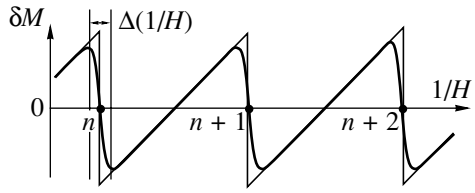

Fig. 9. Oscillations of magnetization in a layer δk_z without taking into account (saw-tooth curve) and taking into account (smooth curve) the finite width of magnetic subbands.

Second, in each series of the oscillations, level broadening effect takes place, leading to a decrease in the oscillation amplitude and to a relative decrease in the contribution from high harmonics. In order to estimate this effect quantitatively, we make use of the fact that in a magnetic field $p/q = 1/q'$, where $q' \gg 1$ and the magnetic subband width is $\Delta E_n \ll \hbar\omega_c$. Here, $\omega_c = eH/mc$ and $m = (1/2\pi)\partial S/\partial\varepsilon$, where $S(\varepsilon, k_z)$ is the area bounded by a semiclassical trajectory in the cross section $k_z = \text{const}$. We use a simple model of the spectrum in the nth 3D magnetic subband: $e_n(k_x, k_y, k_z) = \varepsilon_{\perp n}(k_x, k_y) + \hbar^2 k_z^2/2m$, where $\varepsilon_{\perp n}(k_x, k_y)$ is the energy spectrum in the nth 2D magnetic subband, and assume that $T = 0$. We also assume that the center of the magnetic subband $\varepsilon_{\perp n}(k_x, k_y)$ in this field region coincides with the position of the nth Landau level $\hbar\omega_c(n + 1/2)$. The number of magnetic subbands may change rapidly upon a change in the magnetic field, but all of them are grouped near the Landau levels. In this case, we can disregard the fine structure of these subbands and speak of the broadening of the nth level, which remains practically constant and equal to ΔE_n. Such a pattern of the spectrum in weak fields is typical of any model of the spectrum of 2D Bloch electrons in a weak magnetic field.

It is well known [24] that the magnetic moment of an electron system with equidistant Landau levels in a layer of thickness δk_z near a certain k_z is given by

$$\delta M = -\frac{\varepsilon_F'}{H}(n - n_0),$$

where $\varepsilon_F' = \varepsilon_F - \hbar^2 k_z^2/2m$ and $(n - n_0)$ is the difference in the populations of the layer δk_z in the given field H and in zero field. In the absence of Landau level broadening, δM is a periodic function of $1/H$ with period $e\hbar/mc\varepsilon_F'$ depending on k_z. For values of H determined by the condition $\hbar\omega_c(n + 1/2) = \varepsilon_F'$, when the upper Landau level passes through ε_F', the electron concentration in the layer δk_z changes jumpwise by $(eH/4\pi^2 c\hbar)\delta k_z$, while the magnetization δM experiences s discontinuity with the jump $(e\varepsilon_F'/4\pi^2 c\hbar)\delta k_z$ (the saw-tooth curve in Fig. 9).

In the case when the system of discrete Landau levels is transformed into a system of narrow magnetic subbands, the form of the dependence of δM on $1/H$ changes. This is due to the fact that with increasing field, the next magnetic subband is vacated from electrons not instantaneously, but over the interval ΔH, where $\Delta H/H = \Delta E_n/\hbar\omega_c$, in which the given subband passes through ε_F'. The behavior of the curve $\delta M(1/H)$ is determined in this case by the density of states in the subbands grouped near the Landau level. The qualitative behavior of magnetization is illustrated in Fig. 9. The transition region corresponds to the passage of the subband with number n, $n + 1$, $n + 2$, … through ε_F'. Obviously, the smoothness of the function $\delta M(1/H)$ leads to a qualitative change in the form of the Fourier spectrum for the oscillations, namely, to the suppression of higher harmonics. In experiments, oscillations with periods determined by the condition $\partial\varepsilon_n/\partial k_z = 0$ must be observed, which corresponds to the value $k_z = 0$ in the model under investigation.

Concluding this section, we will calculate the total energy of the electron gas. This will enable us to determine the constant component of the magnetic moment. The calculations will be made for zero temperature since $kT \ll \hbar\omega_c$ in a strong magnetic field. We consider the case when the field $\mathbf{H} \parallel (0, 0, 1)$ and disregard the second term in relation (9), which makes small corrections to the total energy of the electron gas. The values of the total energy in dimensionless units are given below:

p/q:	0	1/4	1/3	1/2	2/3	3/4	1
E:	–163	–170	–173	–200	–173	–170	–163

These data indicate that the total energy of electrons in magnetic fields p/q and $1 - p/q$ is the same; in the p/q interval from 0 to 1/2, the energy decreases with increasing field. The fact that the total energy decreases in a strong field is associated with a significant rearrangement of the spectrum (cf. (9) and (10)). For $p/q = 1/2$, when $E_F \approx 0$, the maximum of the density of states corresponds to $\varepsilon < 0$ (in this case, $E_F \approx 0$), while in zero field, the density of states for the dispersion relation (9) attains its maximum value for $\varepsilon > 0$, and $E_F \approx 1$. This leads to a decrease in the total energy on the interval of p/q from 0 to 1/2. The minimum value is attained for $p/q = 1/2$, i.e., for the number of magnetic flux quanta piercing a unit cell equal to the occupation ν of the Brillouin zone by electrons. Since $p/q = \nu = 1/2$ and the density of states at the Fermi level is small and varies smoothly with energy, we can expect that the curve $E(H)$ exhibits a smooth behavior near its minimum. This distinguishes the 3D problem from the 2D problem [25], in which the derivative dE/dH suffers a discontinuity in the region of the minimum. A similar result was obtained in [15, 16], where the problem on a 3D Bloch electron in a simple cubic lattice placed in a magnetic field with various orientations was considered. It should be noted that in the problem of a 2D Bloch electron in a magnetic field, a local energy minimum also takes place for $\nu = p/q$ [25], when each electron corresponds to a magnetic flux quantum. In the 3D problem, such an interpretation is ruled out. The decrease in the total energy and, in particular, the existence of a minimum in a 3D crystal in a strong magnetic field indicates that monovalent metals with an fcc lattice (Cu, Ag, Au), which are diamagnetic in weak fields, become paramagnetic in an ultrastrong field. Indeed, for absolute zero temperature, the constant component of the magnetic moment on the p/q interval from 0 to 1/2 is positive:

$$M = -\frac{\Delta E}{\Delta H} > 0;$$

i.e., the electron gas is paramagnetic. The absolute value of susceptibility defined as $\kappa = M/H$ is comparable in magnitude with that for a weak magnetic field: $\kappa \approx 10^{-6}$–10^{-5}. It should be noted that the inclusion of the spin in the model under investigation near $p/q = 1/2$ is immaterial since the density of states in the vicinity of the Fermi level (and, hence, the Pauli paramagnetism) has the minimum value.

6. CONCLUSIONS

In the present work, the strong-coupling approximation is used to construct the explicit expression for the one-electron wave function which is an eigenfunction of the magnetic translation operator in a 3D crystal in a constant magnetic field. The Harper-type 3D equation determining the amplitudes of the wave function at the sites of a magnetic cell is derived and the energy spectrum in magnetic subbands is determined. The transformation of the Fermi surface upon the application of a strong magnetic field with the numbers of magnetic quanta $p/q = 1/2$ and 1/3 is studied for a monovalent metal with an fcc lattice. The magnetic field was oriented along the crystallographic directions (0, 0, 1) and (1, 1, 0). A strong effect of the magnitude and orientation of the magnetic field on the topology of the Fermi surface is discovered. It is shown that a metal–semiconductor-type transition occurs in the (0, 0, 1) direction for even values of q, while for odd q, the upper magnetic subband is filled approximately by half and the metallic properties of the initial material are preserved.

The form of the de Haas–Van Alphen oscillations is considered. It is shown that in magnetic fields corresponding to the magnetic flux value $p/q + 1/q'$, where p and q are prime integers and $q' \gg 1$, regular series of the de Haas–Van Alphen oscillations similar to magnetic oscillations in weak fields for $p/q \ll 1$ must be observed. The effect of the Landau level broadening on the magnetic oscillation spectrum is considered. The dependence of the total energy on the magnitude of a magnetic field directed along (0, 0, 1) is analyzed and a

minimum is observed for $p/q = 1/2$. This indicates that the electron gas in a crystal placed in a magnetic field may exhibit not only diamagnetic, but also paramagnetic properties.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (grant no. 01-02-17102), by the Ministry of Education of the Russian Federation (grant no. E00-3.1-413), and by CRDF Foundation and the Ministry of Education of the Russian Federation (project no. REC-001).

REFERENCES

1. B. A. Boyko, A. I. Bykov, M. I. Dolotenko, *et al.*, in *Book of Abstracts of the VIII International Conference on Megagauss Magnetic Field Generation and Related Topics*, Tallahassee, USA, p. 149.
2. P. G. Harper, Proc. Phys. Soc. London, Sect. A **68**, 874 (1955).
3. G. E. Zil'berman, Zh. Éksp. Teor. Fiz. **32**, 296 (1957) [Sov. Phys. JETP **5**, 208 (1957)]; Zh. Éksp. Teor. Fiz. **34**, 515 (1958) [Sov. Phys. JETP **7**, 355 (1958)].
4. M. Ya. Azbel', Zh. Éksp. Teor. Fiz. **46**, 929 (1964) [Sov. Phys. JETP **19**, 634 (1964)].
5. D. R. Hofstadter, Phys. Rev. B **14**, 2239 (1976).
6. V. Ya. Demikhovskiĭ and A. A. Perov, Zh. Éksp. Teor. Fiz. **114**, 1795 (1998) [JETP **87**, 973 (1998)].
7. V. Ya. Demikhovskii and A. A. Perov, Phys. Low-Dimens. Struct. **7/8**, 135 (1998).
8. H. Silberbauer, J. Phys.: Condens. Matter **4**, 7355 (1992).
9. P. S. Sandhu, Ju. H. Kim, and J. S. Brooks, Phys. Rev. B **56**, 11566 (1997).
10. S. Y. Han, J. S. Brooks, and Ju. H. Kim, Phys. Rev. Lett. **85**, 1500 (2000).
11. D. Weiss, M. L. Roukes, A. Mensching, *et al.*, Phys. Rev. Lett. **66**, 2790 (1991).
12. T. Schlösser, K. Ensslin, J. P. Kotthaus, *et al.*, Semicond. Sci. Technol. **11**, 1582 (1996).
13. D. Peter, D. Mayou, and M. Cyrot, Phys. Rev. Lett. **65**, 386 (1990).
14. G. Montambaux and M. Kohmoto, Phys. Rev. B **41**, 11417 (1990).
15. H. Hasegawa, J. Phys. Soc. Jpn. **59**, 4384 (1990).
16. Z. Kunszt and A. Zee, Phys. Rev. B **44**, 6842 (1991).
17. V. Ya. Demikhovskii, A. A. Perov, and D. V. Khomitsky, Phys. Lett. A **267**, 408 (2000).
18. I. M. Lifshits, M. Ya. Azbel', and M. I. Kaganov, *Electron Theory of Metals* (Nauka, Moscow, 1971; Consultants Bureau, New York, 1973).
19. R. E. Peierls, Z. Phys. **80**, 763 (1933).
20. J. Zak, Phys. Rev. A **134**, 1602 (1964); **134**, 1607 (1964).
21. E. M. Lifshitz and L. P. Pitaevskiĭ, *Course of Theoretical Physics*, Vol. 5: *Statistical Physics* (Nauka, Moscow, 1978; Pergamon, New York, 1980), Part 2.
22. A. P. Cracknell and K. C. Wong, *Fermi Surface: Its Concept, Determination and Use in Physics of Metals* (Clarendon, Oxford, 1973; Atomizdat, Moscow, 1978).
23. D. Shoenberg, *Magnetic Oscillations in Metals* (Cambridge Univ. Press, Cambridge, 1984; Mir, Moscow, 1986).
24. C. Kittel, *Quantum Theory of Solids* (Wiley, New York, 1963; Nauka, Moscow, 1967).
25. Y. Hasegawa *et al.*, Phys. Rev. B **41**, 9174 (1990).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 93, No. 1, 2001, pp. 180–187.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 1, 2001, pp. 205–213.

SOLIDS
Electronic Properties

Pressure-Induced Variation of the Magnetic Structure of the Surface of $La_{0.6}Sr_{0.4}MnO_3$ Granules

V. Yu. Tarenkov, A. I. D'yachenko, and V. N. Krivoruchko*
Donetsk Physicotechnical Institute, National Academy of Sciences of Ukraine, Donetsk, 83114 Ukraine
**e-mail: krivoruc@host.dipt.donetsk.ua*
Received February 7, 2001

Abstract—The transport properties of textured films and tunneling junctions of $La_{0.6}Sr_{0.4}MnO_3$, defined by the surface state of the granule, are studied in low magnetic fields (below 100 Oe) and at pressures of up to 10 kbar. Tunneling junctions of two types are investigated, namely, mechanical break junctions and $La_{0.6}Sr_{0.4}MnO_3$–insulator–superconductor junctions. Although only one electrode represents the magnetic material in the latter case, all samples exhibit a low-field magnetoresistive effect. Hydrostatic compression suppresses the magnetoresistive effect to considerably change the transport properties of ceramic and tunnel samples. The reasons for such behavior are discussed in connection with the model of spin-polarized inelastic tunneling of charge carriers through a potential barrier formed both by the intergranular region and by the surface of contacting granules. Reasons are given for the fact that it is most probable that the magnetic state of the barrier and its height vary under the effect of pressure because of the transition of the surface of granules to the metallic state.

1. INTRODUCTION

It has now been found experimentally that mono- and polycrystalline samples of lanthanum manganites exhibit fairly diverse magnetoresistive properties. While the resistance of monocrystalline samples varies little in low magnetic fields, polycrystalline and ceramic systems are characterized by a marked decrease in resistance even in a low field (see, for example, [1–4]). In early studies, this phenomenon was interpreted within the phenomenological model [5], in which the insulating barrier transparency to charge carriers depended on the relative orientation of electrode magnetization alone. Later, a number of forcible experimental and theoretical arguments appeared, which pointed to more complicated physics of the tunneling effect in junctions formed by magnetics, in particular, in tunneling systems of lanthanum manganites [6–15]. Namely, the pattern of current flow in manganite materials is defined both by the magnetic state of the granule proper and by the structure of the intergranular barrier. This is evidenced, for example, by marked changes of the transport characteristics as functions of temperature [6, 8, 11, 12, 14], by the dependence of the tunneling junction properties on the barrier material for one and the same material of the electrodes [10, 11], by the presence of magnetoresistive properties even when only one of the electrodes of the junction is magnetic [10, 11, 15, 16], and so on. These results indicate that the intergranular boundaries both serve as the potential barrier and define the pattern of tunneling of spin-polarized electrons from deep within the granule.

At present, several models have been suggested which take into account the importance of the intergranular potential barrier in forming the transport properties of ceramic samples of lanthanum manganites. However, in spite of the efforts made in this direction, no full understanding has existed until now of the importance of the intergranular boundaries as regards the manifestation of the effect of spin-dependent tunneling; this lack of understanding shows up in some inconsistency between the suggested models. For example, the model suggested in [1, 2] is based on spin-polarized tunneling between ferromagnetic granules through an insulating nonmagnetic barrier of the granule surface. In contrast, the authors of [4, 6] assumed the emergence of magnetic polarization of the granule surface due to magnetization inside the granule. A high probability of tunneling through paramagnetic impurity states in the intergranular barrier is suggested in [17, 18]. Ziese *et al.* [11, 19] and Svistunov *et al.* [15] suggested a description assuming the presence of a spontaneous magnetically ordered state (clusters) in the barrier. In parallel with the foregoing, models [14, 20] are developed in which an additional tunneling barrier between granules arises because of curving of zones between the inner volume of the granule and its surface after the transformation of the granule content to the magnetically ordered state.

Of interest from the standpoint of investigating the importance of the surface of granules as the potential barrier are the procedures that cause a variation of the pattern of flow of tunneling current through the barrier. In this respect, unique possibilities are offered by the method of the hydrostatic compression of the sample, because the reaction of the conductance of the junction between metals to pressure depends considerably on

the barrier structure (see, for example, [21–23]). Note that most of the previous tunneling experiments with lanthanum manganites pursued the objective of producing the maximum magnetoresistive effect in low magnetic fields.Thus, the capabilities of the tunneling effect as a tool for spectroscopic investigations were not fully realized.

It is the objective of our investigation to test the validity of the assumptions made in [6, 11, 15, 19] about the important part played by the magnetic state of the boundary region of a granule (interface) in forming the transport characteristics both of tunneling junctions and of ceramic samples of lanthanum manganites. We employed the procedure of hydrostatic compression of samples and the available data on the variation of the transport properties of manganites under pressure [24–28]. We expected to observe manifestations of the internal spatially nonuniform magnetic structure of the interface in the tunneling characteristics of junctions, as well as the sensitivity of these characteristics to low magnetic fields and pressures. Indeed, the spectroscopic peculiarities of tunneling enabled us to analyze the observed total suppression of low-field magnetoresistive effect under conditions of hydrostatic compression and relate this phenomenon to the structure of spatially nonuniform state of the manganite microsrystal being investigated.

2. MEASUREMENT PROCEDURE AND SAMPLES

The possibility of obtaining tunneling junctions on a single granule was demonstrated by the results of numerous investigations of superconducting cuprate perovskites similar to manganites in morphology and mechanical properties (see, for example, [29] and the references cited there), while the procedure we employed for preparing stable microcrystal junctions made it possible to perform investigations at high hydrostatic pressures.

The procedure for preparing single microcrystal junctions consists in making a ceramic plate sized $0.1 \times 1 \times 10$ mm^3, whose granules are aligned in one direction and are intimately mated with one another to form a highly textured structure. The plate is glued to an elastic substrate. The substrate is bent until a crack appears in the ceramic plate that passes through all granules in the region of deformation. When the external load is relieved, the plate returns to the initial position, the crack "closes," and the microcrystals are tightly pressed against one another on the lines of break. The most "correct" alignment of the break of microcrystals must be expected in the plate region in which the shear deformation is minimal. This is apparently one of the reasons why such a procedure results in the realization of only one effective junction of the microcrystal–microcrystal type. In the literature, junctions prepared by a similar procedure came to be known as break junctions [29–31]. The choice of a single junction with minimal tunneling resistance from the competing junctions is further assisted by the very specific nature of the tunneling effect in which the value of current depends exponentially on the barrier thickness (a variation of the barrier thickness by 1 Å usually causes a variation of the junction resistance several times over). The small thickness of the plate according to the present procedure is an additional important factor in preparing microcrystal–microcrystal junctions.

We investigated $La_{0.6}Sr_{0.4}MnO_3$ ceramic samples prepared by the procedure described in [15]. In particular, the reduced temperature and short time of annealing made it possible to prepare plates whose intergrain bonds exhibited the tunneling pattern of current flow, as reflected in the observation of a clearly defined low-field magnetoresistive effect due to spin-dependent tunneling of electrons. The resistance of initial plates at room temperature was in the range from 2 to 10 Ω. The resistance of current and potential junctions was $R \sim 10^{-7}$ Ω cm^2.

In order to check the quality of prepared plates of $La_{0.6}Sr_{0.4}MnO_3$, the temperature dependence of their resistance was measured. The latter had a characteristic maximum in the neighborhood of 360 K, associated with the metal–dielectric transition. Also measured was the magnetoresistive effect, $[\rho(T, 0) - \rho(T, H)]/\rho(T, 0)$, in a low magnetic field; at $T = 77$ K and $H = 100$ Oe, this effect was 3–4%.

As in [15], the tunneling junctions on plates of $La_{0.6}Sr_{0.4}MnO_3$ ceramic were prepared using two methods. Junctions of the nonmagnetic metal–ceramic type (Pb–I–LaSrMnO junctions, where I is the insulator) were formed by pressing a droplet of lead into the surface of ceramic plates. Such junctions are formed at a break of microcrystals of manganite and penetration of metal (Pb) into the break region. The choice of lead as electrode made possible the observation of the energy gap of Pb upon transition to the superconducting state. The presence of superconducting gap was unambiguous proof of the tunneling mechanism of the flow of current in Pb–I–LaSrMnO junctions.

Symmetric LaSrMnO–I–LaSrMnO junctions of the type of break junction were prepared by the procedure described above. Note that another reason for the emergence of an additional contribution to the tunneling barrier in junctions of this type is apparently the depletion of the break surface of the ceramic plate in charge carriers. As in the case of cuprate superconductors, the oxygen concentration plays an important part in the position of the lanthanum manganite system in the phase diagram. In the case of mechanical break of ceramic, oxygen diffuses from the break surface, which leads to the formation of a dielectric (semiconductor) interlayer on the granule surface. It is this fact that promotes the formation of tunneling junction.

The junctions were prepared at both nitrogen and room temperatures; however, their quality did not depend much on the temperature at which the break occurred. The characteristics of junctions prepared from plates made in a single production cycle differed lit-

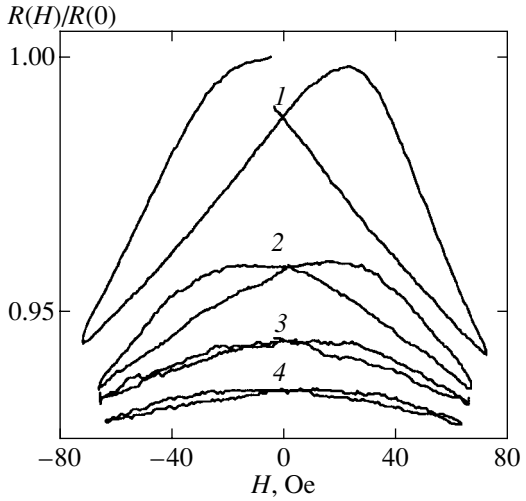

Fig. 1. The effect of pressure on the magnetoresistive properties of a $La_{0.6}Sr_{0.4}MnO_3$ ceramic plate in low magnetic fields: P = 0, 3, 6.5, and 9.8 kbar (curves *1–4*, respectively).

tle from one another. Samples with the junction resistance of several tens of ohms were selected for tunneling studies, which exceeded greatly the plate resistance and obviated the problem of four-probe connection of the junction [23]. The hydrostatic compression of samples was performed in a chamber of the piston–cylinder type with a kerosene-oil transmission medium. A manganin pressure cell and a copper-wire thermometer were placed inside the chamber for monitoring the parameters.

3. EXPERIMENTAL RESULTS

We started our investigations with studying the transport characteristics of $La_{0.6}Sr_{0.4}MnO_3$ ceramic plates in the absence of external pressure and the variation of these characteristics under the effect of pressure. In accordance with the literature data on samples of the composition being investigated [24–28], a strong effect of pressure on conductance was observed in the entire measured temperature range. Figure 1 gives the reduced resistance $R(H, P)/R(0, 0)$ of a $La_{0.6}Sr_{0.4}MnO_3$ ceramic plate as a function of magnetic field at T = 77 K for different values of applied pressure P. In view of the fact that, at low temperatures, the magnetic state inside granules varies little under the effect of pressure, it is natural to assume that the observed variation of $R(H)$ with pressure is due to the effect of hydrostatic compression on the conductance of the intergrain boundaries (interface). At the same time, it is indicative of the significant contribution made by the interface to the sample impedance.

A similar effect of pressure on conductance is observed for tunneling junctions. Figures 2 and 3 give the $R(H, P)/R(0, 0)$ characteristics for an asymmetric (LaSrMnO–I–Pb) and a symmetric (LaSrMnO–I–LaSrMnO) junction depending on magnetic field at T = 77 K and for different values of applied pressure. One can see in Fig. 2 that the magnetoresistive properties of a junction are retained even if only one of the junction sides is of a magnetoresistive material. With a pressure $P \approx 8$ kbar, the same as for a bulk sample, the effect disappears almost completely.

Note that, in hysteretic dependences $R(H)$ for bulk samples (Fig. 1) and for tunneling junctions (Figs. 2 and 3), the maxima of resistance in magnetic fields $H \approx$ 30 Oe virtually coincide. Hysteresis on the $R(H)$ curves is caused by the residual magnetization of microcrystals of the sample, and the field of maximum resistance corresponds to the field of coercivity $H_C \approx 30$ Oe.

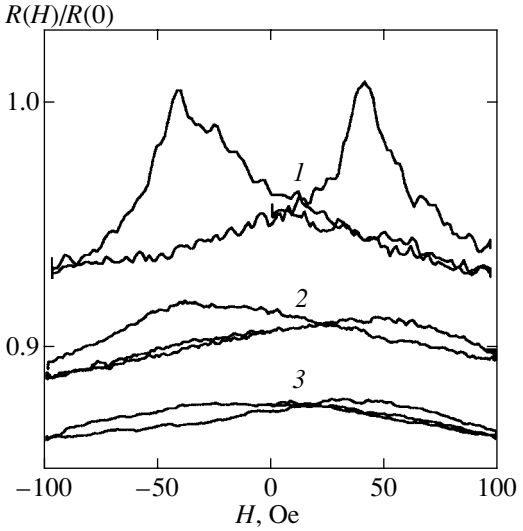

Fig. 2. The effect of pressure on the magnetosensitive part of tunneling conductance of an asymmetric (Pb–I–$La_{0.6}Sr_{0.4}MnO_3$) junction in low magnetic fields: P = 0, 3, and 6 kbar (curves *1–3*, respectively).

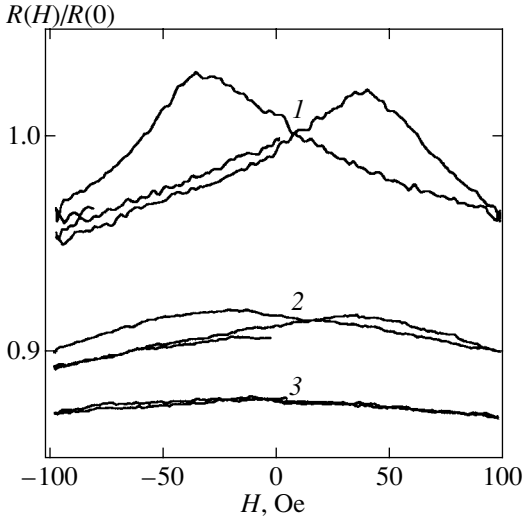

Fig. 3. The effect of pressure on the magnetosensitive part of tunneling conductance of a symmetric ($La_{0.6}Sr_{0.4}MnO_3$–I–$La_{0.6}Sr_{0.4}MnO_3$) junction in low magnetic fields: P = 0, 3, and 6.5 kbar (curves *1–3*, respectively).

Figures 4 and 5 give the junction conductance as a function of bias voltage at $T = 77$ K for different values of pressure for symmetric and asymmetric contacts, respectively. The junctions of both types demonstrate the linear pattern of conductance at low temperatures, $\sigma = \sigma_0(1 + \gamma|V|)$, in a wide range of values of bias voltage, $|eV| \le E_C \sim 100$ meV. It is shown in the subsequent section that the Coulomb interaction of electrons and holes on localized levels in the barrier forms this anomalously strong linear dependence of the junction conductance on voltage. The insets in Figs. 4 and 5a give the conductance of tunneling contacts at $T = 4.2$ K and zero pressure in the region of low values of bias voltage. The tunneling pattern of the flow of current is demonstrated especially clearly by the manifestation of superconducting energy gap of lead, $\Delta_s \approx 1.4$ meV, in the conductance of asymmetric contact (indicated by arrows in Fig. 5a). The strong diffusion of the gap characteristic in the density of states for lead is apparently due both to the effect of localized levels on the tunneling process and to the injection of spin-polarized electrons to superconductor. Also observed for different junctions were singularities behind the energy gap of lead (see Fig. 5a), which were due, in all probability, to inelastic tunneling of electrons through impurity states in the barrier with emission of real phonons (magnons) [23]. One can see in Figs. 4 and 5 that, as the hydrostatic compression increases, the normalized conductance $\sigma(V)/\sigma(0)$ becomes less sensitive to voltage. This fact points to rearrangement of the intergrain potential barrier and to the variable pattern of tunneling of electrons under the effect of pressure.

As was noted above, the junctions being investigated meet all of the criteria of "tunneling," namely, the presence of a superconducting energy gap and the manifestation of phonon (magnon) singularities and a Coulomb gap. Nevertheless, the temperature dependence of the resistance of the LaSrMnO–I–Pb junction (curve *1* in Fig. 6) differs considerably from the classical tunneling dependence $R(T)$, which is characterized by a weak temperature decrease of resistance, $\sigma(T) = \sigma_0 + \gamma T + \alpha T^2$ [23]. Such a variation of $\sigma(T)$ is typical in the temperature range $k_B T \sim 0.1$ eV (~1000 K). In our case, at $T \approx 250$ K, an anomalously fast increase in the junction resistance is observed with decreasing temperature, which is hard to explain within the classical tunneling effect. We attribute the unusually high rate of increase in the tunneling resistance even at $T \sim T^* < 300$ K to the opening of the pseudogap Δ_p on the sides or at the interface of the tunneling junction at a temperature $T < T^*$. (Note that, while the Coulomb gap E_C characterizes the "two-dimensionality" of the barrier, the main contribution to the pseudogap Δ_p is made by three-dimensional Coulomb correlations. The presence of a pseudogap in compounds with a colossal magnetoresistive effect was directly found recently with the aid of the ARPES method [32]).

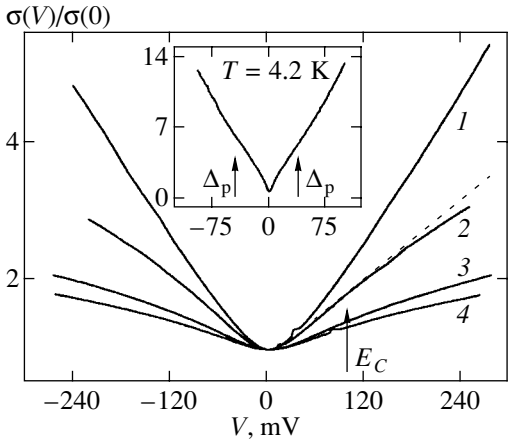

Fig. 4. The variation of reduced conductance $\sigma(V)/\sigma(0)$ of a symmetric $La_{0.6}Sr_{0.4}MnO_3$–I–$La_{0.6}Sr_{0.4}MnO_3$ tunneling junction at $T = 77$ K under the effect of pressure: $P = 0$, 3, 6.5, and 10 kbar (curves *1–4*, respectively); the arrows indicate the energy of the Coulomb gap, $E_C \approx 100$ meV. The inset gives the same correlation for $P = 0$ kbar and $T = 4.2$ K in the range of low values of bias voltage; the arrows indicate the position of pseudogap $\Delta_p \approx 40$ meV.

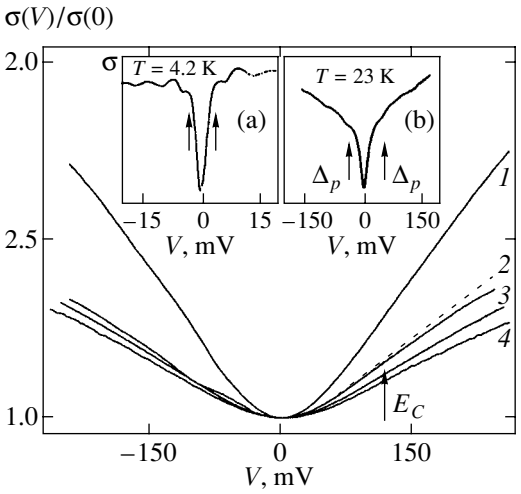

Fig. 5. The variation of reduced conductance $\sigma(V)/\sigma(0)$ of an asymmetric (Pb–I–$La_{0.6}Sr_{0.4}MnO_3$) tunneling junction at $T = 77$ K under the effect of pressure: $P = 0$, 3, 6.5, and 10 kbar (curves *1–4*, respectively); the arrows indicate the energy of Coulomb gap, $E_C \approx 100$ meV. The inset a gives the same correlation for $P = 0$ kbar and $T = 4.2$ K in the range of low values of bias voltage; the arrows indicate the energy gap for Pb $\Delta_S \approx 1.4$ meV. The inset b gives the same correlation for $P = 0$ kbar and $T = 23$ K; the arrows indicate the position of pseudogap $\Delta_p \approx 40$ meV.

The reflection of such a pseudogap in the correlation $\sigma(V)$ for LaSrMnO–I–Pb is given in the inset in Fig. 5b. A similar but less pronounced singularity is also observed in the case of symmetric LaSrMnO–I–LaSrMnO junctions (see the inset in Fig. 4). Curiously, the value

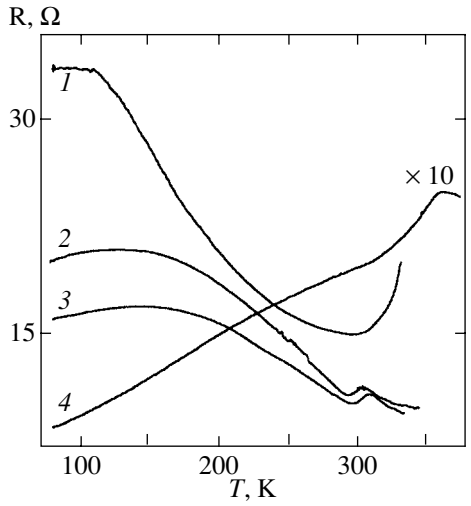

Fig. 6. The temperature dependence of the resistance of an asymmetric (Pb–I–$La_{0.6}Sr_{0.4}MnO_3$) tunneling junction at P = 0, 3, and 6 kbar (curves *1–3*, respectively) and of a ceramic plate at P = 0 kbar (curve *4*).

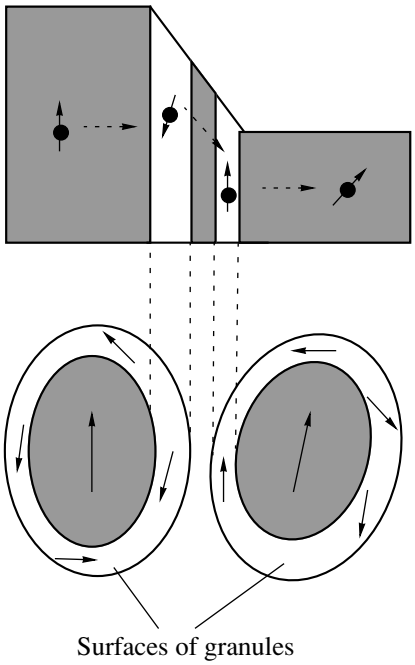

Fig. 7. A model of tunneling junction between two granules of lanthanum manganite. The tunneling barrier is formed both by the intergranular layer and by the surface regions of granules.

of pseudogap $\Delta_p \approx 40$ meV and the value of temperature $T^* \approx 280$ K, determined by these singularities, are in adequate agreement with the BCS theory, $2\Delta_p = 3.5T^*$. As the pressure increases, the $R(T)$ curve is somewhat smoothed (see curves *2* and *3* in Fig. 6); however, the behavior of $R(T)$ retains its singularity (abrupt increase) at the point $T \sim T^* = 280$ K. Therefore, the pressure dependence of the value of pseudogap remains virtually unvaried. The observed smoothing of the $R(T)$ characteristic of tunneling junctions under the effect of pressure points to the transformation of the magnetic structure and to the metallization of the surface layers of the junction.

Figure 6 (see curve *4*) further gives the temperature dependence of resistance of $La_{0.6}Sr_{0.4}MnO_3$ ceramic plate in zero external field (note the scale variation). This dependence is typical of the class of compounds being treated [24–28]. A comparison of curves *1* and *4* reveals that the behavior of the temperature dependence of tunneling contact resistance is entirely different from that of the temperature dependence of ceramic. This is due to the fact that, in the latter case, the sample resistance is formed both by the interface regions and by the internal regions of granules; as the temperature increases, this contribution to resistance becomes determining.

4. DISCUSSION OF THE RESULTS

We will treat in more detail the dependence of the junction conductance on the bias voltage and temperature. The term $q|V|$ ("line background") is due to resonance tunneling of electrons through localized states located at the center of tunneling barrier [33, 34]. We will demonstrate that such processes are inelastic; i.e., they must be accompanied by emission of some boson mode.

We will treat the tunneling of electrons through a system of localized states in a barrier, characterized by the importance of the Coulomb interaction of electrons (or holes) at impurity centers (see Fig. 7). We will assume that a positive voltage is applied to the right-hand electrode of a tunneling junction. We will reckon the values of energy E_i of localized states starting from the level of the chemical potential of this electrode. We will analyze only the resonance processes, i.e., the processes of tunneling through localized levels located at the center of the tunneling barrier. For qualitative analysis, we will restrict ourselves to the case of zero temperature, when localized levels with energy $E_i \geq 0$ take part in the process of tunneling. We assume the system as a whole to be electroneutral.

The characteristic lifetime of an electron on an impurity center with energy $E_i \geq 0$ is of the order of $\hbar/\Gamma$, where $\Gamma = \Gamma_0 \exp(-d/a_0)$ is the energy width of the impurity state, a_0 is its radius, and d is the tunneling barrier thickness (the parameter Γ_0 is estimated in [35]). We will assume that the electron tunnels to unfilled level 1 with energy E_1 from the left-hand side of the junction. In order to preserve electroneutrality in a system of barrier states, it is necessary that simultaneously (during time $\tau \ll \hbar/\Gamma$ in quantum-mechanical sense), an electron from filled level 2 (with energy E_2) should go to the right-hand side of the junction. Therefore, a positively charged "hole" emerges on level 2; the

attraction of electron 1 to this hole reduces the energy of the system of barrier states by $e^2/4\pi\varepsilon\varepsilon_0 r_{12}$. By summing up the energy losses, we will arrive at the inequality

$$\delta E_{12} = E_1 - E_2 - e^2/4\pi\varepsilon_0\varepsilon r_{12} > 0, \quad (1)$$

where r_{12} is the distance between the 1st and 2nd impurities, and ε_0 and ε are dielectric constants. Note the peculiar nature of the tunneling process being analyzed. The incoming and outgoing electrons are spaced apart at a distance r_{12} which may be fairly long (especially, in the case of low voltage V across the junction). In fact, the system of impurity levels "captures" the first electron and emits the other electron.

The fact that the increment of energy given by Eq. (1) is positive for the overwhelming majority of tunneling processes implies that the distribution of levels of energy E_i of localized states in the barrier corresponds to equilibrium. According to Eq. (1), states with low energies E_1 and E_2 must be spaced at a considerable distance r_{12} from each other. This explains the decrease in the level density $g(E)$ for low values of energy (the effect of Coulomb gap in impurity semiconductors [36]). The exact distribution of $g(E)$ over the energies E is obtained as a result of numerical simulation of the above-identified tunneling by the Monte Carlo method. The result of [33, 34] brings about the following dependence for the Coulomb gap: $g(E) \propto |E|^\gamma$, where the exponent γ depends on the ratio of the width W of the scatter of impurity states with respect to energies E_i to the Coulomb gap "width" $E_C \approx e^2/4\pi\varepsilon\varepsilon_0 r_D$ (r_D is the mean distance between impurity states in the central interlayer of the barrier).

The observed value of the Coulomb gap $E_C \approx 100$ meV (see Figs. 4 and 5) enables one to estimate the parameters of impurity states of the interface. We assume that the charges in the interface are screened similarly with those in cuprate oxides. Then, at $\varepsilon \approx 4$, we obtain $r_D = 36$ Å, which corresponds to the concentration of localized states $N_l = r_D^{-3} \approx 2 \times 10^{19}$ cm^{-3}. In semiconductors, the transition from metallic to activation conductance usually occurs at values of concentration satisfying the Mott criterion, $N_l a^3 \approx k_M$, where a is the radius of impurity states. In view of the fact that the states in the interface are at the percolation threshold, for $r_D = 36$ Å and $k_M \approx 0.02$, we obtain the radius $a = 9$ Å, which is very close to the estimate of the radius of polaron states in manganites ($a = 8$ Å) (see, for example, [37, 38]). The value of the Coulomb gap $E_C \approx 100$ meV agrees with the values of activation gaps, obtained as a result of measurements of the thermopower, and with estimates of the characteristic energy of local single-impurity states [38]. Corresponding to the observed variation of tunneling conductance in the region of the Coulomb gap $\Delta\sigma \approx 1\ \Omega^{-1}$ is $N = \Delta\sigma/\sigma_C \approx 6 \times 10^3$ localized levels with energy width $\Gamma = N(2\pi^2 a S\rho_1)^{-1}$ meV (here, $\sigma_C = e^2/\pi\hbar \approx 77.5 \times 10^{-5}\ \Omega^{-1}$ is the "conductance quantum," $\rho_1 = N_1/W$ is the density of states for localized levels, and $W \approx 2E_C$ is the width of the impurity zone).

With narrow energy distribution of impurity states ($W/E_C \le 0.7$), the exponent $\gamma < 1$, and, with a further decrease in W, a wide minimum appears on the function $g(E)$ with $|V| \le E_C$ at the expense of the so-called Coulomb blockade of the tunneling current. If $W/E_C \ge 1$, then $\gamma \approx 1$ to 1.3, and the function of tunneling conductance of junction acquires a line background,

$$\sigma(V) = dJ/dV = \sigma_0 + q|eV|^\gamma,$$

where q is a constant and σ_0 is the background nonresonance part of junction conductance.

The excess energy δE_{12} given by Eq. (1) and arising upon injection of electron 1 and emission of electron 2 may be transferred with the aid of magnons, phonons, or other Bose-like excitations that are possible in a system of charged impurity states. Their presence provides for the synchronization of the process of entry of electron 1 and exit of electron 2. Superficially, this process appears to be an inelastic tunneling of an electron from the left-hand side of the junction to the right-hand side, from the state with energy E_1 to that with energy E_2, for which the expression for the energy difference $\Delta E_{12} = \delta E_{12} + e^2/4\pi\varepsilon_0\varepsilon r_{12}$ is valid for an impurity system as well.

We will analyze the effect of temperature on tunneling current through localized states. In the general case, the inelastic tunneling current may be represented as [23, 33, 34] $J = J_1 + J_2$, where $e = \hbar = 1$,

$$J_1 \sim \int d\omega F(\omega)[S(V-\omega) - S(-V-\omega)],$$
$$S(E) = \int_{-\infty}^{\infty} dE' f(E')(1 - f(E+E')) \quad (2)$$
$$= \frac{E}{1 - \exp(-E/kT)},$$

$f(E)$ denotes the Fermi distribution functions, and $F(\omega) \equiv F(\omega, T)$ is the effective spectral function of distribution of boson excitations in a system of impurity states. For simulation using the Monte Carlo method, the spectrum of effective excitations in the impurity system was approximated by the expression

$$F(\omega, T) = \frac{2}{\pi}\arctan\frac{\omega}{aT} \quad (3)$$

(the value of the parameter a was determined by varying the integral and was in the range $a = 1$ to 2). For $|eV| \gg kT$, the conductance $\sigma_1 = dJ_1/dV \propto e|V|$.

The current component is

$$J_2 \sim \int d\omega F(\omega) n(\omega) [S(V-\omega) - S(\omega - V) + S(V+\omega) - S(-\omega - V)] = VG_2T,$$
$$\sigma_2 = dJ_2/dV \propto G_2T, \quad G_2 = 2\int_0^\infty \frac{F(x,T)}{\exp(x)-1}dx, \tag{4}$$

and $n(\omega)$ denotes the Bose distribution functions.

For $|eV| \gg kT$, we have $F(\omega, T) \approx 1$; therefore, at high voltages, the junction admittance is $\sigma(V) \sim \sigma_0 + q|eV|$. On the other hand, at $V = 0$, the conductance is

$$\sigma(T) \sim 2\int_0^\infty d\omega F(\omega)(S'(-\omega) + n(\omega)) = 2kT\alpha_T,$$
$$\alpha_T = \frac{1}{4}\int_0^\infty \frac{xF(x,T)}{\sinh^2(x/2)}dx. \tag{5}$$

Therefore, the assumption of inelastic behavior of resonance tunneling with the emission of collective mode with the spectrum $F(\omega, T)$ agrees with experiment if $F(\omega, T) \longrightarrow 1$ at $\hbar\omega \gg kT$. As a result, for the effective level density $g(E)$, we derive the temperature dependence

$$g(E) = G_2T + \int_0^\infty d\omega F(\omega, T).$$

Further, the value of a in expression (3) was preassigned, and the constants G_2 given by Eq. (4) and α_T given by Eq. (5) were calculated. It was found that the parameter G_2 and the coefficient α_T in the temperature dependence of tunneling conductance $\sigma(T)_{V=0} = \sigma_0 + 2q\alpha_T(kT)$ were related as $\alpha_T \approx G_2^{0.83}$. The experimentally observed value of the parameter $\alpha_T = 1.1$–1.2 is attained at $G_2 \approx 1.2$.

It was assumed during analysis that the line background of $q|V|$ is due to resonance tunneling of electrons through localized states (this provides for greater effect). The process of electron injection from the left-hand side of the junction to level 1 and of the electron exit from level 2 to the right-hand side is treated as an inelastic process during which two excitations form simultaneously, namely, magnon (phonon) + electron-hole pair ("exciton"). The effective spectrum of such combined Bose excitation is given by an expression of the type of Eq. (3). It brings about the emergence of the line background of $q|V|$ as a function of junction conductance. The observed independence of the parameter q of the magnetic field leads one to conclude that for the junctions investigated by us, in the case of resonance tunneling, neither the energy nor the spin of tunneling electrons are preserved. This is possible only if the inelastic tunneling is accompanied by the magnon emission in the region of the junction interface.

5. CONCLUSION

We have studied the effect of hydrostatic compression on the transport properties of a $La_{0.6}Sr_{0.4}MnO_3$ ceramic and of $La_{0.6}Sr_{0.4}MnO_3$–I–Pb and $La_{0.6}Sr_{0.4}MnO_3$–I–$La_{0.6}Sr_{0.4}MnO_3$ tunneling junctions. An analysis of the experimental data has revealed that the pressure does not cause a variation of the tunneling behavior of the passage of electrons in the structures being studied. To the contrary, the temperature dependence of the current passing through a junction assumes a more "classical tunneling form". At the same time, the magnetoresistive effect is suppressed, although the magnetic properties of the junction sides are retained. A conclusion has been drawn that the region of the junction interface (of nanometer width) is in the vicinity of the metal–insulator phase transition boundary. As is known, this latter interface may be shifted by applying a pressure of the order of $P \sim 10$ kbar [24–28]. It was this pressure range that was realized in our investigation. Proceeding from the obtained data and the results of analysis, it is natural to assume that the surface layer of electrodes of the investigated tunneling junctions and ceramic granules is in the phase of a magnetic insulator (semiconductor). The pressure transfers these regions to the metallic state, so that the magnetic state of the internal regions of a granule extends to its surface areas, and the tunneling assumes a "standard pattern."

Therefore, in ceramic systems of lanthanum manganites, the formation of intergrain tunneling junctions is responsible for the spin-dependent tunneling of electrons and for the low-field magnetoresistive effect. In this way, the granule surface represents an additional tunneling junction. Most effective is the resonance tunneling through impurity levels located at the center of a potential barrier formed both by the true (insulating) barrier and by the interface of contacting granules. The tunneling through such states leads to the formation of a Coulomb gap in the density of states of charge carriers. For a fairly large barrier thickness, the resonance processes with the boson emission (absorption) prevail. The processes of tunneling through the impurity states of a nanoscale magnetic cluster make a decisive contribution to the low-field magnetoresistive effect.

REFERENCES

1. H. Y. Hwang, S.-W. Cheong, N. P. Ong, and B. Batlogg, Phys. Rev. Lett. **77**, 2041 (1996).
2. Yu Lu, X. W. Li, G. Q. Gong, and G. Xiao, Phys. Rev. B **54**, 8357 (1996).
3. J. M. D. Coey, M. Viret, and S. V. Molnar, Adv. Phys. **48**, 167 (1999).
4. J. E. Evetts, M. G. Blamire, N. D. Mathur, *et al.*, Philos. Trans. R. Soc. London, Ser. A **356**, 1593 (1998).
5. M. Julliere, Phys. Lett. A **54A**, 225 (1975).

6. N. Zhang, W. Ding, W. Zhong, *et al.*, Phys. Rev. B **56**, 8138 (1997).
7. J. Y. T. Wei, N.-C. Yeh, R. P. Vásquez, and A. Gupta, J. Appl. Phys. **83**, 7366 (1998).
8. K. Steenbeck, T. Eick, K. Krisch, *et al.*, Appl. Phys. Lett. **73**, 2506 (1998).
9. S. Lee, H. Y. Hwang, B. I. Shraiman, *et al.*, Phys. Rev. Lett. **82**, 4508 (1999).
10. J. M. De Teresa, A. Barthelemy, A. Fert, *et al.*, Phys. Rev. Lett. **82**, 4288 (1999).
11. M. Ziese, Phys. Rev. B **60**, R738 (1999).
12. P. Lyu, D. Y. Xing, and J. Dong, Phys. Rev. B **60**, 4235 (1999).
13. A. Tiwari and K. P. Rajeev, Phys. Rev. B **60**, 10591 (1999).
14. R. Gross, L. Alff, B. Buchner, *et al.*, J. Magn. Magn. Mater. **211**, 150 (2000).
15. V. M. Svistunov, Yu. V. Medvedev, V. Yu. Tarenkov, *et al.*, Zh. Éksp. Teor. Fiz. **118**, 629 (2000) [JETP **91**, 547 (2000)].
16. H. B. Peng, X. X. Zhang, Z. Xie, *et al.*, Phys. Rev. B **61**, 8955 (2000).
17. F. Guinea, Phys. Rev. B **58**, 9212 (1998).
18. R. Jansen and J. C. Lodder, Phys. Rev. B **61**, 5860 (2000).
19. M. Ziese, G. Hyedon, R. Hohne, *et al.*, Appl. Phys. Lett. **74**, 1481 (1999).
20. N. I. Solin, S. V. Naumov, and A. A. Samokhvalov, Fiz. Tverd. Tela (St. Petersburg) **42**, 899 (2000) [Phys. Solid State **42**, 925 (2000)].
21. V. G. Bar'yakhtar, O. V. Grigut', A. V. Vasilenko, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **47**, 457 (1988) [JETP Lett. **47**, 540 (1988)].
22. V. M. Svistunov, V. Yu. Tarenkov, A. I. D'yachenko, *et al.*, Zh. Éksp. Teor. Fiz. **113**, 1397 (1998) [JETP **84**, 763 (1988)].
23. E. L. Wolf, *Principles of Electron Tunneling Spectroscopy* (Oxford Univ. Press, New York, 1985; Naukova Dumka, Kiev, 1990).
24. H. Y. Hwang, T. T. M. Palstra, S.-W. Cheong, and B. Batlogg, Phys. Rev. B **52**, 15046 (1995).
25. K. Khazeni, Y. X. Jia, Li Lu, *et al.*, Phys. Rev. Lett. **76**, 295 (1996).
26. J. J. Neumeier, M. F. Hundley, J. D. Thompson, and R. H. Heffner, Phys. Rev. B **52**, R7006 (1995).
27. R. Wang, R. Mahesh, and M. Itoh, Phys. Rev. B **60**, 14513 (1999).
28. J. M. De Teresa, M. R. Ibarra, J. Blasco, *et al.*, Phys. Rev. B **54**, 1187 (1996).
29. A. I. D'yachenko, V. Yu. Tarenkov, R. Szymczak, *et al.*, Phys. Rev. B **61**, 1500 (2000).
30. J. Moreland and J. W. Ekin, J. Appl. Phys. **68**, 4876 (1990).
31. N. Miyakawa, P. Guptasarma, J. F. Zasadzinski, *et al.*, Phys. Rev. Lett. **80**, 157 (1998).
32. T. Saitoh, D. S. Dessau, Y. Moritomo, *et al.*, Phys. Rev. B **62**, 1039 (2000).
33. A. I. D'yachenko, V. Yu. Tarenkov, A. V. Abaleshev, and V. M. Svistunov, Fiz. Tverd. Tela (St. Petersburg) **38**, 1690 (1996) [Phys. Solid State **38**, 933 (1996)].
34. A. I. D'yachenko, Fiz. Tekh. Vys. Davleniĭ **8**, 47 (1998).
35. A. I. Larkin and K. A. Matveev, Zh. Éksp. Teor. Fiz. **93**, 1030 (1987) [Sov. Phys. JETP **66**, 580 (1987)].
36. A. L. Efros and B. I. Shklovski, J. Phys. C **8**, L49 (1975).
37. M. Hennion, F. Moussa, G. Biotteau, *et al.*, Phys. Rev. Lett. **81**, 1957 (1998); Phys. Rev. B **61**, 9513 (2000).
38. V. M. Loktev, Yu. G. Pogorelov, Fiz. Nizk. Temp. **26**, 231 (2000) [Low Temp. Phys. **26**, 171 (2000)].

Translated by H. Bronstein

Journal of Experimental and Theoretical Physics, Vol. 93, No. 1, 2001, pp. 188–196.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 1, 2001, pp. 214–223.

NONLINEAR PHYSICS

Big Entropy Fluctuations in Statistical Equilibrium: the Macroscopic Kinetics¶

B. V. Chirikov* and O. V. Zhirov**
Budker Institute of Nuclear Physics, Novosibirsk, 630090 Russia
**e-mail: B.V.Chirikov@inp.nsk.su*
***e-mail: zhirov@inp.nsk.su*
Received November 20, 2000

Abstract—Large entropy fluctuations in the equilibrium steady state of classical mechanics are studied in extensive numerical experiments in a simple strongly chaotic Hamiltonian model with two degrees of freedom described by the modified Arnold cat map. The rise and fall of a large separated fluctuation is shown to be described by the (regular and stable) "macroscopic" kinetics, both fast (ballistic) and slow (diffusive). We abandon a vague problem of the "appropriate" initial conditions by observing (in a long run) a spontaneous birth and death of arbitrarily big fluctuations for any initial state of our dynamical model. Statistics of the infinite chain of fluctuations similar to the Poincaré recurrences is shown to be Poissonian. A simple empirical relationship for the mean period between the fluctuations (the Poincaré "cycle") is found and confirmed in numerical experiments. We propose a new representation of the entropy via the variance of only a few trajectories ("particles") that greatly facilitates the computation and at the same time is sufficiently accurate for big fluctuations. The relation of our results to long-standing debates over the statistical "irreversibility" and the "time arrow" is briefly discussed. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION: MACROSCOPIC VERSUS MICROSCOPIC FLUCTUATIONS

Fluctuations are an inseparable part of statistical laws. This has been well known since Boltzmann. What is apparently less known are the peculiar properties of rare big fluctuations (BF) that are different from, and even in a sense opposite to, the properties of small stationary fluctuations. In this paper, we consider the simplest type of chaotic dynamical systems, namely, a Hamiltonian system with a finite number of the degrees of freedom that admits the (stable) statistical equilibrium (SE). This class of dynamical models is still popular (since Boltzmann!) in debates over the dynamical foundations of statistical mechanics (see, e.g., "Round Table on Irreversibility" in [1, 2]).

A sufficiently simple picture of BFs in such systems is well understood by now, although not yet well known. To Boltzmann, this picture was the basis of his fluctuation hypothesis for our Universe. It is also well understood that this hypothesis is totally incompatible with the present structure of the Universe because it would immediately imply the notorious "heat death" (see, e.g., [3]). For this reason, one may even term such systems the heat death models. Nevertheless, they can be and actually are widely used in describing and studying local statistical processes in thermodynamically closed systems. The latter term means the absence of any heat exchange with the environment. We note, however, that under conditions of the exponential instability of motion, which are typical of chaotic systems, the only dynamically closed system would be the "entire Universe." In particular, this excludes the hypothetical "velocity reversal" that also is popular in debates over "irreversibility" since Loschmidt (for a discussion, see, e.g., [4]).

In any case, dynamical models with the SE do not tell us the whole story of either the Universe or even a typical macroscopic process therein. The principal solution of this problem, unknown to Boltzmann, is quite clear now: the "equilibrium-free" models are required. Various classes of such models are intensively studied today. Moreover, the celebrated cosmic microwave background tells us that our Universe was born already in the state of a heat death, which, however, became unstable due to the well-known Jeans gravitational instability [5]. This resulted in developing a rich variety of collective processes, or synergetics, the term recently introduced or, better to say, put in use by Haken [6]. The most important peculiarity of such a collective instability is that the total overall relaxation (to somewhere?) with the ever increasing total entropy is accompanied by an also increasing phase space inhomogeneity of the system, particularly with respect to temperature. In other words, the entire system and its local parts become more and more nonequilibrium to the extent of the birth of a secondary dynamics that can be, and sometimes is, as perfect as, e.g., the celestial mechanics (see, e.g., [4, 7, 8] for a general discussion).

We stress that all these inhomogeneous nonequilibrium structures are not BF like in the SE but are a result

¶This article was submitted by the authors in English.

of a regular collective instability; therefore, they are immediately formed under a certain condition. In addition, they are typically dissipative structures according to Prigogine [9] due to the energy and entropy exchange with the infinite environment. The latter is the most important feature of such processes, and at the same time the main difficulty in studying the dynamics of those models both theoretically and in numerical experiments, which are so much simpler for SE systems.

In the latter case, a BF consists of two symmetric parts: the rise of a fluctuation followed by its return, or relaxation, back to the SE (see Figs. 1 and 2). Both parts are described by the same kinetic (e.g., diffusion) equation, the only difference being in the sign of time. This relates the time-symmetric dynamical equations to the time-antisymmetric kinetic (but not statistical!) equations. The principal difference between the two types of equations, sometimes overlooked, is that the kinetic equations are generally understood as describing the relaxation only, i.e., the increase of the entropy in a closed system, whereas in fact they do so (at least, in the SE) for the rise of BF as well, i.e., for the entropy decrease. All this was qualitatively known already to Boltzmann [10]. The first simple example of a symmetric BF was considered by Schrödinger [11]. A rigorous mathematical theorem for the diffusive (slow) kinetics was proved by Kolmogorov in 1937 in the paper entitled "Zur Umkehrbarkeit der statistischen Naturgesetze" (Concerning Reversibility of Statistical Laws in Nature) [12] (see also [13]). Regrettably, the principal Kolmogorov theorem still remains unknown to both the participants of heated debates over "irreversibility" and the physicists actually studying such BFs (see, e.g., [14]).

At present, there exists a well-developed ergodic theory of dynamical systems (see, e.g., [15]). In particular, it proves that the relaxation (correlation decay, or mixing) eventually proceeds in both directions of time for almost any initial conditions in a chaotic dynamical system. However, the relaxation must not always be monotonic, which simply means a BF on the way, depending on the initial conditions. To eliminate this apparently confusing (to many) "freedom," we take a different approach to the problem: instead of discussing the "true" initial conditions and/or a "necessary" restriction of them, we start our numerical experiments at arbitrary initial conditions (most likely corresponding to the SE) and observe what the dynamics and statistics of BF are like. This approach is obviously based on the fundamental hypothesis that all the statistical laws are contained in, and can be principally derived from, the underlying fundamental (Hamiltonian) dynamics. To the best of our knowledge, there is as yet no contradiction to this principal hypothesis. We note, however, that this approach can be directly applied to fluctuations in finite systems with a statistical equilibrium only (see [4] and [16] for a discussion). In these and only these systems, infinitely many BFs grow up spontaneously, independently of the initial conditions of the motion. This is similar to the well-known Poincaré recurrences (see Section 4).

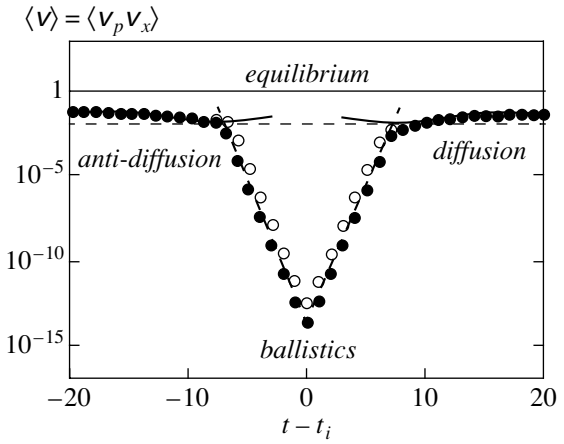

Fig. 1. Mixed kinetics for two BFs of different sizes. Filled/open circles show the time dependence of the mean variance $\langle v(t - t_i) \rangle$ around the BF maximum at $t = t_i$; the upper horizontal straight line is the equilibrium and the lower line indicates the empirical value of the dynamical scale $v_d = 0.015$, Eq. (3.4), with the parameter $F_d \approx 1/3$. The two oblique straight lines represent the expected fast kinetics, Eq. (3.3), and the two solid curves do so for the initial diffusive kinetics, Eq. (3.5). The respective run parameters and results are given by $C = 15$, $N = 1$, $v_b = 3.9 \times 10^{-11}/6.25 \times 10^{-10}$ ($v_{xb} = v_{pb}$), $v(0) = 1.96 \times 10^{-14}/3.1 \times 10^{-13}$, $n = 1971/4459$, $w = 500$. The average period between successive fluctuations is $\langle P \rangle \approx 1.4 \times 10^7/3.5 \times 10^6$ iterations.

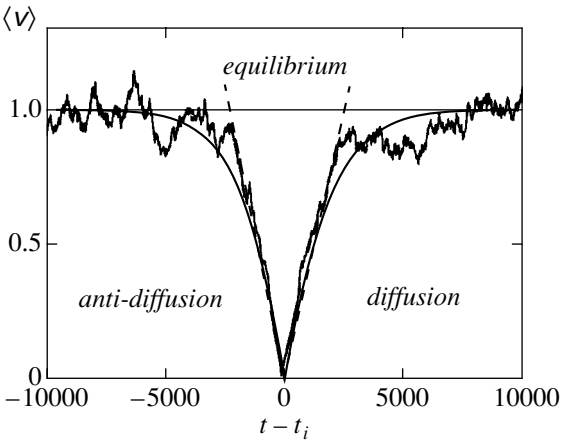

Fig. 2. The same as in Fig. 1 for a typical diffusive kinetics (anti-diffusion/diffusion): the solid curve shows the average over all $n = 20259$ fluctuations in a run and the wiggle line is the same for the first 28 fluctuations. Two oblique straight lines represent the expected initial diffusive kinetics, Eq. (3.5), with $\tau_d = 0$ and the empirical value $v_d^{(\mathrm{emp})} = 0.045$, while the theory (3.15) gives $v_d = 0.02$. Other run parameters/results are given by $C = 50$, $N = 5$, $v_b = 0.0256$, $w = 10^4$, $\langle P \rangle \approx 7.7 \times 10^5/8.7 \times 10^5$, and $B = 306/348$; $\langle P \rangle / w \approx 77/87$.

In spite of essential restrictions, simple SE models allow us to better understand the mechanism and the

role of BFs in statistical physics. In addition to the removal of the vague problem of initial conditions, these models are very helpful in clarifying the relation between macroscopic and microscopic descriptions of chaotic systems. In particular, a spontaneous rise of a BF out of the SE is a macroscopic event as well as is its subsequent relaxation back to the SE, even in a system with a few degrees of freedom. Similarly to other macroscopic processes, BFs are not only perfectly regular by themselves but also surprisingly stable against any perturbations, either regular or chaotic. Moreover, the perturbations must not be small. At a first glance, this looks very strange in a chaotic, highly unstable dynamics. The resolution of this apparent paradox is that the dynamical instability of motion affects the BF instant of time only. The BF evolution is determined by the kinetics independently of its mechanism, from a purely dynamical one, as in model (2.2) used in this paper, to a completely noisy (stochastic) one. As a matter of fact, the fundamental Kolmogorov theorem [12] is precisely related to the latter case but remains valid in a much more general situation. The surprising stability of BFs is similar to the less known concept of robustness for the Anosov (strongly chaotic) systems [17] whose trajectories are only slightly deformed under a small perturbation (see [4] for a discussion).

In this paper, we consider a particular type of BFs characterized by a large concentration of "particles" in a small phase space domain of the dynamical system. In other words, "our" fluctuations are localized in phase space and separated in time. A more accurate definition of these fluctuations is given in Section 3 (see Eq. (3.6)). The same fluctuations in a stochastic model (with noise) were studied in detail in [14]. Obviously, there exist many other fluctuations with their own peculiarities (see, e.g., [18]). The primary object of our studies is the macroscopic kinetics of big fluctuations in the background of small stationary microscopic fluctuations. A brief outline of our results was presented in [16].

2. A HAMILTONIAN MODEL: MOST SIMPLE BUT STRONGLY CHAOTIC

The systems with an SE can be described in terms of models that are very simple as regards both the theoretical analysis and numerical experiments (of which the latter are even more important for us). In the present paper, we use one of the most simple and popular models specified by the so-called Arnold cat map (see [19, 20]):

$$\begin{aligned} \bar{p} &= p + x \mod 1, \\ \bar{x} &= x + \bar{p} \mod 1, \end{aligned} \tag{2.1}$$

which is a linear canonical map on a unit torus. It has no parameters and is chaotic and even ergodic. The rate of the local exponential instability, the Lyapunov exponent $\lambda = \ln(3/2 + \sqrt{5}/2) = 0.96$, implies a fast (ballistic) kinetics with the relaxation time $t_r \sim 1/\lambda \approx 1$. Throughout the paper, t denotes the time in the map iterations.

A minor modification of this map,

$$\begin{aligned} \bar{p} &= p + x - 1/2 \mod C, \\ \bar{x} &= x + \bar{p} - C/2 \mod 1, \end{aligned} \tag{2.2}$$

where C is a circumference of the phase space torus, allows studying both the fast (exponential) ballistic kinetics (for $C = 1$) and the slow (diffusive) relaxation in p (for $C \gg 1$) with the characteristic time $t_p \sim C^2/4D_p \gg 1$, where $D_p = 1/12$ is the diffusion rate in p. In contrast to the slow diffusion in p, the relaxation time in x does not depend on C ($t_r \sim 1$) and the subsequent values of x are therefore practically uncorrelated. Map (2.2) has the (unstable) fixed point at $x = x_0 = 1/2$ and $p = p_0 = C/2$.

A convenient characteristic of the BF size is the rms volume (area) in the 2D phase space (x, p)

$$\sigma(t) = \sigma_p(t)\sigma_x(t) \tag{2.3}$$

occupied by a group of N trajectories (particles). In the ergodic motion at equilibrium, $\sigma = \sigma_0 = C/12$. Because of a severe restriction to small $N \lesssim 10$ in the numerical experiments (see below), we have to use simple (average) characteristics like (2.3) only. On the other hand, these are precisely the macroscopic variables in which we are interested.

In what follows, we also restrict ourselves to a particular case of BFs with the fixed prescribed position in the phase space,

$$x_{fl} = x_0 = \frac{1}{2}, \quad p_{fl} = p_0 = \frac{C}{2}. \tag{2.4}$$

The variance of the phase space size $v = \sigma^2 = \sigma_p^2\sigma_x^2$ is then determined by

$$\sigma_p^2 = \langle p^2 \rangle - p_0^2, \quad \sigma_x^2 = \langle x^2 \rangle - x_0^2, \tag{2.5}$$

where the brackets $\langle \ldots \rangle$ denote averaging over N trajectories. In the ergodic motion at equilibrium, $v = v_{SE} = C^2/12^2$. In what follows, we use the dimensionless measure $\tilde{v} = v/v_{SE} \longrightarrow v$ and omit the tilde. In the diffusive approximation of the kinetic equation, the variable $v(t)$ is especially convenient because it varies proportionally to time. Moreover, $v \longrightarrow v_p$ in this case because of a quick relaxation $v_x \longrightarrow 1$ in x.

Among all the advantages of v, the relation of this variable to the fundamental concept of the entropy is highly desirable. The standard definition of the entropy, which can be traced back to Boltzmann, reads

$$S = -\langle \ln f(x, p) \rangle + S_0, \tag{2.6}$$

where $f(x, p)$ is a coarse-grained distribution function, or the phase-space density, and S_0 an arbitrary constant to be fixed later. We note that the distribution calculated from any finite number of trajectories is always a

coarse-grained one. However, the direct application of Eq. (2.6) requires too many trajectories, especially for a small-size BF. Nevertheless, precisely in the latter case, which is the main problem under consideration, we have found a simple approximate relation

$$S(t) \approx \frac{1}{2}\ln v(t) \tag{2.7}$$

that gives at least a rough estimate for the entropy evolution [16]. Moreover, if the distribution is Gaussian,

$$f(x, p) \longrightarrow f(p) = \frac{\exp(-(p-p_0)^2/2v)}{\sqrt{2\pi v}}, \tag{2.8}$$

estimate (2.7) becomes exact because it is directly derived from the definition of the entropy in Eq. (2.6). The two relations for the entropy are compared in the end of Section 3 for a typical BF.

A great advantage of (2.7) is that the computation of S does not require very many trajectories as does the distribution function. In fact, even a single trajectory is sufficient, as is demonstrated by Fig. 1 in [16] and Fig. 1 in this paper!

A finite number of trajectories used for calculating the variance v is similar to a coarse-grained distribution, as required in relation (2.6), but with a free bin size that can be arbitrarily small.

We can now turn to the numerical experiments.

3. MACROSCOPIC KINETICS: COMPLETE, REGULAR, AND STABLE

In this section, we consider the regular BF kinetics. The data were obtained by simultaneously running N trajectories for a very long time in order to collect sufficiently many BFs for a reliable separation of the regular part of BFs, or the kinetic subdynamics according to Balescu (see [21] and references therein), from the stationary fluctuations. The separation was done by the plain averaging of the individual v_i values ($i = 1, \ldots, n$) over all the n BFs collected in a run.

The size of the BF chosen for the subsequent analysis is fixed by the condition that

$$v(t) < v_b \tag{3.1}$$

at some time instant $t \approx t_i$, the moment of a BF. Here, a prescribed value v_b is the main input parameter of the run. This condition actually determines the border of the entire fluctuation domain (FD) as $0 < v < v_b$.

The event of entering the FD is the macroscopic "cause" of the BF whose obvious "effect" will be the subsequent relaxation to the equilibrium. However, the main point of our study is that the second "effect" of the same "cause" was preceding the rise of the BF in an apparent contradiction with the "causality principle" (for a discussion, see [16] and Section 4 below). In any event, the second effect requires the permanent memory of trajectory segments within some time window w, which is another important input parameter of the run.

The exact procedure of data processing during the run is as follows. Starting from arbitrary (random) initial conditions, selection rule (3.1) is checked at each iteration. Suppose that it is satisfied at some instance t_{in} when the bundle of trajectories enters the FD. In the first approximation, we could consider it as the fluctuation maximum (or the variance minimum) $t_i = t_{\text{in}}$, where the subscript i is the number of the current fluctuation in a run. However, this simple procedure would cause an asymmetry with respect to $t = t_i$. A better choice would be given by the rule $t_i = (t_{\text{in}} + t_{\text{out}})/2$, where t_{out} is the time instance of the exit from the FD. Instead, we have accepted a more complicated procedure that better restores the true BF symmetry, as we hope. Starting from the moment t_{in}, we search for the minimum of $v(t)$ inside a sufficiently large interval $t_{\text{in}} < t < t_{\text{in}} + w$. If a minimum is found at some $t = t_{\min}$, we check that it also is the minimum inside the next interval $t_{\min} < t < t_{\min} + w$. If this is the case, we identify this minimum with the BF top and set $t_i = t_{\min}$; otherwise, we set $t_{\min}$ equal to the time of a better minimum and repeat the last step. Obviously, the parameter w must be small compared to $\langle P\rangle$, the mean period of the BF, but sufficiently long for the trajectory to leave FD (3.1). Typically, we chose $w \gtrsim C^2$, the total diffusion time. After fixing the current t_i value, the computation within the interval $t_i < t < t_i + w$ was completed, and only then the search for the next BF is continued.

As mentioned above, there are two quite simple limiting cases of generally very complicated kinetics, namely, the fast (ballistic) and the slow (diffusive) limits. An example of both in one run for $N = 1$ (!) is presented in Fig. 1 for two fluctuations of different sizes. In this case, general condition (3.1) was checked separately for p and x,

$$v_p(t) < v_{pb} \quad \text{and} \quad v_x(t) < v_{xb}, \tag{3.2}$$

with $v_{pb} = v_{xb} \sim 10^{-5}$ and $v_b = v_{pb}v_{xb} \sim 10^{-10}$.

The fast part of the kinetics is approximately described by

$$v(\tau) \approx v(0)\exp(4\lambda\tau), \tag{3.3}$$

where $\tau = t - t_i$, λ is the Lyapunov exponent (see Section 2) and $v(0) \sim 10^{-13}$ is the minimal variance averaged over all n fluctuations observed in the run. We note that the latter value is considerably smaller than the border value $v_b \sim 10^{-10}$. This is because of the penetration of trajectories into the FD. Interestingly, the ratio $v_b/v(0) = 2000$ is the same for both runs in Fig. 1.

A surprisingly sharp crossover to the diffusive kinetics, clearly seen in Fig. 1, is related to the dynamical scale of the diffusion corresponding to a certain size v_d of the increasing variance at which the exponential growth stops. Roughly, it occurs at the time instance

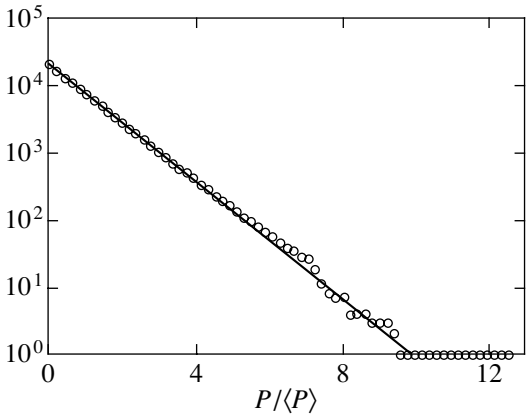

Fig. 3. The histogram of integrated distribution (3.9) for data in Fig. 2. Each circle shows the number of periods $P_m > m\Delta P$, for $m = 0, 1, \ldots$ $P_0 = n$, $\Delta P = 1.5 \times 10^5$; $P_{\min}/w = 1.0027$; $P_{\max}/\langle P\rangle = 12.63$; $\langle P\rangle = 765084$. The straight line is the expected distribution $n\exp(-P/\langle P\rangle)$.

$\tau = \tau_d$, when $|x - x_0| \sim |p - p_0| \sim 1/2$, whence $\nu_{xd} \sim 12/4 = 3$ and $\nu_{pd} \sim 3/C^2$. We can therefore characterize the dynamical scale as

$$\nu(\tau_d) = \nu_d = F_d \nu_{pd} \nu_{xd} = \frac{9F_d}{C^2}, \qquad \tau_d = \frac{\ln(\nu_d/\nu(0))}{4\lambda}, \tag{3.4}$$

where F_d is an empirical factor and τ_d is found from Eq. (3.3). The data in Fig. 1 imply the dynamical scale $\nu_d \approx 0.015$ independently of ν_b, which gives the empirical factor $F_d \approx 1/3$.

In the diffusion region ($\nu > \nu_d$), the initial kinetics is described by a simple relation for the free diffusion (see Section 2),

$$\nu(\tau) \approx \frac{\tau \pm \tau_d}{C^2} + \nu_d, \quad \tau_d < \tau \ll C^2, \tag{3.5}$$

which is also shown in Fig. 1. It involves two corrections, τ_d and ν_d, due to the exponential ballistic kinetics. The first one (with opposite signs for the two symmetric parts of the fluctuation) takes the "lost" time after (or prior to) the antidiffusion (diffusion) into account, while the second correction describes a finite fluctuation size at the crossover from (to) the diffusion. The mean empirical value $\tau_d = 7$ used in Fig. 1 is close to the value $\tau_d = 6.5$ found from Eq. (3.4) with another empirical quantity, $\nu_d = 0.015$.

The large ratio

$$B = \frac{\langle P\rangle}{C^2} \gg 1 \tag{3.6}$$

of the mean fluctuation period $\langle P\rangle$ to the characteristic time of the diffusion relaxation (see Eq. (3.5)) is the definition of a big fluctuation. It guarantees the time separation of successive fluctuations.

We now turn to the main subject of our study, the purely diffusive kinetics of BFs. For this, we first eliminate the x-statistics by excluding ν_x from selection condition (3.1), which now reads

$$\nu(t) = \nu_p < \nu_{pb} = \nu_b. \tag{3.7}$$

Next, the variance ν_b must now exceed the new dynamical border,

$$\nu_b > \nu_d = \nu_{pd} \approx f_p \frac{12}{C^2} \tag{3.8}$$

with some empirical factor $f_p \approx 1$ (see Eq. (3.4) and the discussion below).

A typical example of a diffusive BF is shown in Fig. 2. Both the regular macroscopic kinetics of the antidiffusion/diffusion and the irregular fluctuations around are clearly seen. We note that their size rapidly decreases toward the BF maximum. It may even seem that the motion becomes regular in that region, hence the term "optimal fluctuational path" [14]. In fact, the motion remains diffusive down to the dynamical scale $\nu \sim \nu_d$ in Eq. (3.8).

Even though a separate BF is sufficiently regular, the time instance of its spontaneous appearance t_i and, hence, the individual period P are random in the chaotic system. Due to the statistical independence of BFs under condition (3.6), the expected distribution in P is Poissonian (Fig. 3),

$$f(P) = \frac{\exp(-P/\langle P\rangle)}{\langle P\rangle}. \tag{3.9}$$

The principal characteristic of the period statistics, $\langle P\rangle$, can be estimated as follows. From the ergodicity of motion in the N-dimensional momentum space, we have

$$\Phi = \frac{T_s}{t_f} = \frac{\langle T_s\rangle}{\langle P\rangle} = \frac{\mathcal{P}_{fl}}{\mathcal{P}_{eq}}. \tag{3.10}$$

This is an exact relation (in the limit as $t_{\rm run} \longrightarrow \infty$), with T_s being the total sojourn time of trajectories within the FD (under the condition $\nu(t) < \nu_b$) during the entire run time $t_{\rm run}$ and $\langle T_s\rangle$ the same per fluctuation. Both ratios are equal to the ratio of the N-dimensional momentum volume $\mathcal{P}$ of the fluctuation at $\tau = 0$ to that in the equilibrium. The ratio Φ was also measured during the run. It follows that

$$\langle P\rangle = \frac{\langle T_s\rangle}{\Phi}. \tag{3.11}$$

The next, more difficult step is the valuation of $T_s = 2T_{\rm ex}$ from the diffusion equation, where $T_{\rm ex}$ is the exit (or entrance due to symmetry) time from (or to) the FD.

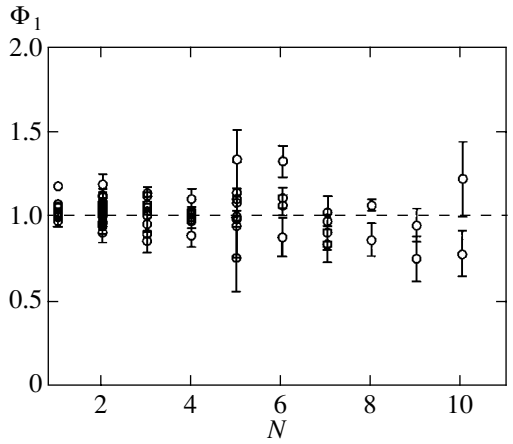

Fig. 4. The comparison of the directly measured ratio Φ_{emp} given by Eq. (3.10) with the theoretical approximation Φ_{th}, Eq. (3.12) for $N = 1$–10: $\Phi_1 = \Phi_{emp}/\Phi_{th}$; the average over 71 runs is $\langle\Phi_1\rangle = 1.015 \pm 0.11$ (the standard deviation); the bars show statistical errors $1/\sqrt{n}$ for each run; the total number of fluctuations in all runs is 127346.

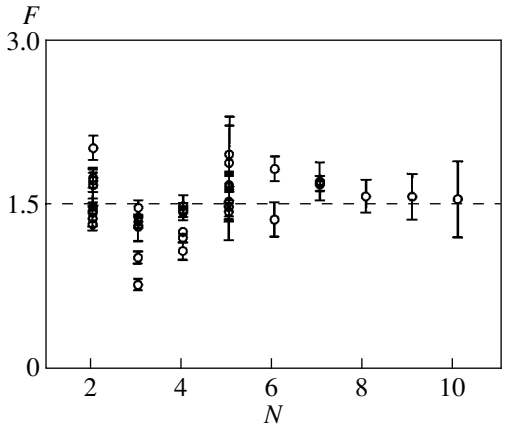

Fig. 5. The comparison of the empirical data for 36 runs selected from 61 runs computed for $N = 2$–10 by the two rules, Eq. (3.6) with $B > 7$ and Eq. (3.15) with $A = 6$, to theoretical relation (3.14) with the main fitting factor F_m, $m = 1, \ldots, 36$ (see text). The average value is $\langle F\rangle = 1.51(1 \pm 0.17)$ (the standard deviation); the bars show statistical errors $F_m/\sqrt{n}$ for each run; the total number of fluctuations in 36 runs is 34429.

A simple crude estimate is $T_{ex} \sim \nu_b/D_p = \nu_b C^2$ (see Section 2). However, the first numerical experiments have already revealed that the actual exit time is much shorter, roughly by the factor $1/N^2$. A plausible explanation is that inside the FD, the distribution is concentrated in a relatively narrow layer at the surface of the N-dimensional sphere determined by the selection condition $\nu(t) < \nu_b$ in Eq. (3.7). The relative width of the layer $\sim 1/N$ then implies the observed factor $\sim 1/N^2$. Further, the ratio

$$\Phi(\nu_b, N) = \nu_b^{N/2}\phi(N), \tag{3.12}$$

with the geometrical function

$$\phi(N) \approx \left(\frac{\pi e}{6}\right)^{N/2}\frac{(1 - 1/6N)}{\sqrt{\pi N}}, \tag{3.13}$$

admits a relatively accurate approximation down to $N = 1$ (see Fig. 4).

Collecting all the above formulas, we arrive at our final empirical relation

$$\langle P\rangle \approx \frac{F}{\Phi}\frac{2\nu_b AC^2}{N^2} \approx F\frac{2AC^2}{N^2}\frac{\nu_b^{1-N/2}}{\phi(N)} \tag{3.14}$$

with two fitting factors, A for the layer width and F for all the other approximations made above. The two factors cannot be united in one because the former enters a new expression for the dynamical scale that naturally generalizes Eq. (3.8). Together with inequality (3.6) for a big fluctuation, the new dynamical scale was used in selecting purely diffusive BFs described by Eq. (3.14). The corresponding inequality reads (cf. Eq. (3.8))

$$\nu_b > \nu_d, \quad \nu_d\frac{A}{N^2} \approx f_p\frac{12}{C^2}, \tag{3.15}$$

which means that even a small part ($A/N^2 < 1$) of the FD must exceed the dynamical scale.

All the empirical parameters were optimized as follows. The values of two factors, B in Eq. (3.6) and f_p in (3.15), are not crucial; larger values of these factors correspond to a better selection of purely diffusive BF but reduce the amount of the empirical data available. A compromise was found at $B = 7$ and $f_p = 1$, which leaves 36 runs of 61 done and 34429 of the total 75053 BFs computed with $N = 2$–10 for comparison to Eq. (3.14). This was executed as follows. For each selected run with the parameters N, C, and ν_b and the computed values $\langle P\rangle$ and Φ, the empirical factor F (which was assumed to be a constant) was calculated from the first equation in (3.14). The value of A was chosen by minimizing the relative standard deviation to $\Delta F/\langle F\rangle = 0.17$. For a given set of data, the result was $A \approx 6$. The final dependence $F(N)$ is shown in Fig. 5, where the bars are the statistical errors $F/\sqrt{n}$ for each run.

Coming to the analysis of our main theoretical result, the second equation in (3.14), we first remark that it does not describe a single trajectory ($N = 1$). This is because we excluded ν_{xb} from selection condition (3.7) (cf. Eq. (3.2)) and thus reduced the phase space dimension to the minimal value, unity. In this case, a single trajectory repeatedly crosses the FD with the period $P \sim C^2$, the entire diffusion time around the phase space torus, which is independent of the FD size.

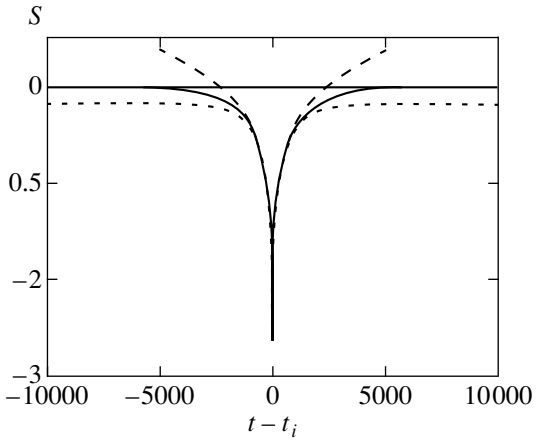

Fig. 6. The macroscopic kinetics of the BF entropy: the lower line is the "exact" entropy given by Eq. (2.6), to be compared with approximation (2.7), the middle line; the upper line is the same approximation for the diffusion theory, Eq. (3.5) with $\tau_d = 0$ and the empirical value $\nu_d^{(\mathrm{emp})} = 0.02$. The run parameters/results are $C = 50$, $N = 5$, $\nu_b = 0.01$, $w = 10^4$, $n = 4580$, $\langle P\rangle \approx 3.3 \times 10^6$, $B = 1314$; $\langle P\rangle/w \approx 329$. The number of partition bins for calculating (2.6) is $N_p = 401$.

More formally, this also follows from Eq. (3.14), because condition (3.6) cannot be satisfied for small ν_b.

For two trajectories ($N = 2$), the period does not depend on ν_b, and for the data in Fig. 5, we have the ratio $\langle P\rangle/C^2 \approx 8.7$. Because of fluctuations, the actual values of this ratio are in the interval 7.4–11.0, still not too big for a BF. Apparently, this leads to a relatively large scattering of points with $N = 2$, which also persists for $N = 3$.

The main dependence in Eq. (3.14), the exponential of N, is readily derived from a graphic picture of N statistically independent particles gathering together inside a small domain with the probability $\sim 1/P \sim \nu_b^{N/2}$. Such estimates are known for the Poincaré recurrences since Boltzmann [10]. The estimate is especially vivid in the geometrical picture of the N-dimensional sphere of the radius $\sqrt{\nu_b}$ considered above. Our empirical relation (3.14) considerably improves the simple estimate by including a weaker power-law dependence, which is evident in Fig. 5.

In our studies described above, we fixed the position of a BF in phase space, Eq. (2.4). If we lift this restriction, the probability of a BF increases by the factor $\nu_b^{-1/2}$, which corresponds to decreasing N by one ($N \longrightarrow N - 1$) because only $N - 1$ trajectories then remain independent. With the latter change, all the above relations presumably remain valid.

Our main relation (3.14) describes the diffusive kinetics for $\nu_b > \nu_d$, Eq. (3.15), when a BF is not too big. In the opposite case $\nu_b \ll \nu_d$ of a very big fluctuation, as in Fig. 1, the dependence $\langle P(\nu_b)\rangle$ becomes much simpler (see Eqs. (3.11)–(3.13) and [16]):

$$\langle P(\nu_b)\rangle = \frac{\langle T_s\rangle}{\Phi} \approx \frac{2}{\nu_b^{N/2}\phi(N)} \approx 2\nu_b^{-N/2}. \qquad (3.16)$$

This is explained by a fast exponential kinetics near the BF top (Fig. 1), which implies the shortest exit time $T_{\mathrm{ex}} \approx 1$, and hence, $T_s \approx 2$. Indeed, for both BFs in Fig. 1, we have the empirical value $\langle P\rangle\Phi = 1.98$.

In the conclusion of this section, we show in Fig. 6 the macroscopic kinetics of the BF entropy, both the "exact" one in Eq. (2.6), calculated for the partition of the entire interval ($0 < p < C$) into $N_p = 401$ bins, and the one given by our approximation (2.7). Both entropies were calculated for the same 5 trajectories in one run. The necessary statistics for the exact entropy was obtained at the expense of a large number $n = 4580$ of fluctuations in the run. To compare the two entropies, we must adjust the constant S_0 in Eq. (2.6). As is easily verified, Gaussian distribution (2.8) leads exactly to relation (2.7) if

$$S_0 = -\frac{1}{2}\ln(2\pi e) \approx -1.4189 \approx -\sqrt{2}. \qquad (3.17)$$

Approximation (2.7) is valid for the most part of the BF except a relatively small domain near the equilibrium, where the distribution in p approaches the homogeneous one. The exact entropy (with constant (3.17)) in the equilibrium is

$$S_{SE} = -\frac{1}{2}\ln\left(\frac{\pi e}{6}\right) \approx -0.18 \qquad (3.18)$$

instead of zero in approximation (2.7). The difference is relatively small, the larger the fluctuation. In the main part of the BF, our simple relation for the entropy in Eq. (2.7) reproduces exact relation (2.6) to a surprisingly good accuracy. This confirms that the distribution in p is indeed very close to the Gaussian one in Eq. (2.8), as expected.

4. CONCLUSION: THERMODYNAMIC ARROW?

We have presented the results of extensive numerical experiments on big entropy fluctuations (BFs) in a statistical equilibrium (SE) of classical dynamical systems and discussed their peculiarities.

All numerical experiments were carried out on the basis of a very simple model given by Arnold cat map (2.1) on a unit torus with only two minor, but important and helpful, modifications:

(1) expanding the torus in the p direction, Eq. (2.2), for a more impressive diffusive kinetics of BFs out of the equilibrium (Fig. 2), and

(2) inserting a special (unstable) fixed point for a better demonstration of the exponential ballistic kinet-

ics (Fig. 1). In addition, this point was used as a fixed position of BFs, which relates our studies of BFs to another interesting and important problem, the Poincaré recurrences (see Eq. (2.2)).

The most important distinction of our approach is that we have abandoned the vague question of the initial conditions, in particular, a "necessary" restriction of those in statistical physics. Instead, we started our numerical experiments at arbitrary initial conditions (most likely corresponding to the SE), and did observe the dynamics and statistics of BFs. In other words, we studied the spontaneous BFs only.

It is also important that such a spontaneous rise of a BF out of the SE and its subsequent relaxation back to the SE can be considered as a statistical macroscopic event, even in a system with a few degrees of freedom as the one in Eq. (2.2). The term "macroscopic" refers to average quantities including variance, entropy, mean period, distribution function, and the like.

We consider a particular class of BFs that we call the Boltzmann fluctuations. They are obviously symmetric under the time reversal (see Figs. 1, 2, and 6), and therefore, at least in this case, there is no physical reason at all for the concept of the notorious "time arrow." Nevertheless, a related concept—the *thermodynamic* arrow pointing in the direction of the average increase of entropy—makes sense in spite of the time symmetry [16]. The point is that the BF characteristic relaxation time is determined by the model parameter C only and does not depend on the BF itself. On the contrary, the expectation time for a given BF, or the mean period between successive fluctuations, rapidly grows with the BF size and with the number of trajectories (or the degrees of freedom), Eq. (3.14). A large ratio of the two quantities, $B = \langle P \rangle / C^2 \gg 1$, is our definition of a big fluctuation, Eq. (3.6). A similar result was recently obtained in [22], but the authors missed the principal difference between the time arrow and the thermodynamic arrow.

A related notion of the causality arrow, which by definition points from an independent macroscopic cause to its effect, also makes some physical sense (see [16] and Section 3 for a discussion). For the Boltzmann BFs considered in the present paper, the directions of both arrows coincide independently of the direction of time. In our opinion, the last statement is the most important, philosophical "moral" that the principally well-known Boltzmann fluctuations teach us.

Even though we discuss and interpret our empirical results in terms of entropy (S), which is the most fundamental concept in statistical physics, we actually use another entropy-like quantity, the variance $V(t)$ for a group of N trajectories, Eq. (2.5). One reason is technical: the computation of V is much simpler than that of $S(t)$, which is either very time-consuming in numerical experiments (for exact S given by (2.6)) or approximate in accordance with (2.7). In addition, for diffusive kinetics, in which we are mainly interested, the variance is a natural variable that makes the BF picture most simple and comprehensible.

Originally, we planned to cover both sides of the BF phenomenon, the regular macroscopic kinetics and the accompanying microscopic fluctuations (noise) around. However, our numerical experiments revealed a much more complicated structure of the latter, as an example in Fig. 2 demonstrates. The dependence $V(t)$ looks like a fractal curve on a variety of time scales, ranging from the minimal one ~1 iteration up to $\sim C^2$, which is comparable to that of the BF itself. This interesting problem certainly requires and deserves further studies.

Only the fluctuations in classical mechanics are considered in this paper. General quantum fluctuations are quite different. However, according to the Correspondence Principle, the dynamics and statistics of a quantum system in the semiclassical region are close to the classical ones at the appropriate time scales, the longest of which corresponds to the diffusive kinetics and ensures the transition to the classical limit (see [4, 23] for details). Curiously, the computer classical dynamics that is the simulation of a classical dynamical system on digital computer is of a qualitatively similar character. This is because any quantity is discrete ("overquantized") in the computer representation. As a result, the correspondence between the classical continuous dynamics and its computer representation in numerical experiments is generally restricted to certain finite time scales as in quantum mechanics (see the first two references in [23]).

The discreteness of the computer phase space leads to another peculiar phenomenon: generally, the computer dynamics is irreversible due to the rounding-off operation unless a special algorithm is used in numerical experiments. However, this does not affect the statistical properties of the chaotic computer dynamics. In particular, the statistical laws remain time-reversible in the computer representation in spite of the (nondissipative) irreversibility of the underlying dynamics. This simple example demonstrates that contrary to a common belief, the statistical reversibility is a more general property than the dynamical reversibility.

REFERENCES

1. J. Lebowitz, I. Prigogine, and D. Ruelle, Physica A (Amsterdam) **263**, 516 (1999).
2. J. Lebowitz, Physica A (Amsterdam) **194**, 1 (1993).
3. L. D. Landau and E. M. Lifshitz, *Statistical Physics* (Nauka, Moscow, 1995; Pergamon, Oxford, 1980), Part 1.
4. B. V. Chirikov, in *Law and Prediction in the Light of Chaos Research*, Ed. by P. Weingartner and G. Schurz (Springer-Verlag, Berlin, 1996), p. 10; Open Systems and Information Dynamics **4**, 241 (1997); chao-dyn/9705003 (1997); Wiss. Z. Humboldt-Univ. Berl., Ges.-Sprachwiss. Reihe **24**, 215 (1975).
5. J. Jeans, Philos. Trans. R. Soc. London, Ser. A **199**, 1 (1929).

6. H. Haken, *Synergetics: an Introduction* (Springer-Verlag, Berlin, 1978; Mir, Moscow, 1980).
7. A. Turing, Philos. Trans. R. Soc. London, Ser. B **237**, 37 (1952); G. Nicolis and I. Prigogine, *Self-Organization in Non-Equilibrium Systems* (Wiley, New York, 1977; Mir, Moscow, 1979).
8. A. Cottrell, in *The Encyclopedia of Ignorance*, Ed. by R. Duncan and M. Weston-Smith (Pergamon, Oxford, 1977), p. 129.
9. P. Glansdorf and I. Prigogine, *Thermodynamic Theory of Structure, Stability, and Fluctuations* (Wiley, New York, 1971; Mir, Moscow, 1972).
10. L. Boltzmann, *Vorlesungen über Gastheorie* (J. A. Barth, Leipzig, 1896-98; Gostekhizdat, Moscow, 1956); *Lectures on Gas Theory* (Univ. of California Press, Berkeley, 1964).
11. E. Schrödinger, *Über die Umkehrung der Naturgesetze* (Sitzungsber. Preuss. Akad. Wiss., 1931), p. 144.
12. A. N. Kolmogoroff, Math. Ann. **112,** 155 (1936); **113**, 766 (1937).
13. A. M. Yaglom, Dokl. Akad. Nauk SSSR **56**, 347 (1947); Mat. Sb. **24**, 457 (1949).
14. D. G. Luchinsky, P. McKlintock, and M. I. Dykman, Rep. Prog. Phys. **61**, 889 (1998).
15. I. P. Kornfel'd, Ya. G. Sinaĭ, and S. V. Fomin, *The Theory of Ergodicity* (Nauka, Moscow, 1980).
16. B. V. Chirikov, Zh. Éksp. Teor. Fiz. **119**, 205 (2001) [JETP **92**, 179 (2001)].
17. D. V. Anosov, Dokl. Akad. Nauk SSSR **145**, 707 (1962).
18. L. Schulman, Phys. Rev. Lett. **83**, 5419 (1999); G. Casati, B. V. Chirikov, and O. V. Zhirov, Phys. Rev. Lett. **85**, 896 (2000); D. Evans, D. Searles, and E. Mittag, cond-mat/0008421 (2000); B. V. Chirikov and O. V. Zhirov, cond-mat/0009125 (2000).
19. V. I. Arnold and A. Avez, *Ergodic Problems of Classical Mechanics* (Benjamin, New York, 1968; Izhevsk, 1999).
20. A. J. Lichtenberg and M. A. Lieberman, *Regular and Stochastic Dynamics* (Springer-Verlag, New York, 1992).
21. R. Balescu, *Equilibrium and Nonequilibrium Statistical Mechanics* (Wiley, New York, 1975; Mir, Moscow, 1978).
22. R. Metzler, W. Kinzel, and I. Kanter, cond-mat/0007382 (2000).
23. B. V. Chirikov, F. M. Izrailev, and D. L. Shepelyansky, Sov. Sci. Rev., Sect. C **2**, 209 (1981); B. V. Chirikov, in *Lectures in Les Houches Summer School on Chaos and Quantum Physics, 1989* (Elsevier, Amsterdam, 1991), p. 443; G. Casati and B. V. Chirikov, in *Quantum Chaos: Between Order and Disorder*, Ed. by G. Casati and B. V. Chirikov (Cambridge Univ. Press, Cambridge, 1995), p. 3; Physica D (Amsterdam) **86**, 220 (1995); B. V. Chirikov, in *Proceedings of the International Conference on Nonlinear Dynamics, Chaotic and Complex Systems*, Zakopane, 1995, Ed. by E. Infeld, R. Zelazny, and A. Galkowski (Cambridge Univ. Press, Cambridge, 1997), p. 149; B. V. Chirikov and F. Vivaldi, Physica D (Amsterdam) **129**, 223 (1999).

Journal of Experimental and Theoretical Physics, Vol. 93, No. 1, 2001, pp. 24–32.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 1, 2001, pp. 31–40.

ATOMS, SPECTRA, RADIATION

Laser Synthesis of Chiral Molecules in Isotropic Racemic Media

S. S. Bychkov, B. A. Grishanin*, and V. N. Zadkov

Department of Physics, International Laser Center, Lomonosov Moscow State University, Moscow, 119899 Russia
**e-mail: Grishan@comsim1.ilc.msu.su*

Received November 27, 2000

Abstract—The problem of directional laser synthesis of enantiomers in an isotropic racemic mixture of chiral molecules is analyzed taking into account rotational degrees of freedom of molecules. It is shown that the chirality of the polarization structure of a laser field is the most general necessary condition for the laser distillation of an isotropic non-racemic mixture of chiral molecules with the isotropic distribution over the Eulerian angles. In the approximation of the electric-dipole interaction, the required field configuration can be provided due to noncoplanarity of the polarization vectors of laser-pump components. A relevant experimental scheme is proposed and calculated for the transformation of an isotropic racemic mixture with the help of a three-component pulsed laser field. It is shown that the possibility of the laser control of chirality in an isotropic medium corresponds to nonzero information on coupling between the input and output in the laser field-chiral molecular state information channel. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The problem of directional enrichment of a racemic mixture of left-oriented (*L*-enantiomers) and right-oriented (*D*-enantiomers) chiral molecules with the help of laser radiation is being recently actively discussed. At present, two approaches exist to the solution of this problem. One of them consists in the selection of enantiomers of one type without changing their nuclear configuration (the scenario of laser selection of enantiomers in a racemic mixture). This mechanism was proposed in [1], where the mixture deracemization caused by a circularly polarized laser beam with a nonuniform intensity profile over the beam cross section was demonstrated.

Another approach is based on photoinduced synthesis of enantiomers of one type from enantiomers of the other type (laser synthesis of enantiomers in a racemic mixture). Synthesis of enantiomers, or the so-called scenario of laser distillation of enantiomers in a racemic mixture, is obviously possible only in the case if the photoinduced dynamics of the *L*-enantiomer differs from that of the *D*-enantiomer. Several models of distillation of a racemic mixture have been proposed [2–5]. In particular, its was suggested to produce deracemization of the initially racemic mixture of enantiomers by irradiating it by circularly polarized light [2, 4]. However, the estimates of the percent ratio of enantiomers in this case give an unsatisfactory result, the difference being only 10^{-6}%. In this respect, a method for coherent control of chirality proposed by Shapiro *et al.* [5] seems to be more efficient. However, a theoretical analysis performed in [5] can be applied only to media with oriented molecules and requires a qualitative modification in the case of isotropic media (in gas, liquid, or amorphous states). Note also that the coherent control of chirality is efficient only in the case of small chiral molecules with the mass $M \lesssim 10^3 m_H$, where m_H is the proton mass. For example, hydrogen peroxide vapor, solutions of alkaloid molecules, etc. can be used as active media. In the case of more complex enantiomers, the redistribution of the energy of laser pulses over many degrees of freedom becomes substantial and coherent control of chirality is strongly hindered.

In this paper, we study the role of the isotropic distribution of molecules over rotational degrees of freedom in the process of laser distillation of enantiomers from a racemic mixture of chiral molecules for any possible distillation scenarios (Section 2). We showed that, in the case of the isotropic distribution of chiral molecules over orientation Eulerian angles in a racemic mixture, the scenario of coherent synthesis of enantiomers suggested in [5] could not be successful. In Sections 3 and 5, we suggest and analyze a modified scenario of laser distillation of enantiomers from a racemic mixture of chiral molecules, which allows the efficient synthesis of enantiomers of a specified symmetry. The qualitative interpretation of the described criterion of physical implementation of schemes of directional laser synthesis in the terms of the information theory is presented in Section 4. In the Conclusion, the main results of the paper are presented.

1063-7761/01/9301-0024$21.00

2. ANALYSIS OF SCENARIOS OF LASER SYNTHESIS OF ENANTIOMERS FROM A RACEMIC MIXTURE TAKING INTO ACCOUNT THE ROTATIONAL SYMMETRY OF MOLECULES

We assume that the ground electronic state of the nuclear configuration of a stable enantiomer[1] depends on one reaction coordinate θ and Eulerian angles $\mathbf{O} = (\varphi, \vartheta, \tilde{\theta})$, which characterize the rotation of a molecule as a whole. Below, we assume for simplicity that the states of all other intramolecular degrees of freedom are fixed, so that the variation of these states is ignored in calculating the molecular dynamics. In the case of stable molecules, a two-well potential of chiral molecules has a high barrier, so that the tunneling splitting is virtually absent and the eigenstates $|L\rangle = \psi_L(\theta)$ and $|D\rangle = \psi_D(\theta)$ of the Hamiltonian in the ground electronic state are presented with equal probabilities in a thermodynamically equilibrium statistical racemic mixture of *L*- and *D*-enantiomers corresponding to these states. The presence of a high potential barrier in practically interesting cases of large molecules is caused by a large mass *M* of the enantiomer. In this case, the rotation of the enantiomer as a whole is classical, because $\hbar\omega_{n,n+1} \ll kT$, where $\omega_{n,n+1}$ are frequencies of rotational transitions. The free dynamics of such an enantiomer is described in terms of states which depend on the only reaction coordinate θ (i.e., $|L\rangle \longrightarrow \psi_L(\theta)$, $|D\rangle \longrightarrow \psi_D(\theta)$), while the role of Eulerian angles $\mathbf{O}$ is reduced to the classical averaging over the initial distribution, which does not change during laser irradiation.

Even this simplest model of the free dynamics of an isotropic medium reveals the fundamental role of rotational symmetry. Its consideration shows that it is impossible to realize coherent control of chirality in an isotropic racemic mixture for the polarization configuration of the fields, in which all the polarization vectors lie in the same plane; i.e., they are coplanar.

The main quantum-mechanical variable that determines the excess of *L*-enantiomers in the mixture, i.e., the degree of its deviation from racemism, is the chirality operator $\hat{\chi} = |L\rangle\langle L| - |D\rangle\langle D|$. The $|L\rangle$ and $|D\rangle$ states are the eigenvectors of the operator $\hat{\chi}$, and its eigenvalues are ± 1. In the matrix representation in the basis of the states $|L\rangle$ and $|D\rangle$, the chirality operator is represented by the Pauli matrix $\hat{\sigma}_z$. The general algorithms of the construction of the chirality operator in the coordinate representation can be found in papers [6, 7], according to which the operator $\hat{\chi}$ is a pseudoscalar, i.e., $\hat{R}^{-1}\hat{\chi}\hat{R} = -\hat{\chi}$, where $\hat{R}$ is the inversion operator corresponding to the reflection $\mathbf{r} \longrightarrow -\mathbf{r}$ of the radius vector.

The inversion operator $\hat{R}$ in the subspace of the eigenfunctions $|L\rangle$ and $|D\rangle$ of the ground electronic state is described by relations $\hat{R}|L\rangle = |D\rangle$ and $\hat{R}|D\rangle = |L\rangle$. The excess of *L*-enantiomers, or the degree of chirality at fixed Eulerian angles $\mathbf{O}$, is expressed in terms of the chirality operator $\hat{\chi}$, the transition superoperator $\mathcal{S}_I(\mathbf{O})$, which describes the transformation of the density matrix $\hat{\rho}_0 \longrightarrow \hat{\rho}_t$ for the time *t*, and the operator of free time evolution

$$\hat{U}_0(\tau) = \exp\left[-\frac{i}{\hbar}\hat{H}_0\tau\right],$$

where $\hat{H}_0$ is the molecular Hamiltonian. The superoperator nature of the transformation of the density matrix is caused by the presence, in the general case, of decay processes. For an ensemble of molecules distributed over the orientation angles, the degree of chirality includes an additional averaging over the Eulerian angles:

$$\chi = \langle \mathrm{Tr}[\hat{\chi}\hat{U}_0\mathcal{S}_I(\mathbf{O})\hat{\rho}_0\hat{U}_0^{-1}]\rangle_{\mathbf{O}}. \quad (1)$$

Below, we assume in this averaging that the angular distribution is isotropic.

The transition superoperator, which describes the photoinduced dynamics of a chiral molecule taking into account relaxation processes over the reaction coordinate and neglecting the rotation of the molecule as a whole, has the form

$$\mathcal{S}_I(\mathbf{O}) = T\exp\left\{\int_0^t\left(-\frac{i}{\hbar}[\hat{U}_0^{-1}\hat{H}_I\hat{U}_0, \odot] + \mathcal{L}_r\right)d\tau\right\}, \quad (2)$$

where *T* is the time-ordering operator; the expression of type $[\hat{A}, \odot]$ with the substitution symbol $\odot$ describes the superoperator of commutation of the transformed density matrix with the operator $\hat{A}$; $\hat{H}_I = -\mathbf{E}\hat{\mathbf{d}} - \mathbf{E}\hat{\boldsymbol{\mu}}$ is the interaction Hamiltonian of the molecule with a multicomponent laser field,

$$\mathbf{E} = \mathrm{Re}\sum_k \mathcal{E}_k(t)e^{-i\omega_k t};$$

$\hat{\mathbf{d}}$ is the dipole moment operator of the nuclear configuration of the molecule; $\hat{\boldsymbol{\mu}}$ is the electronic dipole moment operator; and $\mathcal{L}_r$ is the relaxation Liouvillian, which takes into account relaxation processes in the molecule [8].

In the superoperator representation, i.e., as applied to the operators of quantum-mechanical quantities, the inversion a molecule is performed by the superoperator

[1] We consider an enantiomer stable if the time of tunneling between its different chiral states is much longer than at least the time of the experiment.

$\mathcal{R}^{-1} = \hat{R}^{-1} \odot \hat{R}$, while the inversion of density matrices is performed by the superoperator $\mathcal{R} = \hat{R} \odot \hat{R}^{-1}$. For a racemic mixture of initially unexcited molecules, we have

$$\hat{\rho}_0 = \frac{1}{2}(|L\rangle\langle L| + |D\rangle\langle D|),$$

and, taking into account the symmetry of the molecular Hamiltonian with respect to inversion, we have $\mathcal{R}^{-1}\hat{H}_0 = \hat{H}_0$ (in the model under study, which ignores weak interactions, the Hamiltonian $\hat{H}_0$ is represented by a scalar). Then, it follows from the identity

$$-\chi = \langle \mathrm{Tr}[(\mathcal{R}^{-1}\hat{\chi})\hat{U}_0\mathcal{S}_I\hat{\rho}_0\hat{U}_0^{-1}]\rangle_{\mathbf{O}}$$

and taking into account the symmetry of the initial state ($\mathcal{R}\hat{\rho}_0 = \hat{\rho}_0$), that

$$-\chi = \langle \mathrm{Tr}[\hat{\chi}\hat{U}_0(\mathcal{R}\mathcal{S}_I\mathcal{R}^{-1}\hat{\rho}_0)\hat{U}_0^{-1}]\rangle_{\mathbf{O}}, \qquad (3)$$

where

$$\mathcal{R}\mathcal{S}_I\mathcal{R}^{-1} = T$$

$$\times \exp\left\{\int_0^t\left(-\frac{i}{\hbar}[\hat{U}_0^{-1}(\mathcal{R}^{-1}\hat{H}_I)\hat{U}_0, \odot] + \mathcal{R}\mathcal{L}_r\mathcal{R}^{-1}\right)d\tau\right\}.$$

Using expression (3), we show that the effect of laser distillation of enantiomers from a racemic isotropic mixture completely vanishes when a multicomponent laser field is not chiral; i.e., the polarization vectors of the laser-field components are coplanar. The information meaning of this qualitative result is explained in Section 4.

The inversion of the interaction Hamiltonian $\hat{H}_I$ in (3) is equivalent to the inversion of the dipole moment operator $\mathcal{R}^{-1}\hat{\mathbf{d}} = -\hat{\mathbf{d}}$, which in turn is equivalent to the inversion of the vector of the electromagnetic field strength **E**. Because the relaxation Liouvillian $\mathcal{L}_r$ is invariant with respect to the inversion $\mathcal{R} \odot \mathcal{R}^{-1}$, it follows directly from (3) that the degree of chirality χ changes its sign upon inversion of the polarization configuration of the multicomponent laser field. The inversion operation can be decomposed into the mirror reflection relative to an arbitrary plane and the corresponding rotation. If the polarization configuration of the laser field is coplanar, then the inversion operation is reduced to a simple rotation because the polarization configuration of the field in this case is invariant with respect to the mirror reflection relative to the plane in which all the polarization vectors of the laser field lie. When the distribution of molecules is isotropic, the degree of chirality χ, according to (1), is also invariant with respect to any rotation of the polarization configuration of the field. It follows from all the above statements that, when the laser field is not chiral, the degree of chirality χ is invariant with respect to inversion. In this case, it follows from (3) that $\chi = -\chi$, i.e., $\chi = 0$.

Thus, the analysis presented above shows that in the case of the coplanar polarization configuration of a laser field and the isotropic distribution over Eulerian angles, the $D \longrightarrow L$ and $L \longrightarrow D$ transitions are always mutually compensated and the racemic nature of the initial state is retained.

In the scenario of laser distillation of enantiomers from an isotropic racemic mixture, the symmetry of the $L \longrightarrow D$ and $D \longrightarrow L$ transitions, which appears upon averaging over isotropically distributed Eulerian angles, can be violated by two methods. The first method can be based on the additional orientation of molecules over Eulerian angles by the laser field. This scheme was considered in papers [9, 10] as applied to hydrogen dioxide vapor. However, in the case of heavier molecules, such an orientation can be achieved only at very high laser-field strengths, approaching the intra-atomic field strength. The second method considered in Section 3 can be based, in accordance with the above analysis, on the use of a multicomponent laser field with the noncoplanar polarization configuration, which acts on the intramolecular degrees of freedom of enantiomers. The generalization of the scenario of laser distillation [5] under study is based on the use of three linearly polarized subpicosecond laser pulses whose polarization vectors are noncoplanar. The noncoplanar polarization vectors should be used because only such a system of vectors possesses chirality, which is inherent only in the three-dimensional space and is absent in the one-dimensional and two-dimensional cases. Such a field contains information on chirality even in the case of the local interaction with a molecule, i.e., in the dipole approximation. The circularly polarized field in this approximation (i.e., considered in a single point) does not reveal its chirality, which is related to the propagation direction of the field and is manifested only upon nonlocal interaction.

3. SCENARIO OF LASER SYNTHESIS OF ENANTIOMERS IN ISOTROPIC RACEMIC MEDIA

In the scheme of coherent control of chirality in the racemic mixture of chiral molecules [5], the following model of the free dynamics of molecules is adopted. The potential over the reaction coordinate θ corresponding to the excited electronic state is quasi-harmonic, with a minimum at $\theta = 0$. The first states of the nuclear configuration of the excited electronic level $|1\rangle$ and $|2\rangle$ are described by the wave functions that are symmetric and antisymmetric relative to $\theta = 0$, respectively. The frequency of the transition from the $|1\rangle$ state to the $|2\rangle$ state lies in the IR region. It was shown, ignoring the limitations considered in Section 2, which are specific for isotropic media, that the scenario of laser

distillation of enantiomers from a racemic mixture of chiral molecules can be realized using two subpicosecond pulses (Fig. 1a). The subpicosecond laser pulse $\mathscr{E}_2$ with frequency ω_2 is used for the preparation of the coherent superposition of the states $|1\rangle$ and $|2\rangle$. Under such a condition, the asymmetric transfer of populations of the *L* and *D* states can be induced via the states $|1\rangle$ and $|2\rangle$ by the subpicosecond laser pulse $\mathscr{E}_1$ with frequency ω_1. Because the polarization vectors of the pulsed laser field are always coplanar in this scheme, the expected effect of the transformation of the racemic mixture in an isotropic medium will completely vanish, according to Section 2.

Consider the generalization of the scheme of laser distillation [5] to the isotropic racemic mixture of chiral molecules using the same free dynamics model (Fig. 1b). The coherent superposition of the states $|1\rangle$ and $|2\rangle$ is induced by the biharmonic pump with linearly polarized amplitudes $\mathscr{E}_2(t)$ and $\mathscr{E}_3(t)$ of the pump components. The pumping of the excited electronic level by the linearly polarized pulse $\mathscr{E}_1(t)$ results in the violation of the symmetry of transfer of the *L* and *D* states via the upper levels $|1\rangle$ and $|2\rangle$. The polarization vectors of the pulses are chosen so that the polarization configuration of the multicomponent laser field

$$\mathbf{E}(t) = \{\mathscr{E}_1(t), \mathscr{E}_2(t), \mathscr{E}_3(t)\}$$

is noncoplanar (see Fig. 2).

In deriving the interaction Hamiltonian $\hat{\hat{H}}_I$ in the rotating-wave approximation for a four-level system shown in Fig. 1b, we take into account the fact that the transition between vibronic states also includes the dynamics of a nuclear subsystem; i.e., this transition is not a Franck–Condon transition. By denoting the states of the nuclear configuration of the molecule by ket vectors $|L\rangle$, $|D\rangle$, $|1\rangle$, and $|2\rangle$ (which depend, in the coordinate representation, only on the reaction coordinate θ), and the states of the electronic subsystem by $|g\rangle$ and $|e\rangle$, we determine the corresponding Rabi frequencies and the relations between them for fixed Eulerian angles. The Rabi frequencies of vibronic transitions are

$$\Omega_1 = \frac{1}{\hbar}\langle 1|\langle e|(\mathscr{E}_1\cdot\hat{\boldsymbol{\mu}})|g\rangle|L\rangle = \frac{1}{\hbar}\langle 1|\langle e|(\mathscr{E}_1\cdot\hat{\boldsymbol{\mu}})|g\rangle|D\rangle,$$

$$\frac{1}{\hbar}\langle 1|\langle e|(\mathscr{E}_1\cdot\hat{\boldsymbol{\mu}})|g\rangle|D\rangle \sim \frac{1}{\hbar}\langle 2|\langle e|(\mathscr{E}_1\cdot\hat{\boldsymbol{\mu}})|g\rangle|D\rangle,$$

$$\frac{1}{\hbar}\langle 2|\langle e|(\mathscr{E}_1\cdot\hat{\boldsymbol{\mu}})|g\rangle|D\rangle = -\frac{1}{\hbar}\langle 2|\langle e|(\mathscr{E}_1\cdot\hat{\boldsymbol{\mu}})|g\rangle|L\rangle.$$

The Rabi frequency Ω_2 for a biharmonic pump of the transition between vibrational states $|1\rangle$ and $|2\rangle$ has the form [11]

$$\Omega_2 \propto \frac{1}{\hbar^2}\sum_p \frac{(\mathscr{E}_2\cdot\langle 1|\hat{\mathbf{d}}|p\rangle)(\langle p|\hat{\mathbf{d}}|2\rangle\cdot\mathscr{E}_3)}{\omega_2-\omega_{1p}},$$

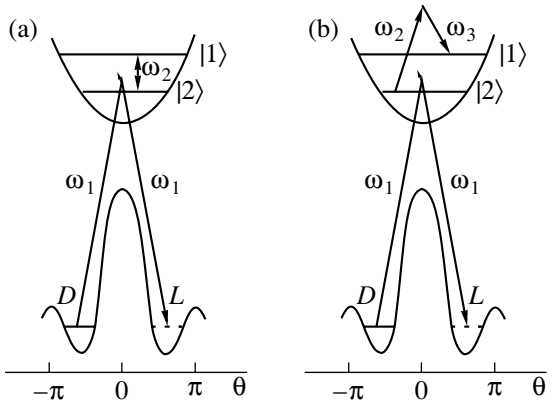

Fig. 1. (a) Schematic of laser distillation of a racemic mixture of stable enantiomers using two laser pulses at frequencies ω_1 and ω_2 [5]. (b) Schematic of laser distillation of an isotropic racemic mixture using three laser pulses at frequencies ω_1, ω_2, and ω_3 with noncoplanar polarization vectors. The variable θ describes the reaction coordinate responsible for the chiral state of a molecule in the ground electronic state.

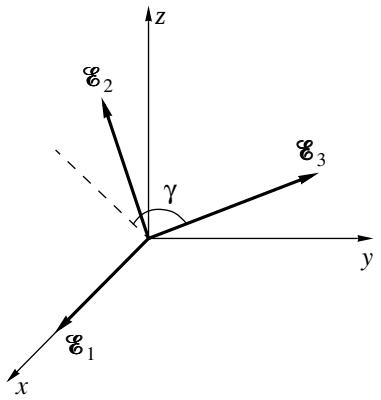

Fig. 2. Polarization configuration of a multicomponent laser field: γ is the angle between the field component $\mathscr{E}_3$ at frequency ω_3 and the field projection on the plane formed by components $\mathscr{E}_1$ and $\mathscr{E}_2$ at frequencies ω_1 and ω_2, respectively.

where $|p\rangle$ are the intermediate vibrational states of the upper electronic term.

The dependence of Rabi frequencies on Eulerian angles is manifested in the matrix elements of the corresponding dipole operators. It is important to note that after averaging over the isotropically distributed Eulerian angles, the expected distillation of the racemic mixture completely vanishes for any polarization configuration of the laser field if the geometric configuration of matrix elements $\langle g, L|\hat{\boldsymbol{\mu}}|1, e\rangle$, $\langle 1|\hat{\mathbf{d}}|p\rangle$, and $\langle p|\hat{\mathbf{d}}|2\rangle$ of the transition dipole moments is coplanar. This statement can be proved in the same way as the

necessity of using the noncoplanar configuration of the laser field in scenarios of laser distillation of an isotropic racemic mixture (see Section 2). The only difference is that in this case, all the operations of reflection and rotations are directly applied to the dipole moments of molecules. Therefore, the method of distillation of a racemic mixture under study can be applied only to molecules with the noncoplanar geometric configuration of the transition dipole moments.

The modified Hamiltonian $\hat{H}_0$ of the free dynamics of molecules in the rotating-wave approximation and the interaction Hamiltonian $\hat{H}_I$ expressed in the terms of the state vectors $|L\rangle$, $|D\rangle$, $|1\rangle$, and $|2\rangle$ have the form

$$\begin{aligned}\hat{H}_0 &= \hbar\omega_1(|1\rangle\langle 1| + |2\rangle\langle 2|),\\ \hat{H}_i &= \hbar\Omega_1\sum_{k=1}^{2}[|k\rangle\langle L| + |L\rangle\langle k|\\ &+ (-1)^{k-1}(|k\rangle\langle D| + |D\rangle\langle k|)]\cos\omega_1 t\\ &+ \hbar\Omega_2\sum_{k\neq l}|k\rangle\langle l|\cos\omega_{12}t + \hbar(\Delta_1|1\rangle\langle 1| - \Delta_2|2\rangle\langle 2|),\end{aligned} \tag{4}$$

where ω_1 is the frequency of the laser pulse $\mathscr{E}_1$, ω_{12} is the frequency of the transition between vibrational levels $|1\rangle$ and $|2\rangle$ of the upper electronic term, which exactly coincides with the Raman frequency of the biharmonic pump. In the rotating-wave approximation, the Hamiltonian $\hat{H}_I(t)$ in the interaction representation $\hat{\tilde{H}}_I = \hat{U}_0^{-1}\hat{H}_I\hat{U}_0$ is averaged in time and is time-independent for the given choice of the unperturbed Hamiltonian $\hat{H}_0$. As a result, we obtain from the second relation in (4) in the matrix representation in the basis of the states $|L\rangle$, $|D\rangle$, $|1\rangle$, and $|2\rangle$

$$\hat{\tilde{H}}_I = \frac{\hbar}{2}\begin{pmatrix} 0 & 0 & \Omega_1 & -\Omega_1\\ 0 & 0 & \Omega_1 & \Omega_1\\ \Omega_1 & \Omega_1 & 2\Delta_1 & \Omega_2\\ -\Omega_1 & \Omega_1 & \Omega_2 & -2\Delta_2\end{pmatrix},$$

where Δ_1 and Δ_2 are the detunings from the resonances at the $|L\rangle|g\rangle \longrightarrow |1\rangle|e\rangle$ and $|L\rangle|g\rangle \longrightarrow |2\rangle|e\rangle$ transitions, respectively. Let us represent now the transition superoperator (2) as an analytic function of the Hamiltonian $\hat{\tilde{H}}_I$ and restrict ourselves to the consideration of rectangular laser pulses of duration τ_p, for which the parameters of the interaction Hamiltonian are independent of time. We will also neglect relaxation processes, assuming that the pump pulses are substantially shorter than the characteristic relaxation time. In this case, the transition superoperator (2) has the form

$$\mathscr{S}_I(\mathbf{O}) = \exp\left(-\frac{i}{\hbar}[\hat{\tilde{H}}_I, \odot]\tau_p\right). \tag{5}$$

In the final calculations, we used the following relations: $\Omega_2 \ll \omega_{12}$, Δ_1 and $\Omega_1 \sim \omega_{12}$. The first relation determines the maximum admissible order of magnitude of the field strengths $\mathscr{E}_2$ and $\mathscr{E}_3$ in the case of the resonance approximation, while the second one provides the possibility of coherent control of chirality determined by the field $\mathscr{E}_1$.

Let us represent the Hamiltonian $\hat{H}_1$ as a sum of the Hamiltonian $\hat{H}_1$, which determines the one-photon transfer of populations of the states L and D via states $|1\rangle$ and $|2\rangle$, and the Hamiltonian $\hat{H}_2$, which determines the induction of the coherent superposition of the states $|1\rangle$ and $|2\rangle$. Because in the model of photoinduced dynamics under study, the racemism of the mixture is disturbed by the simultaneous action of all the three pulses, we will retain in expansion (5) the terms containing both $\hat{H}_1$ and $\hat{H}_2$. Moreover, taking into account that $\hat{H}_1 \gg \hat{H}_2$, we will retain in the first-order approximation in (5) only the terms that linearly depend on $\hat{H}_2$. The corresponding analytic expression for the effective transition superoperator has the form

$$\begin{aligned}\mathscr{S}_I(\mathbf{O}) &\approx \exp\left\{-i\frac{\tau_p}{\hbar}[\hat{H}_1, \odot]\right\}\\ &- 2i\frac{\sin\left\{\frac{\tau_p}{\hbar}[\hat{H}_1, \odot]\right\}}{[\hat{H}_{I=1}, \odot]}[\hat{H}_2, \odot],\end{aligned} \tag{6}$$

where

$$\hat{H}_1 = \frac{\hbar}{2}\begin{pmatrix} 0 & 0 & \Omega_1 & -\Omega_1\\ 0 & 0 & \Omega_1 & \Omega_1\\ \Omega_1 & \Omega_1 & 2\Delta_1 & 0\\ -\Omega_1 & \Omega_1 & 0 & -2\Delta_2\end{pmatrix},$$

$$\hat{H}_2 = \frac{\hbar}{2}\begin{pmatrix} 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & \Omega_2\\ 0 & 0 & \Omega_2 & 0\end{pmatrix}.$$

To elucidate the role of the field noncoplanarity, we consider the expansion of expression (6) in powers of the Hamiltonian $\hat{H}_1$ by omitting the expansion terms

that contain only either Hamiltonian $\hat{H}_1$ or $\hat{H}_2$. These terms do not describe Raman transitions induced by all the pulses applied (see Fig. 1b) and, therefore, do not affect the process of laser distillation of the racemic mixture. Taking this circumstance into account, the required expansion has the form

$$\mathcal{S}_I(\mathbf{O}) \longrightarrow \sum_{n=2}^{\infty}\left(-\frac{i\tau_p}{\hbar}\right)^n \times([\hat{H}_2, [\hat{H}_1 \ldots [\hat{H}_1, \odot]]] + \ldots + [\hat{H}_1, [\hat{H}_1 \ldots [\hat{H}_2, \odot]]]). \quad (7)$$

By substituting (7) into (1), we finally obtain the degree of chirality

$$\chi \sim \sum_{n=2}^{\infty}\tau_p^n \sum_{i=0}^{n} b_{in}\Delta_1^i \langle \Omega_1^{n-1-i}\Omega_2\rangle_{\mathbf{O}}, \quad (8)$$

where τ_p is the pulse duration and b_{in} are the dimensionless coefficients, which are independent of the parameters of the laser field and a molecule.

It is important to note that it follows from (8) that the degree of chirality linearly depends on parameters of the biharmonic pump of the $|1\rangle \longrightarrow |2\rangle$ transition and, hence, is zero in the absence of the coherent superposition of states $|1\rangle$ and $|2\rangle$, i.e., for $\hat{H}_2 \longrightarrow 0$. One can easily obtain from (8) the important dependence of the degree of chirality on the parameter characterizing the noncoplanarity of the polarization configuration of a multicomponent laser field

$$\mathbf{E} = \{\mathcal{E}_1(t), \mathcal{E}_2(t), \mathcal{E}_3(t)\}.$$

In this case, the specific form of coefficients b_{in} is not important, although it can be obtained analytically using the methods of computer algebra. The parameter of interest to us is, for example, the angle γ between a plane formed by the vectors $\mathcal{E}_1(t)$, $\mathcal{E}_2(t)$, and $\mathcal{E}_3(t)$ (Fig. 2). The analysis of coherent laser synthesis of a nonracemic mixture in the case of the isotropic distribution of molecules over Eulerian angles performed in Section 2 showed that the degree of chirality changes its sign upon mirror reflection of the polarization configuration of a laser field with respect to an arbitrary plane. For the method of transformation of a racemic mixture under study, this means that upon the mirror reflection $\gamma \longrightarrow -\gamma$ of the polarization configuration of a laser field, we have $\chi \longrightarrow -\chi$; i.e., it follows from the general considerations that the degree of chirality χ is an odd function of γ. To obtain the exact dependence of the degree of chirality on the angle γ, we consider the expression $\langle \Omega_1^k \Omega_2\rangle_{\mathbf{O}}$ for arbitrary values of k. Let, for definiteness, the vector $\mathcal{E}_1(t)$ be directed along the x axis of the laboratory coordinate system. Then, for odd $k \geq 1$ (in the case of even k, the corresponding terms are set to zero), after averaging the expressions containing the powers of trigonometric functions over the Eulerian angles, we obtain

$$\langle \Omega_1^k \Omega_2\rangle_{\mathbf{O}} \propto \epsilon_{x\alpha\beta} e_{2\alpha} e_{3\beta}, \quad (9)$$

where $\alpha, \beta = x, y, z$; $\epsilon_{x\alpha\beta}$ is the antisymmetric Levi–Civita tensor; and $\mathbf{e}_2$ and $\mathbf{e}_3$ are the unit polarization vectors of the biharmonic-pump components.

Taking into account (9), we obtain from (8) that the degree of chirality is proportional to the mixed product of the unit polarization vectors of a multicomponent field and, hence, $\chi \sim \sin\gamma$. The dependence of the degree of chirality on the other parameters (the frequency detuning Δ_1 and laser-pulse duration τ_p) for the model of photoinduced dynamics under study can be analyzed only partially. Figure 3 shows the dependence $\chi(\Delta_1, \tau_p)$ that was numerically calculated for the case of mutually orthogonal polarization vectors $\mathcal{E}_1$, $\mathcal{E}_2$, and $\mathcal{E}_3$.

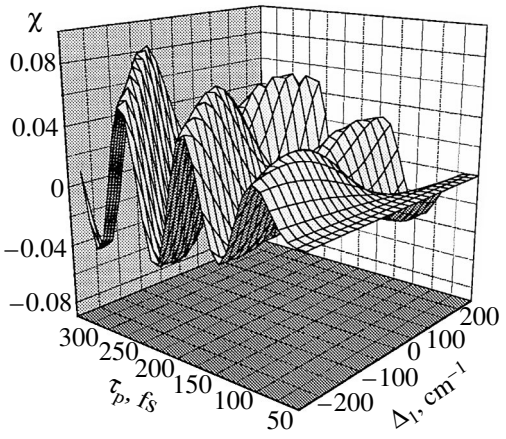

Fig. 3. Dependence of the degree of chirality χ on the frequency detuning Δ_1 of the laser field $\mathcal{E}_1$ and the pulse duration τ_p for the case of mutually orthogonal polarization vectors $\mathcal{E}_1(t)$, $\mathcal{E}_2(t)$, and $\mathcal{E}_3(t)$ of laser pulses. The following parameters were used in calculations: $\Omega_1 = \omega_{12} = 100\ \text{cm}^{-1}$, $\Omega_2 = 10\ \text{cm}^{-1}$.

4. CRITERION FOR CONTROL OF THE CHIRAL STATE OF AN ISOTROPIC MEDIUM

To gain a better qualitative insight into the meaning of the results obtained, it is expedient to consider the experiment in the most general terms, ignoring the details of the laser excitation scheme and the structure of molecules under study. At this general level, the system under study represents an information channel [12] with a controlling laser field at its input. The output of the channel is the chiral state of a molecule after laser excitation, which is averaged over the initial density matrix and orientation angles and is reduced over eigenstates $|L\rangle$ and $|D\rangle$ of the chiral variable $\hat{\chi}$. The latter became a purely classical quantity after this reduc-

tion and has values of ±1 that correspond to states $|L\rangle$ and $|D\rangle$. Along with states $|L\rangle$ and $|D\rangle$, which describe only the ground electronic state and are of interest to us, we should also introduce the states in the form of the third state $|e\rangle$ that sums up the former states. Although these additional states are of no interest to us in themselves, they provide a complete set of possible states. In our case, the state $|e\rangle$ corresponds to the excited electronic states, which provide the conservation of the total probability in the process of laser excitation. The result of calculations of the excitation process in the most general form is described by the conditional ("transient") distribution of the channel probabilities

$$\mathcal{P}(\hat{\chi}|\mathbf{E}) = \{p(L|\mathbf{E}), p(D|\mathbf{E}), p(E|\mathbf{E})\}, \qquad (10)$$

where $\mathbf{E} = \{\mathcal{E}_1, \mathcal{E}_2, \ldots\}$ is the set of fields represented in the interaction Hamiltonian and e denotes excited electronic states. Probabilities (10) in fact enter into expression (1), which describes the averaging of the chirality operator. Assuming that the initial density matrix

$$\hat{\rho} = \rho_L|L\rangle\langle L| + \rho_D|D\rangle\langle D|$$

is diagonal, these probabilities are expressed in terms of the matrix elements of the evolution superoperator as

$$\begin{aligned} p(L|\mathbf{E}) &= S_{LL}(\mathbf{E})\rho_L + S_{LD}(\mathbf{E})\rho_D, \\ p(D|\mathbf{E}) &= S_{DD}(\mathbf{E})\rho_D + S_{DL}(\mathbf{E})\rho_L, \\ p(e|\mathbf{E}) &= 1 - p(L|\mathbf{E}) - p(D|\mathbf{E}), \end{aligned} \qquad (11)$$

where

$$S_{\alpha\beta} = \mathrm{Tr}|\alpha\rangle\langle\alpha|(\mathcal{S}_I|\beta\rangle\langle\beta|) = \langle\alpha|(\mathcal{S}_I|\beta\rangle\langle\beta|)|\alpha\rangle$$

for $\alpha, \beta = L, D$. The matrix elements $S_{\alpha\beta}$ describe the conditional probabilities of chiral states at the input for the fixed initial chiral states of a molecule, over which the averaging is performed for the channel under study with the laser field at its input.

The information on the laser field for the channel represented by distribution (10) is contained not only in the degree of chirality of the ground electronic state but also in the total probability of preparing the molecule in this state, because a fraction of molecules undergo transitions to the excited electronic state upon laser excitation. Obviously, it is interesting to reduce the input information only to the relative distribution of molecules in the ground state. By describing the state of the molecule immediately after termination of laser pulses, we can do this by restricting ourselves to the consideration of the relative distribution of probabilities for the ground state only, which is specified by conditional probabilities

$$\begin{aligned} p(L|\mathbf{E}) &= \frac{S_{LL}(\mathbf{E})\rho_L + S_{LD}(\mathbf{E})\rho_D}{S_{LL}(\mathbf{E})\rho_L + S_{LD}(\mathbf{E})\rho_D + S_{DD}(\mathbf{E})\rho_D + S_{DL}(\mathbf{E})\rho_\lambda}, \\ p(D|\mathbf{E}) &= \frac{S_{DD}(\mathbf{E})\rho_D + S_{DL}(\mathbf{E})\rho_L}{S_{LL}(\mathbf{E})\rho_L + S_{LD}(\mathbf{E})\rho_D + S_{DD}(\mathbf{E})\rho_D + S_{DL}(\mathbf{E})\rho_\lambda}. \end{aligned} \qquad (12)$$

Another situation of interest is related to the consideration of the distribution in the ground electronic state, which appears after the free relaxation of the excited electronic state. If we assume that the excited state is not chiral and relaxes after excitation, then the same probabilities equal to $(1/2)p(e|\mathbf{E})$ are added to each of the chiral states after the establishment of equilibrium. Taking this into account, we obtain the following equilibrium distribution:

$$\begin{aligned} p(L|\mathbf{E}) &= \frac{1}{2} + \frac{S_{LL}(\mathbf{E})\rho_L + S_{LD}(\mathbf{E})\rho_D}{2} \\ &\quad - \frac{S_{DD}(\mathbf{E})\rho_D + S_{DL}(\mathbf{E})\rho_L}{2}, \\ p(D|\mathbf{E}) &= \frac{1}{2} - \frac{S_{LL}(\mathbf{E})\rho_L + S_{LD}(\mathbf{E})\rho_D}{2} \\ &\quad + \frac{S_{DD}(\mathbf{E})\rho_D + S_{DL}(\mathbf{E})\rho_L}{2}. \end{aligned} \qquad (13)$$

The field $\mathbf{E}$ in expressions (12) and (13) is described by a probability distribution $\mathcal{P}(d\mathbf{E})$ in the space of the laser-field parameters. As applied to a series of experiments with a set of different parameters of the laser field, this distribution corresponds to the relative frequencies of the appearance of different values of $\mathbf{E}$. The distribution $\mathcal{P}(d\mathbf{E})$ reflects the nature and amount of information contained in the laser field. In accordance with the experimental conditions, we can study the particular cases of this distribution by restricting ourselves, for example, to fixed amplitudes of the field vectors and considering the distribution over their orientations. Taking into account the laser-field chirality established in Section 2, of special interest is the simplest variant of the input information, when the fields under study are distributed only between two configurations obtained upon the reflection. In this case, the laser field in distributions (12) and (13) is represented in fact only by the two-value parameter $\alpha = 1, 2$: $\mathbf{E} \longrightarrow \alpha$, which is a minimal requirement for the reflection of the chirality of the polarization configuration. Correspondingly, the conditional probability distributions are represented by the 2×2 matrices with matrix elements $p(\beta|\alpha)$, where β describes the chiral state of the mole-

cule and α describes the similar chiral variable of the laser field.

Consider the case of a racemic initial mixture ($\rho_L = \rho_D = 1/2$). In this case, relations (12) and (13) are expressed only in terms of combinations of the matrix elements of the transition superoperator $S_L = S_{LL} + S_{LD}$ and $S_D = S_{DD} + S_{DL}$, which transfer to each other upon the substitution $L \rightleftarrows D$, by realizing the representation of the inversion superoperator $\mathcal{R}$. If the transition superoperator $\mathcal{S}_I$ is invariant with respect to this transformation, then expressions (12) and (13) give the value 1/2 for all conditional probabilities; i.e., the distribution of the resulting chiral variables is independent of the chirality of the input configuration. It is obvious that the corresponding amount of information on the field chirality contained at the output vanishes at any distribution of the rest of the field parameters. Therefore, the criterion for the control of the chiral state is the presence of the information on the chiral variable of the laser-field configuration in the channel under study. This is possible only when the laser field itself contains this information. As shown in Section 2, this condition is satisfied when the polarization configuration of the laser field is chiral, its chirality being realized as the configuration noncoplanarity in the case of the electric-dipole interaction. However, this conclusion is not applied in the general form to the channel (11), in which, along with the intrinsic chirality of the ground electronic state, the degree of the state depletion after laser excitation is also taken into account.

5. DISCUSSION OF THE RESULTS

Analysis of the general relations described in Section 3 and the quantitative calculations of the degree of chirality based on this analysis can be summarized as the following qualitative results, which characterize the proposed scenario of laser distillation of enantiomers in an isotropic racemic mixture of chiral molecules.

First, according to (8), upon the use of three laser fields, the degree of chirality χ is proportional to the sine of the angle γ, $\chi \sim \sin\gamma$, and for $\gamma = 0$, it is impossible to transform a racemic mixture in the case of the isotropic distribution of chiral molecules over rotational degrees of freedom (see Section 2). It is obvious that the maximum degree of chirality is achieved when the polarization vectors of laser pulses are mutually orthogonal ($\gamma = \pi/2$).

Second, the optimum duration τ_p of linearly polarized pulses is determined by the Rabi frequency Ω_1 of the vibronic excitation pulse, the above analysis being valid under the condition $\tau_p \lesssim 2\pi/\Omega_1$. Otherwise, the approximate expansion (5) of the transition superoperator is not applicable and the dependence of χ on the parameters Δ_1 and τ_p becomes complicated (multifrequency), so that the coherent synthesis of enantiomers in a racemic mixture is hindered. For $\tau_p \lesssim 2\pi/\Omega_1$, it follows from (8) that the degree of chirality $\chi \sim \Omega_2/\Omega_1$ and, hence, it does not exceed 10%.

Third, analysis of the dependence of the degree of chirality on the detuning Δ_1 and the laser-pulse duration τ_p (Fig. 3) shows that the maximum degree of chirality $\chi_{max} \sim 8\%$ is achieved at $\Delta_1 \sim -150$ cm^{-1} and $\tau_p \sim 250$ fs. Our calculations gave the following values of the excitation-pulse intensity corresponding to the Rabi frequencies used for estimating dipole moments $\mu_{1,D}$, $d \sim ea_B$ (a_B is the Bohr radius). For the vibronic excitation pulse, we obtained $I_1 \sim 10^9$ W/cm^2; for biharmonic pump pulses, we have $I_2, I_3 \sim 10^{12}$ W/cm^2 if the frequencies of these pulses lie in the optical region, and $I_2, I_3 \sim 10^9$ W/cm^2 if their frequencies lie in the IR region.

6. CONCLUSIONS

The analysis performed in the paper has shown that in the case of the isotropic distribution of chiral molecules over the rotational degrees of freedom, coherent control of chirality is possible only when the structure of a multicomponent laser field is chiral in itself, as is the case of in fact the nonlocal magnetic-dipole [2] or quadrupole interaction of a molecule with a circularly polarized field. In the electric-dipole approximation, i.e., upon the local interaction, the molecule cannot distinguish the circular and linear polarizations, and the chirality of the field is manifested only in the case of the noncoplanar polarization configuration (Section 2). This result corresponds to the nonzero information on the chiral state of the polarization conformation of the field, which is contained in the chiral molecular state after laser excitation (Section 4). This information is also contained in the laser-field structure itself. The information corresponding to the experimental scheme under study can be used as a dimensionless characteristic of the efficiency of laser synthesis of molecules with the specified chirality. For stable enantiomers with the mass $M \lesssim 10^3 m_H$, we proposed and calculated the scheme of laser distillation of enantiomers from a racemic mixture, which can be also applied, in contrast to the scheme [5], in the case of the isotropic distribution of chiral molecules over the Eulerian angles (Sections 3 and 5). The efficiency of the scheme depends on the polarization configuration of the laser field (Section 5) and the intensities of laser pulses. It is shown that the maximum efficiency is achieved with the help of laser pulses of moderate intensity.

REFERENCES

1. J. Kucirka and A. Shekhtman, Phys. Lett. A **221**, 273 (1996).
2. J. Shao and P. Hanggi, J. Chem. Phys. **107**, 9935 (1997).
3. R. Marquardt and M. Quack, Z. Phys. D **36**, 229 (1996).
4. A. Salam and W. Meath, J. Chem. Phys. **106**, 7865 (1997).

5. M. Shapiro, E. Frishman, and P. Brumer, Phys. Rev. Lett. **84**, 1669 (2000).
6. A. Harris, R. Kamien, and T. Lubensky, cond-mat/9901174 (1999).
7. M. Osipov, B. Pickup, and D. Dunmur, Mol. Phys. **84**, 1193 (1995).
8. B. A. Grishanin, *Quantum Random Processes*, http://comsim1.ilc.msu.su/people/grishanin/index.html (2000).
9. B. A. Grishanin and V. N. Zadkov, Zh. Éksp. Teor. Fiz. **116**, 1250 (1999) [JETP **89**, 669 (1999)].
10. B. A. Grishanin, V. N. Zadkov, and S. S. Bychkov, in *Technical Digest of XIX European CARS Workshop (*Moscow, 2000).
11. Y. R. Shen, *The Principles of Nonlinear Optics* (Wiley, New York, 1984; Nauka, Noscow, 1989).
12. R. G. Gallager, *Information Theory and Reliable Communication* (Wiley, New York, 1968; Sov. Radio, Moscow, 1974).

Translated by M. Sapozhnikov

Journal of Experimental and Theoretical Physics, Vol. 93, No. 1, 2001, pp. 33–42.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 1, 2001, pp. 41–51.

ATOMS, SPECTRA, RADIATION

Excitation of a Wake Field by a Relativistic Electron Bunch in a Semi-Infinite Dielectric Waveguide

V. A. Balakirev, I. N. Onishchenko, D. Yu. Sidorenko, and G. V. Sotnikov
Kharkov Institute of Physics and Technology, National Scientific Center, 61108 Kharkov, Ukraine
e-mail: sotnikov@kipt.kharkov.ua
Received December 22, 2000

Abstract—A wake field excited by a relativistic electron bunch in a semi-infinite metal waveguide filled with a dielectric consists of the Vavilov–Cherenkov radiation, the "quenching"-wave field, and transient radiation, which interfere with each other. An exact analytic expression for the transient component of the field of a thin relativistic annular bunch is derived for the first time. The evolution of the space distribution of a field excited by a finite-size electron bunch is numerically calculated. The excitation of the wake field by a periodic train of electron bunches in a finite-length waveguide is studied. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The propagation of a wave signal in a dispersion medium is a classical problem in many fields of physics. Such problems are encountered in acoustics [1], solid-state theory [2], plasma physics [3], and radiophysics [4, 5]. The propagation of signals in dispersion media is commonly analyzed by expanding the input-signal phase in a power series [2, 3, 5–7] or by using asymptotic methods (see [8, 9] and references cited in review [2]). The exact solution of this problem was obtained only for the case of sound [1] and electromagnetic waves in the ionosphere [10] or in a flat waveguide [11].

The above-mentioned papers considered the situation when the source of a wave signal was immobile; i.e., a quasi-monochromatic or a pulsed signal was fed at the input $z = 0$ at the instant of time $t = 0$. At the same time, the problem of propagation of electromagnetic radiation during the passage of a charged particle through the interface between two media arises in a natural way. Such problems have been considered in detail in the theory of transient radiation [12]. If the condition for the appearance of Cherenkov radiation is not satisfied in both media, the propagation of transient radiation can also be described within the framework of the formalism used for studying propagation of pulsed signals in a dispersion medium. If this condition is satisfied, transient radiation can interfere with the Vavilov–Cherenkov radiation. The transient radiation in the presence of the Vavilov–Cherenkov radiation has been studied in many papers (see, for example, references in [13]). However, the studies that have been performed so far were restricted to the calculations of total energy losses by transient and Cherenkov radiation. In the cases when the total losses could be divided into components, the Cherenkov radiation propagated into a medium where the condition of its appearance was no longer satisfied was analyzed [14].

The possibility of using Vavilov–Cherenkov radiation for accelerating charged particles in the so-called wake acceleration method [15–18] has inspired a renewed interest in this radiation. In this method, an intense bunch or a train of bunches propagating in a medium excites an electromagnetic field behind it, which then is used for the acceleration of another bunch to higher energies.

The excitation of wake fields is usually theoretically described in slow-wave media that are infinite in the direction of the bunch propagation. In this approach, a wake wave exists within the entire region behind the bunch and propagates at the phase velocity that is equal to the bunch velocity. The consideration of the finiteness of real systems can qualitatively change the space-time structure of an electromagnetic wave if only because the propagation of Cherenkov radiation in a dispersion medium (as a particular type of the wave signal [2, 6, 7, 19]) should be greatly complicated. In addition, the transient radiation appears, which will interfere with Cherenkov radiation.

In this paper, we studied excitation of an electromagnetic wave by an electron bunch and by a train of electron bunches in a semi-infinite slow-wave medium. We considered the problem in the simplest formulation. As a slow-wave medium, we considered a dielectric with $\varepsilon = \text{const}$, which filled a cylindrical waveguide with the metal walls and end. Note that a similar problem on the radiation of a single charge moving along the axis of a semi-infinite waveguide filled with a dielectric has been formulated in paper [20] where asymptotic solutions have been obtained for large times t. Below, we will show that there exists an exact solution to this problem, which is valid at any times t and for an arbitrary longitudinal coordinate z. This allows

us to analyze in detail the process of the formation of the space-time structure of the electromagnetic field excited by an electron bunch or by a train of electron bunches and to give recommendations concerning the use of our results for accelerating charge particles. The results of our paper allow us to determine the length of the localization region of the electromagnetic field excited by the electron bunch and thereby to estimate the duration of the current pulse of accelerated particles.

2. THE FIELD OF A THIN ANNULAR BUNCH

Consider a cylindrical metal waveguide of radius b that is filled with a homogeneous dielectric with the permittivity ε. The waveguide occupies the region $0 \le z < \infty$ in the longitudinal direction and is shorted with a metal wall at the end $z = 0$. We assume that an axially symmetric monoenergetic electron bunch flies into the waveguide from the wall side and then propagates at a constant velocity v_0 along the z axis. The condition $v_0 > c/\sqrt{\varepsilon}$ for the appearance of Cherenkov radiation in the waveguide is satisfied. We neglect the influence of a narrow vacuum drift channel on the electrodynamics of the system and assume for simplicity that the waveguide is completely filled with the dielectric.

To determine the field produced by an electron bunch with an arbitrary density distribution, it is necessary first to find the field of an infinitely short and thin charged ring whose plane is perpendicular to the waveguide axis. The charge density of such a ring can be written in the general case in the form

$$\rho = -\frac{eN}{2\pi r_L v_L}\delta(r - r_L)\delta(t - t_L),$$

where $-e$ is the electron charge, N is the number of electrons in the ring, $v_L(t_0, r_0, z)$ and $r_L(t_0, r_0, z)$ are the Lagrangian velocity and the radius of the annular bunch, respectively, $t_L(t_0, r_0, z)$ is the Lagrangian time of a particle, t_0 is the instant of the annular bunch flying in the waveguide, and r_0 is the initial radius of the ring. In the uniform motion approximation, the velocity and radius of the ring are constant, $v_L = v_0$ and $r_L = r_0$, and the Lagrangian time of the ring is $t_L = t_0 + z/v_0$.

The electromagnetic field excited by the charge in the semi-infinite waveguide satisfies the following boundary conditions. First, the longitudinal component of the electric field vanishes at the side conducting surface $r = b$ of the waveguide: $E_z|_{r=b} = 0$. Second, the transverse component of the electric field should vanish at the metal wall $z = 0$: $E_r|_{z=0} = 0$. Taking this into account, the expression for the Fourier component of the longitudinal component of the electric field of the axially symmetric E-wave has the form [20]

$$E_z^\omega = \frac{2iNe}{\pi b^2 \varepsilon}\sum_n \left\{\left(\frac{\varepsilon}{c^2} - \frac{1}{v_0^2}\right)\frac{\omega \exp[i\omega(z/v_0 + t_0)]}{k_{zn}^2 - \omega^2/v_0^2} - \frac{\lambda_n^2}{b^2}\frac{\exp(ik_{zn}z + i\omega t_0)}{k_{zn}v_0(k_{zn}^2 - \omega^2/v_0^2)}\right\} \times \frac{J_0(\lambda_n r_0/b)J_0(\lambda_n r/b)}{J_1^2(\lambda_n)}, \quad (1)$$

where λ_n is the nth root of the cylindrical Bessel function J_0 and

$$k_{zn}^2 = \varepsilon\omega^2/c^2 - \lambda_n^2/b^2$$

is the longitudinal wave number of free electromagnetic oscillations in the dielectric waveguide.

By performing the inverse Fourier transform, we obtain

$$E_z(t, r, z, t_0, r_0) = \frac{2iNe}{\pi b^2 \varepsilon}\sum_n \left\{I_1^n - \frac{c\lambda_n^2}{\sqrt{\varepsilon}v_0 b^2}\frac{I_2^n}{\varepsilon/c^2 - 1/v_0^2}\right\} \times \frac{J_0(\lambda_n r_0/b)J_0(\lambda_n r/b)}{J_1^2(\lambda_n)}, \quad (2)$$

$$I_1^n = \int_L d\omega\, \omega \frac{\exp[-i\omega t + i\omega(z/v_0 + t_0)]}{(\omega - \omega_0)(\omega + \omega_0)}, \quad (3)$$

$$I_2^n = \int_L d\omega \frac{\exp(-i\omega\tau + i\xi\sqrt{\omega^2 - \alpha^2})}{\sqrt{\omega^2 - \alpha^2}(\omega - \omega_0)(\omega + \omega_0)}, \quad (4)$$

where

$$\omega_0 = \frac{\lambda_n}{b\sqrt{\varepsilon/c^2 - 1/v_0^2}}, \quad \alpha = \frac{\lambda_n c}{b\sqrt{\varepsilon}},$$

$$\tau = t - t_0, \quad \xi = \frac{z\sqrt{\varepsilon}}{c}.$$

The quantity ω_0 is real, and the corresponding poles of integrands in (3) and (4) are located on the real axis ω. The contour of integration L is a straight line in the complex plane ω, which is parallel to the real axis and is located slightly above it [10].

The integral (3) describes the wake field of the charge propagating in an infinite waveguide [21]. It can be easily calculated to be

$$I_1^n = -2\pi i \cos\left[\omega_0\left(t - t_0 - \frac{z}{v_0}\right)\right]\theta\left(t - t_0 - \frac{z}{v_0}\right), \quad (5)$$

where

$$\theta(x) = \begin{cases} 1 & \text{for } x > 0 \\ 0 & \text{for } x \le 0. \end{cases}$$

The integral (4) corresponds to free eigenmodes of the cylindrical waveguide, which appear because the system is limited over z. The consideration of this additional term provides the fulfillment of the boundary condition on the side metal wall. We will show below that integral (4) gives a "quenching" wave and transient radiation. Integrals of this type are encountered in the theory of propagation of signals in dispersion media [2, 6, 7, 19]. They are calculated usually by numerical or asymptotic methods. In [20], the corresponding integral was calculated by the saddle point method, so that only an asymptotic solution was obtained, which is valid for times that are much longer than the time of the wave propagation across the waveguide. Note, however, that the field of free eigenmodes (4) can be found analytically. The analytic solution for a similar integral has been obtained for the first time in the study of the propagation of an electromagnetic signal in an ionized gas [10]. The exact solution was found later in [11] by a method that was slightly different from that used in [10].

Before proceeding to calculating integral (4), note that the function

$$k(\omega) = \sqrt{\omega^2 - \alpha^2}$$

has branch points at $\omega = \pm\alpha$. Therefore, it is necessary to make a cut along the segment $(-\alpha; \alpha)$ in the complex plane ω. Let us choose such a sheet of Riemann surface $k(\omega)$ at which

$$0 < \arg(\omega \pm \alpha) < 2\pi.$$

The signs of the real and imaginary parts of $k(\omega)$ on this sheet and, hence, of $k_{zn}(\omega)$ will be equal to the signs of the real and imaginary parts of ω, respectively. It is this condition that should be satisfied in our problem, because we consider only the waves propagating in the positive direction of the z axis, for which we should have

$$\mathrm{sgn}[\mathrm{Re}(k_{zn})] = \mathrm{sgn}[\mathrm{Re}(\omega)].$$

In this case, the condition

$$\mathrm{Im}(k_{zn}) > 0$$

will be satisfied in the upper half-plane, where the contour of integration L passes; i.e., the wave will decay at $z \longrightarrow +\infty$.

If $\tau < \xi$, we can easily close the contour of integration L of integral (4) through a semicircle of an infinite radius in the upper half-plane ω, the integral over which is zero. It is obvious that the required integral will be also zero because the resulting closed contour contains no singularities.

To calculate integral (4) for $\tau > \xi$, we perform the integration over the closed contour consisting of the initial contour L, the upper and lower banks of the cut, and a semicircle of an infinite radius in the half-plane $\mathrm{Im}\,\omega \le 0$. This contour has a clockwise direction and confines a simply connected region, which contains two poles at points $\omega = \pm\omega_0$. According to the theorem of residues, we have

$$\begin{aligned}(I_2^n + I_{cut} + I_{-\infty}) \\ = -2\pi i(\mathrm{Res}\,F(-\omega_0) + \mathrm{Res}\,F(\omega_0)),\end{aligned} \tag{6}$$

where I_{cut} is the integral over the banks of cut, $I_{-\infty}$ is the integral over the semicircle of an infinite radius in the lower half-plane (equal to zero for $\tau > \xi$), and $\mathrm{Res}\,F(\pm\omega_0)$ is the residue of the integrand in (4) at the corresponding pole $\omega = \pm\omega_0$.

To calculate the integral I_{cut} over the banks of cut, we consider a doubly connected region, whose exterior boundary is the ellipse C_ω^- and the inner boundary are the banks of cut. C_ω^- has focuses at points $\omega = \pm\alpha$ and contains poles $\omega = \pm\omega_0$. According to the theorem of residues, the integration over the boundary of this region gives

$$I_{cut} + I_{el} = -2\pi i(\mathrm{Res}\,F(-\omega_0) + \mathrm{Res}\,F(\omega_0)), \tag{7}$$

where I_{el} is the integral from the integrand in (4) over C_ω^-. By comparing (6) and (7), we can see that

$$I_2^n = I_{el}.$$

Thus, we pass from the integration over the infinite straight line L in the complex plane ω to the integration over the ellipse C_ω^- in the clockwise direction. Note that instead of an ellipse we can use any closed curve without self-intersections, which encloses all the poles and the cut, but the elliptic contour can be most simply transformed.

Let us perform a number of successive transformations in the integral I_{el}. First, we introduce the new variable of integration $p = -i\omega$. Next, we make the change of variable

$$\alpha\zeta = \sqrt{p^2 + \alpha^2} - p.$$

Then, we perform the change of variable

$$\zeta = -\beta w,$$

where

$$\beta = \sqrt{(\tau - \xi)/(\tau + \xi)}.$$

As a result, the integral I_{el} is transformed to the integral

$$I_{el} = \frac{4}{\alpha^2\beta^2}\int\limits_{C_w^+} dw \frac{w\exp[(\alpha/2)\sqrt{\tau^2 - \xi^2}(w - 1/w)]}{(w - w_1)(w - w_2)(w - w_3)(w - w_4)},$$

where

$$w_1 = \frac{i}{\beta}\sqrt{\frac{1 - c/\sqrt{\varepsilon}v_0}{1 + c/\sqrt{\varepsilon}v_0}}, \quad w_2 = \frac{i}{\beta}\sqrt{\frac{1 + c/\sqrt{\varepsilon}v_0}{1 - c/\sqrt{\varepsilon}v_0}},$$

$w_3 = -w_1$, $w_4 = -w_2$, and the contour C_w^+ is a circle in the complex plane w with a center at the point $w = 0$, which is passed in the positive direction and contains no poles $w = w_j$, $j = 1, 2, 3, 4$.

Finally, after simple transformations, we obtain

$$I_{el} = \frac{1}{2\alpha\omega_0\sqrt{\omega_0^2/\alpha^2 - 1}}\int\limits_{C_w^+} dw \exp\left[\frac{\alpha}{2}\sqrt{\tau^2 - \xi^2}\left(w - \frac{1}{w}\right)\right] \times\left\{\frac{1}{w - w_1} - \frac{1}{w - w_2} + \frac{1}{w - w_3} - \frac{1}{w - w_4}\right\}. \tag{8}$$

Because the contour of integration C_w^+ does not contain poles $w = w_j$, the expansions

$$-\frac{1}{w - w_j} = \frac{1}{w_j}\sum_{k=0}^{\infty}\left(\frac{w}{w_j}\right)^k \tag{9}$$

are valid for it.

Note also that in [22],

$$\frac{1}{2\pi i}\int\limits_{C_w^+} dw\, w^k \exp\left[\alpha\sqrt{\tau^2 - \xi^2}\frac{1}{2}\left(w - \frac{1}{w}\right)\right] = (-1)^{k+1}J_{k+1}(\alpha\sqrt{\tau^2 - \xi^2}). \tag{10}$$

By substituting series (9) into (8) and interchanging the integration and summation order, we obtain, taking into account (10), that the required integral (4) is

$$I_2^n = 2\pi i\frac{\sqrt{\varepsilon}v_0 b^2}{c\lambda_n^2}\left(\frac{\varepsilon}{c^2} - \frac{1}{v_0^2}\right)\sum_{m=1}^{\infty}(-1)^m\left(\frac{\tau - \xi}{\tau + \xi}\right)^m \times\left[\left(\frac{1 - c/\sqrt{\varepsilon}v_0}{1 + c/\sqrt{\varepsilon}v_0}\right)^m - \left(\frac{1 + c/\sqrt{\varepsilon}v_0}{1 - c/\sqrt{\varepsilon}v_0}\right)^m\right]J_{2m}(\alpha\sqrt{\tau^2 - \xi^2}). \tag{11}$$

The Lommel function $U_n(q, x)$ of two arguments is defined as [23]

$$U_n(q, x) = \sum_{m=0}^{\infty}(-1)^m\left(\frac{q}{x}\right)^{n+2m}J_{n+2m}(x). \tag{12}$$

Let us rewrite (11) by using the second-order Lommel functions of two arguments

$$I_2^n = 2\pi i\frac{\sqrt{\varepsilon}v_0 b^2}{c\lambda_n^2}\left(\frac{\varepsilon}{c^2} - \frac{1}{v_0^2}\right) \times[U_2(r_2 y, y) - U_2(r_1 y, y)], \tag{13}$$

where

$$y = \alpha\sqrt{\tau^2 - \xi^2}, \quad r_1 = \beta\sqrt{\frac{1 - c/\sqrt{\varepsilon}v_0}{1 + c/\sqrt{\varepsilon}v_0}},$$

$$r_2 = \beta\sqrt{\frac{1 + c/\sqrt{\varepsilon}v_0}{1 - c/\sqrt{\varepsilon}v_0}}.$$

Note that expression (13) gives the exact value of integral (4).

To calculate the Lommel function $U_n(q, x)$ for $q \le x$, it is convenient to use expression (12) because the series in Bessel functions in (12) very rapidly converges. For $q > x$, the expression [23]

$$U_n(q, x) = \cos\left(\frac{q}{2} + \frac{x^2}{2q} - \frac{n\pi}{2}\right) + \sum_{m=0}^{\infty}(-1)^{m+n}\left(\frac{x}{q}\right)^{-n+2+2m}J_{-n+2+2m}(x) \tag{14}$$

can be used. This series also rapidly converges for $q > x$. One can easily see that for $x = q$, expressions (12) and (14) give the same result. Therefore, we can write

$$U_2(ry, y) = \begin{cases} -\sum\limits_{m=1}^{\infty}(-1)^m r^{2m}J_{2m}(y) & \text{for } r \le 1 \\ -\cos\left(\frac{ry}{2} + \frac{y}{2r}\right) + \sum\limits_{m=0}^{\infty}\frac{(-1)^m}{r^{2m}}J_{2m}(y) & \text{for } r > 1. \end{cases} \tag{15}$$

Let us introduce the notation

$$v_{pr} = c/\sqrt{\varepsilon}, \quad v_{gr} = c^2/\varepsilon v_0.$$

For $t - t_0 - z/v_{pr} \ge 0$, the following relations are satisfied

$$0 \le r_1 < 1, \quad 0 \le r_2 \le 1 \text{ for } t - t_0 - z/v_{gr} \le 0, \quad r_2 > 1 \text{ for } t - t_0 - z/v_{gr} > 0. \tag{16}$$

Taking into account (15) and (16) for $t - t_0 - z/v_{pr} \ge 0$ and $t - t_0 - z/v_{gr} \le 0$, we can write

$$U_2(r_2 y, y) - U_2(r_1 y, y) = \sum_{m=1}^{\infty}(-1)^m(r_1^{2m} - r_2^{2m})J_{2m}(y). \tag{17}$$

Correspondingly, for $t - t_0 - z/v_{gr} > 0$, we have

$$U_2(r_2 y, y) - U_2(r_1 y, y) = -\cos\left[\omega_0\left(t - t_0 - \frac{z}{v_0}\right)\right] + J_0(y) + \sum_{m=1}^{\infty}(-1)^m\left(r_1^{2m} + \frac{1}{r_2^{2m}}\right)J_{2m}(y). \tag{18}$$

The first term in the right-hand side of (18) describes an electromagnetic wave in the region

$$0 < z < (t - t_0)v_{gr}$$

which coincides with the Vavilov–Cherenkov radiation field (5) taken with the opposite sign. This is a "quenching" wave [20], which compensates Cherenkov radiation in the corresponding region of z. The rest of the terms in (18) correspond to a fraction of the transient radiation that propagates at a velocity slower than v_{gr}. Expression (17) corresponds to a faster fraction of the transient radiation.

Let us substitute (13) and (5) into (2), taking into account (17) and (18). As a result, the field produced in a semi-infinite waveguide by a charged thin ring moving at a constant velocity can be conveniently written, similarly to [20], as a superposition of the Vavilov–Cherenkov radiation field E_z^{cher} limited in space and the transient radiation field E_z^{trans}:

$$\begin{aligned} &E_z(t, r, z, t_0, r_0) \\ &= E_z^{cher}(t, r, z, t_0, r_0) + E_z^{trans}(t, r, z, t_0, r_0), \end{aligned} \tag{19}$$

$$E_z^{cher}(t, r, z, t_0, r_0) = \frac{4Ne}{b^2\varepsilon}\sum_n \frac{J_0(\lambda_n r_0/b)J_0(\lambda_n r/b)}{J_1^2(\lambda_n)}$$

$$\times \cos\left[\omega_0\left(t - t_0 - \frac{z}{v_0}\right)\right] \tag{20}$$

$$\times \left\{\theta\left(t - t_0 - \frac{z}{v_0}\right) - \theta\left(t - t_0 - \frac{z}{v_{gr}}\right)\right\},$$

$$E_z^{trans}(t, r, z, t_0, r_0) = \frac{4Ne}{b^2\varepsilon}\sum_n \frac{J_0(\lambda_n r_0/b)J_0(\lambda_n r/b)}{J_1^2(\lambda_n)}$$

$$\times \left\{\left[\theta\left(t - t_0 - \frac{z}{v_{pr}}\right) - \theta\left(t - t_0 - \frac{z}{v_{gr}}\right)\right]\right.$$

$$\times \sum_{m=1}^{\infty}(-1)^m(r_1^{2m} - r_2^{2m})J_{2m}(y) + \theta\left(t - t_0 - \frac{z}{v_{gr}}\right)$$

$$\left.\times\left[J_0(y) + \sum_{m=1}^{\infty}(-1)^m\left(r_1^{2m} + \frac{1}{r_2^{2m}}\right)J_{2m}(y)\right]\right\}. \tag{21}$$

The Vavilov–Cherenkov wake field (20), taking into account the "quenching wave," is nonzero for

$$(t - t_0)v_{gr} \le z < (t - t_0)v_0.$$

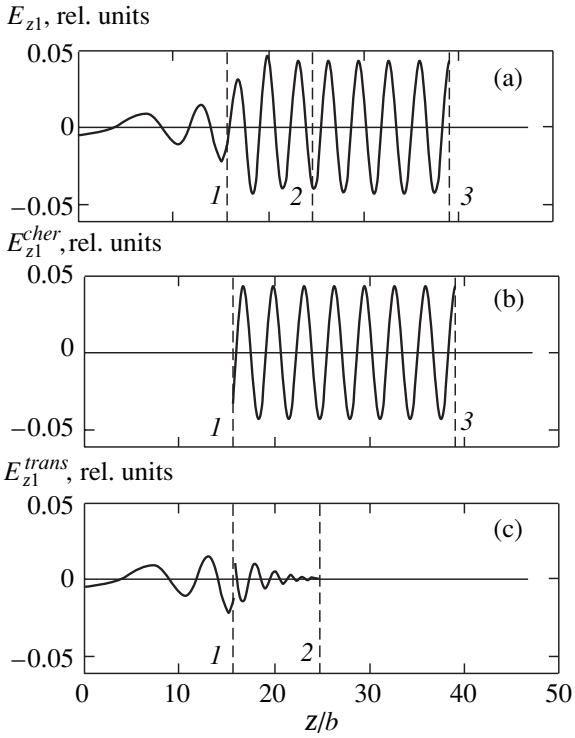

Fig. 1. Structure of the first harmonic of the longitudinal component of the wake field excited by an infinitely short charged disc in a dielectric waveguide: (a) total field E_{z1}, (b) Vavilov–Cherenkov radiation field E_{z1}^{cher}, (c) transient radiation E_{z1}^{trans}. (*1*) $z = z^{gr}$, (*2*) $z = z^{pr}$, (*3*) charge position. The observation time $tc/b = 40$, $t_0 = 0$, $\gamma = 5$, $\varepsilon = 2.6$.

Within this region, the envelope of a Cherenkov signal is constant (see Fig. 1b). The quantity v_{gr} is the group velocity of an electromagnetic wave that is synchronous with the electron bunch. The plane

$$z^{gr} = (t - t_0)v_{gr}$$

is the trailing edge of the wake field. This edge moves behind the electron bunch at the group velocity v_{gr}.

The transient radiation field (21) exists in the region

$$0 \le z < (t - t_0)v_{pr},$$

where v_{pr} is the maximum propagation velocity of electromagnetic signals in a dielectric waveguide. The fastest high-frequency part of the transient signal, the so-called "precursor," propagates precisely at this velocity. The envelope of the transient signal is maximal near the trailing edge (line *1* in Fig. 1c) and decreases away from it. For the precursor (line *2* in Fig. 1c), the envelope tends to zero. Near the rear wall, the envelope is small but nonzero and decreases with time. The transient field (21) undergoes a jump in the trailing-edge plane (see Fig. 1c). This is explained by the fact that we artificially separated the continuous total field (19) into

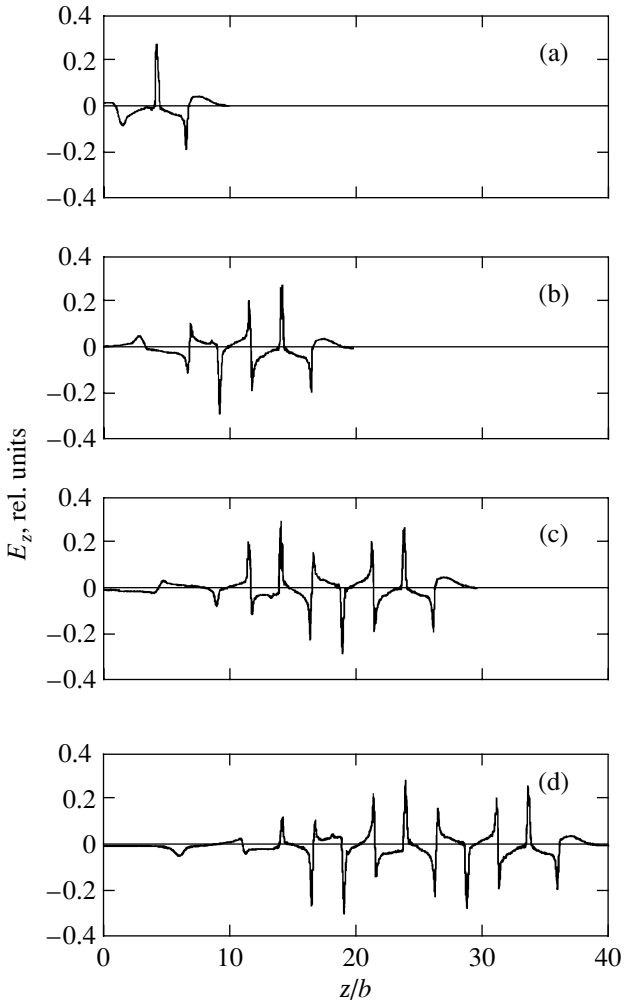

Fig. 2. Evolution of the wake field excited at the waveguide axis by a long relativistic bunch with the asymmetric density profile (22). Ten harmonics are taken into account: $tc/b = 10$ (a), 20 (b), 30 (c), 40 (d); $\gamma = 5$, $\varepsilon = 2.6$, $a/b = 0.143$, $L_b/b = 3.2$.

components. In this case, the separated Vavilov–Cherenkov field (20) also abruptly vanishes after passage through the trailing-edge plane.

As a result, the spatial structure of the field at the instant of time t has the form shown in Fig. 1a. The field is zero in front of the bunch. To the left of $z^{pr} = (t - t_0)v_{pr}$ (line *2*), the field envelope begins to decrease, while for $z = z^{gr}$ (line *1*), it is equal to half the Cherenkov radiation field. In the region $z \ll z^{gr}$, the field is weak and deceases with time.

3. RESULTS OF NUMERICAL CALCULATIONS

A charged bunch of a finite size is characterized by the following parameters: the bunch radius a, the bunch duration T_b, ($T_b = L_b/v_0$, where L_b is the bunch length; we assume that the velocity v_0 is constant and the same for all particles in the bunch), the current density distribution in the bunch $j_z(r_0, t_0)$ ($0 \le t_0 \le T_b$, $0 \le r_0 \le a$, t_0 is the time of arrival of the bunch to the system cross section under study), and a total number N_0 of charged particles in the bunch. Such a bunch can be represented as a set of rings with different charges, radii, and arrival times. The bunch field can be obtained by summing the fields produced by all macroparticles that form the bunch:

$$E_z(t, r, z) = \sum_i E_z(t, r, z, t_{0i}, r_{0i}),$$

where the elementary field $E_z(t, r, z, t_{0i}, r_{0i})$ of the ith macroparticle is determined by expressions (19)–(21).

The methods for accelerating particles by wake fields often use a profiled leading bunch, whose density gradually increases from its head to tail and then abruptly drops to zero. Such a bunch can provide a large transformation coefficient [16], i.e., the ratio of the field excited by the bunch to the strength of the field decelerating particles in the bunch. The transformation coefficient determines the maximum increase in the energy of accelerated particles. Consider, for example, the current density distribution in the form

$$j_z(r_0, t_0) = j_0 J_0\left(\lambda_1 \frac{r_0}{a}\right)\exp\left[-4\left(\frac{t_0}{T_b} - 1\right)^2\right], \qquad (22)$$

where j_0 is the maximum current density in the bunch. Such a distribution is often realized in experiments and is used in numerical calculations. In this case, the total charge and the maximum current density are related by the expression

$$eN_0 = \frac{\pi^{3/2}}{2}\frac{J_1(\lambda_1)}{\lambda_1}\frac{a^2 L_b}{v_0} j_0.$$

This expression is used for determining the characteristic scale of an electric field in numerical calculations, provided the total charge of the bunch is specified.

Figure 2 shows the time evolution of the longitudinal distribution of the wake field excited in a semi-infinite dielectric waveguide by a relativistic electron bunch, whose asymmetric longitudinal profile is described by expression (22). Note that the field behind the bunch greatly exceeds the field in the bunch, which corresponds to a large transformation coefficient. The electric field in Fig. 2 is normalized to the value $4\pi j_0 a/c$, which determines the Coulomb field near the bunch surface. Because many radial harmonics are excited in a dielectric waveguide, whose amplitudes are comparable to each other, we took into account here 10 harmonics (a further increase in their number virtually does not change the numerical results). The inter-

ference of many radial harmonics resulted in the appearance of sharp peaks of the field, whose amplitude greatly exceeds that of any harmonic. One can clearly see that the region where the envelope of the wake field is maximal is located between the charged bunch and the boundary moving behind it—the radiation trailing edge, the velocity of this edge being approximately ε times less than the bunch velocity. Behind the trailing edge, the field is substantially weaker than in front of it and has a different structure. The length of the excited wake field increases with time.

The presence of the trailing edge of the wake field, which propagates at the group velocity, restricts the length of the wave packet and imposes the corresponding requirements on the delay time during which the bunch of charged particles being accelerated should be injected to the accelerating system. The finite length of the region of existence of the excited field restricts the duration of the current pulse of accelerated particles.

A promising method for obtaining wake fields of a large amplitude is the use of a sequence of short low-density bunches, which is modulated so that the fields of individual bunches are added coherently [24]. Consider the generation of a wake field by a sequence of short relativistic bunches in a semi-infinite waveguide. To elucidate a qualitative picture, we first restrict ourselves to the first harmonic of the field. Each bunch of the sequence is simulated by a macroparticle—a thin disc. The charge repetition rate is equal to the resonance frequency of the first harmonic. The number of bunches is limited. One can see from Fig. 3a that the field in an infinite waveguide will increase linearly from the beginning of the sequence to its end, the field amplitude being constant behind the last bunch (line *1*). When the boundary $z = 0$ exists, the situation drastically changes. One can see from Fig. 3c that now the field increases from the beginning of the sequence to line *2* and then decreases, the velocity of the field decrease being increased behind the last bunch (line *1*), and the rear boundary of the wave packet appears, behind which the field tends to zero. Line *2* corresponds to the position of the trailing edge of the wake field excited by the leading bunch of the sequence and determines the number of bunches whose fields can be added coherently; i.e., the fields from all the bunches located to the right from this line are summed in the cross section shown by this line. Until the trailing edge of the field of the first charge is located within the boundaries of the sequence of a finite number of bunches (as in Figs. 3b and 3c), the maximum excited field will be less than the maximum field of the same sequence in an infinite waveguide. When the last charge of the sequence overtakes this edge (Fig. 3d), a region appears in which the fields from all bunches are coherently added and the field amplitude is the same as in an infinite waveguide. This region is bounded by lines *1* and *2* in Fig. 3d.

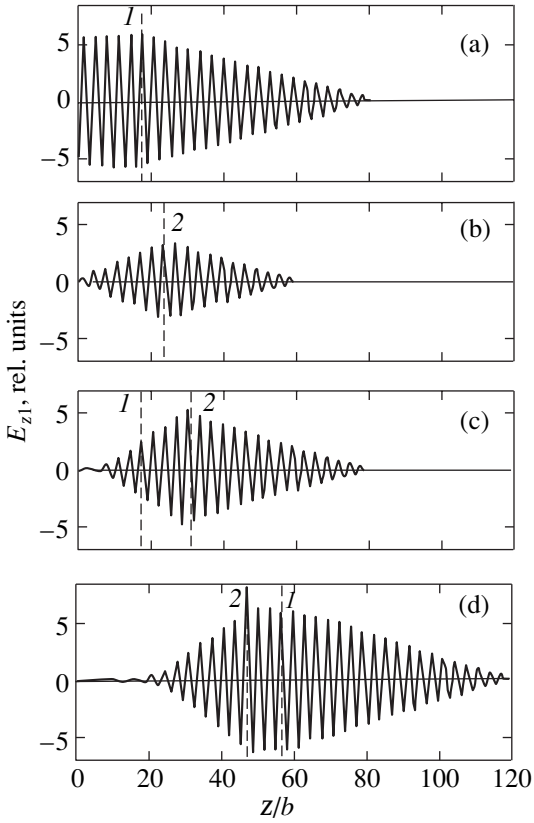

Fig. 3. (a) Structure of the first harmonic of the longitudinal electric wake field excited on the waveguide axis by a sequence of twenty charged discs ignoring the effect of the boundary z = 0; (b, c, d) is the same but for a semi-infinite waveguide: tc/b = 80 (a), 60 (b), 80 (c), 120 (d). (*1*) Coordinate of the last charge z_{20} in the sequence, (*2*) position of the trailing edge of the field produced by the first charge z_1^{gr} of the sequence; $\gamma = 5$, $\varepsilon = 2.6$, $a/b = 0.143$, $\Delta z_0/b = 3.2$.

The spatial distribution of the longitudinal electric field substantially depends on the system length. If a semi-infinite system $0 < z < +\infty$ is considered, then the field excited by a continuous sequence of electron bunches will be maximal near the trailing edge of the wake field produced by the leading bunch of the sequence:

$$z_1^{gr} = (t - t_{01})v_{gr}.$$

In this plane, the fields from the number

$$N_t^{\max} \approx \frac{t(v_0 - v_{gr})}{\Delta z_0} + 1 \qquad (23)$$

of bunches will be added, where Δz_0 is the distance between the adjacent bunches. This maximum will increase with time and propagate behind the bunch at the group velocity. The field will decrease linearly

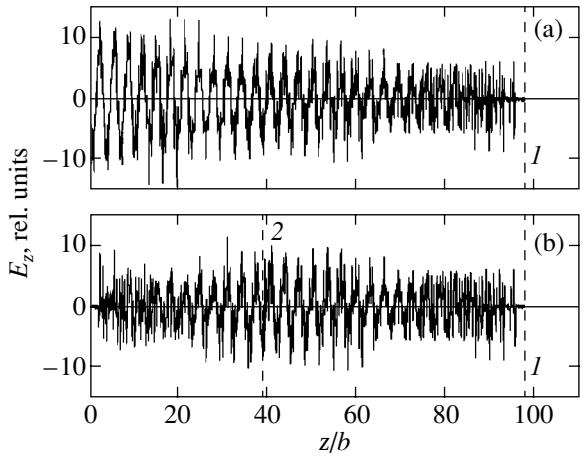

Fig. 4. (a) Structure of the longitudinal electric wake field excited on the waveguide axis by a continuous sequence of charged discs ignoring the effect of the boundary $z = 0$; (b) is the same for a semi-infinite waveguide: $tc/b = 100$. Ten harmonics were considered. (*1*) Coordinate of the first charge z_1 in the sequence, (*2*) position of the trailing edge of the field produced by the first charge z_1^{gr} of the sequence. Altogether, the figure presents 31 bunches; $\gamma = 5$, $\varepsilon = 2.6$, $a/b = 0.143$, $\Delta z_0/b = 3.2$.

behind the maximum and will not exceed the field of one bunch at the system origin ($z \approx 0$) (Fig. 3b).[1]

In a slow-wave system of finite length, a different situation can be realized. Let a waveguide of finite length L_{syst} closed by a metal wall at $z = 0$ be connected at its end, $z = L_{syst}$, to an ideally matched load. In this case, we can neglect the boundary effects at the right end of the waveguide and use expressions (19)–(21) to calculate the field inside the system. Consider the field in the region $0 < z < L_{syst}$ excited by a continuous sequence of charges. As an example, we can take Fig. 3b by assuming that $L_{syst}/b = 20$. The field envelope is maximal at the right boundary of the region $z = L_{syst}$. In this cross section, the fields of the greatest number of bunches are added. We can estimate their number from expression (23), in which we set $t = L_{syst}/v_{gr}$:

$$N_L^{\max} \approx \frac{L_{syst}}{\Delta z_0}\left(\varepsilon - 1 - \frac{\varepsilon}{\gamma^2}\right) + 1. \qquad (24)$$

For $L_{syst} = 70$ cm, $\varepsilon = 2.6$, $\gamma = 5$, $\Delta z_0 = 11.2$ cm, we obtain $N_L^{\max} \approx 10$. Figure 3b confirms this with good accuracy.

When a continuous sequence of bunches modulated at the frequency of the first resonance harmonic excites a multimode wake field in an infinite dielectric waveguide, the following events take place (Fig. 4a). At the beginning of the sequence, the fields from a few bunches are added and the total field has a complicated form because the fields produced by each charge consist of peaks whose repetition rates differ from the modulation frequency of the sequence (the region $70 < z/b < 100$ in Fig. 4a). Then, with distance from the sequence head and an increase in the number of bunches whose fields are added coherently, the resonance modulation frequency begins to be separated, and the field increases and takes a form close to that of a single-mode field (the $0 < z/b < 15$ region in Fig. 4a).

The number of bunches in a limited system, whose fields are added in some cross section, increases with distance between the cross-section plane and a metal waveguide end and reaches the limiting value (24) at the matched output end of the waveguide. For this reason, upon multimode excitation of the wake field by a sequence of bunches, the field at the waveguide input (the $0 < z/b < 20$ region in Fig. 4b near the metal wall) has approximately the same complicated form as that near the sequence head in an infinite waveguide. If the system length is sufficiently long, then oscillations near the ideally matched right end will be close to single-mode oscillations because of the separation of the resonance modulation frequency (the $30 < z/b < 40$ region in Fig. 4b, if $L_{syst}/b = 40$).

4. CONCLUSIONS

We have found the exact analytic solution for the problem of propagation of an electromagnetic signal excited by a charged bunch propagating in a semi-infinite dielectric waveguide. The obtained expressions exactly describe the field structure at any point at an arbitrary instant of time, including the situation when a particle has moved from the wall $z = 0$ by a distance that does not exceed a few wavelengths. This is important in studies of the systems whose length is comparable with their radius.

The presence of the trailing edge of the excited electromagnetic field, which propagates at the group velocity, is essential for the acceleration of a charged particle by the wake field of the charged bunch. In the model of an infinite medium, the accelerated bunch can be injected at an arbitrary period of the wave excited by the leading bunch. In a semi-infinite medium, an increase in the delay time leads either to a decrease in the acceleration rate or completely eliminates the acceleration.

When the wake field is excited by a sequence of bunches in the accelerating system of a finite length, the maximum amplitude of the longitudinal electric field is achieved at the output end of the system. The maximum field strength that can be achieved is determined by the parameters of the waveguide and sequence of charges, and is independent of the number of bunches that have flown through the system.

[1] We can use this figure to illustrate the case of a continuous sequence of bunches because the figure shows the distribution of the field at the instant of time when not all of the bunches have flown in the system.

Let us discuss the range of applicability of the solution obtained. We have used the following basic assumptions to represent the field in the analytic form (19)–(21). First, this is the linear and uniform motion of particles in the bunch parallel to the waveguide axis. Second, we assumed that the frequency dispersion of the permittivity ε is absent. Third, we assumed that the waveguide was completely filled with a dielectric that determined the electrodynamics of the moderating medium.

Having chosen the approximation of the specified linear and uniform motion of particles in the bunch, we neglected the inverse effect of the excited electromagnetic field on the particles. This approximation is often used in studies of the radiation from a single charge or low-current charged bunches [2, 20, 21]. Under real conditions, the region of applicability of this approximation imposes a restriction to the number of electrons in the bunch and the system length. The decelerating longitudinal electric field E_z, which acts on a short bunch with a rectangular density distribution, is of the order of magnitude

$$E_z \sim 8eN_0/b^2\varepsilon.$$

The condition of the neglect of variation in the particle velocity on the system length has the form

$$\frac{8e^2N_0L_{syst}}{mv_0^2\gamma^3\varepsilon b^2} \ll 1.$$

For $b = 4$ cm, $L_{syst}/b = 100$, $\varepsilon = 2.6$, and $\gamma = 5$, this inequality is satisfied for $N_0 \ll 4 \times 10^{12}$. This limiting value considerably exceeds the value achieved at present in experiments. In the wake acceleration experiments, bunches with a much higher energy γ are used, which allows us to apply the approximation of a specified velocity for dense bunches. The transverse motion of charged particles can be ignored upon the application of a strong external longitudinal magnetic field [25].

Neglecting the dependence $\varepsilon(\omega)$ is sufficiently justified and substantially simplifies the calculations. Expressions (19)–(21) describe the field with good accuracy, if the required range of the excited resonance frequencies corresponds to the horizontal part of the dependence $\varepsilon(\omega)$ where there are no absorption regions. This, as a rule, corresponds to the experimental conditions. Thus, the permittivity of polystyrene [26] is virtually constant in the frequency range from 1 to 25 GHz and is equal to 2.55, the losses being very small. The consideration of the frequency dispersion $\varepsilon(\omega)$ will change poles and branch points of integrands in expressions (3)–(4). Also, new poles can appear. In the frequency regions for which the condition $v_0 < c/\sqrt{\varepsilon(\omega)}$ is satisfied, the poles ω_0 will become purely imaginary. Correspondingly, along with the Cherenkov wave field, a quasi-static Coulomb field will emerge, which is strongly localized in the charge region. Because we have $\varepsilon \approx 1$ at $\omega \longrightarrow \infty$, the propagation velocity of the precursor of transient radiation increases: $v_{pr} \longrightarrow c$. Therefore, the field can exist in front of the bunch. However, these effects are determined by oscillations with very high frequencies, whose amplitude is small, so that they can be neglected.

The third assumption is not fundamental. The same treatment, only more cumbersome, can be performed for a waveguide with a finite thickness of the drift channel. The influence of the vacuum drift channel on the dispersion properties of the medium is negligible if its cross section is small compared to the dielectric cross section. In addition, we must keep in mind that the highest radial harmonics of the field that satisfy the condition

$$d\lambda_n \gg b\gamma\sqrt{\varepsilon v_0^2/c^2 - 1}$$

(d is the transverse size of the channel) will be weakly excited by the bunch because of a strong penetration of the field to the vacuum channel.

REFERENCES

1. A. B. Coppens, J. Acoust. Soc. Am. **25**, 525 (1953).
2. M. Elices and F. García-Moliner, in *Physical Acoustics: Principles and Methods*, Ed. by W. P. Mason (Academic, New York, 1968; Mir, Moscow, 1973), Vol. 5.
3. V. L. Ginzburg, *The Propagation of Electromagnetic Waves in Plasmas* (Fizmatgiz, Moscow, 1967; Pergamon, Oxford, 1970).
4. L. A. Zhekulin, Izv. Akad. Nauk SSSR, Otd. Tekh. Nauk **7**, 409 (1938).
5. B. N. Gershman, Zh. Tekh. Fiz. **22**, 101 (1952).
6. A. Sommerfeld, Ann. Phys. (Leipzig) **44** (10), 177 (1914).
7. L. Brillouin, Ann. Phys. (Leipzig) **44** (10), 203 (1914).
8. H. G. Baerwald, Ann. Phys. (Leipzig) **6**, 295 (1930).
9. H. G. Baerwald, Ann. Phys. (Leipzig) **7**, 731 (1930).
10. N. G. Denisov, Zh. Éksp. Teor. Fiz. **21**, 1354 (1951).
11. J. R. Wait and K. P. Spies, Appl. Sci. Res. **16**, 455 (1966).
12. V. L. Ginzburg and V. N. Tsytovich, *Transition Radiation and Transient Scattering* (Nauka, Moscow, 1984).
13. *Bibliography of Works on Transition Radiation of Charged Particles (1945–1982)*, Ed. by G. M. Garibyan (Erevans. Fiz. Inst., Yerevan, 1983).
14. G. M. Garibyan, Zh. Éksp. Teor. Fiz. **33**, 1403 (1957) [Sov. Phys. JETP **6**, 1079 (1958)].
15. P. Chen, J. M. Dawson, R. W. Huff, and T. Katsouleas, Phys. Rev. Lett. **54**, 693 (1985).
16. T. Katsouleas, Phys. Rev. A **33**, 2056 (1986).
17. W. Gai, P. Schoessow, B. Cole, *et al.*, Phys. Rev. Lett. **61**, 2756 (1988).
18. R. Keinigs, M. E. Jones, and W. Gai, Part. Accel. **24**, 223 (1989).
19. V. L. Ginzburg, *Theory of Propagation of Radio-Waves in Ionosphere* (Gostekhizdat, Moscow, 1949).

20. É. L. Burshteĭn and G. V. Voskresenskiĭ, Zh. Tekh. Fiz. **33**, 34 (1963) [Sov. Phys. JETP **8**, 22 (1963)].

21. B. M. Bolotovskiĭ, Usp. Fiz. Nauk **75**, 295 (1961) [Sov. Phys. Usp. **4**, 781 (1962)].

22. M. A. Lavrent'ev and B. V. Shabat, *Methods of the Theory of Functions of a Complex Variable* (Nauka, Moscow, 1973).

23. G. N. Watson, *Treatise on the Theory of Bessel Functions* (Cambridge Univ. Press, Cambridge, 1945; Inostrannaya Literatura, Moscow, 1949).

24. Ya. B. Faĭnberg, V. A. Balakirev, I. N. Onishchenko, *et al.*, Fiz. Plazmy **20**, 674 (1994) [Plasma Phys. Rep. **20**, 606 (1994)].

25. A. A. Rukhadze, L. S. Bogdankevich, S. E. Rosinskiĭ, and V. G. Rukhlin, *Physics of High-Current Relativistic Electron Beams* (Atomizdat, Moscow, 1980).

26. A. R. Hippel, *Dielectric Materials and Applications* (Wiley, New York, 1954; Gostekhizdat, Moscow, 1959).

Translated by M. Sapozhnikov

Journal of Experimental and Theoretical Physics, Vol. 93, No. 1, 2001, pp. 43–54.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 1, 2001, pp. 52–66.

ATOMS, SPECTRA, RADIATION

Perturbation Theory for the Intensity of Stark Lines of a Hydrogen Atom

A. A. Kamenski* and V. D. Ovsiannikov
Voronezh State University, Voronezh, 394693 Russia
**e-mail: alex@kams.vsu.ru*
Received February 8, 2001

Abstract—Perturbation theory for the wave function of a hydrogen-like atom in a homogeneous electric field of strength F makes it possible to obtain the Rayleigh–Schrödinger series with the coefficients of F^N (N = 0, 1, 2, …) being linear combinations of the Sturm function, which represents the unperturbed state, with $8N^2$ functions of the corresponding complete set with indices adjacent to the parabolic quantum number of the initial level. A method for recursive analytic calculation of the coefficients of the linear combination for any order N is developed. General expressions for corrections to the matrix elements and intensities of the radiation transitions between Stark sublevels are obtained. Analytic formulas and numerical values of the corrections up to the fourth order for the Lyman and Balmer series are presented. A comparison with the available data for transitions between the Stark components of Rydberg states is given. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Optical properties of an atom in an electric field depend on the field strength. Shift and splitting of spectral lines, the Stark effect, are caused by the change of energy of atomic levels. For a hydrogen atom, they are calculated in an arbitrary order of perturbation theory for the atom–field interaction [1, 2]. Nevertheless, up to now there are some features of this phenomenon that have not yet been completely studied, so the Stark effect still remains one of the central problems of atomic physics.

Along with the change of frequencies under the influence of the field, a change in the intensity of lines is observed. The effect of the field dependence of intensity provides additional spectroscopic information about the atomic structure, which can be used for optical diagnostics of fields exerted on the atom and for control of the radiation and absorption of light by matter. The intensity of lines can be used to determine the change of the radiation matrix element induced by the field and, thus, find corrections to the wave functions of the initial and final states of the atom.

Determining the dependence of line intensities on the electric field requires new approaches (both theoretical and experimental) to the study of the atomic structure different from those used for determining the dependence of atomic frequencies on the field. Thus, the problem of the change of the matrix elements and intensity of the radiation transitions in the field remains poorly studied, and in many cases no relevant information can be found in the literature.

The change of probability for transitions between Rydberg states in the hydrogen atom was experimentally observed in [3] by the method of ionization spectroscopy. For the theoretical interpretation of the results obtained, the Hamiltonian of the interaction of the atom with the field was numerically diagonalized in [3] in the finite basis of closest energy states. The efficiency of such calculations is limited not only by the incompleteness of the basis states and the necessity to check the accuracy and convergence of the results but also by the absence of any analytic relations that could help analyze the dependence of the effect on the quantum numbers of the initial and final states. Another obstacle is a substantial amount of calculations for every particular transition. The possibility of obtaining simple formulas that represent corrections to the energy of Stark states in terms of parabolic quantum numbers [1, 2] gives reason to expect that similar analytic expressions can be obtained for corrections to the wave functions and matrix elements of radiation transitions.

Expressions for the corrections of the first [4] and second [5] orders have been recently obtained with the help of the Coulomb Green function in parabolic coordinates. The corresponding numerical values of the corrections to the probability of radiation transitions in the field are in good agreement with the experimental data obtained in [3]. However, in strong fields, the first two orders are insufficient for the description of the effect in the vicinity of the ionization threshold. In addition, for transitions between dipoleless states, the corrections of the first or second order (depending on the polarization of radiation) are zero, whereas for estimating the applicability of perturbation theory, at least two nonzero terms of the asymptotic series must be known. This fact stimulates the development of a reliable method for consecutively calculating higher order corrections of perturbation theory for the wave function and matrix

elements similar to corrections to energy. Thus, the half-century-long history of the calculation of the Stark effect by perturbation theory for energy repeats itself nowadays with respect to the calculation of wave functions (the correct expression of the fourth-order correction was first obtained only in 1974 in [6], i.e., almost half a century after the creation of quantum mechanics and derivation of corrections of the first three orders). However, availability of computer algebra systems, which helped calculate Stark energy corrections of an arbitrary high order (up to several dozens and even hundreds) by the early 1980s [1], allows one to substantially reduce the time required to complete the calculation of Stark corrections to the wave functions.

At first sight, these calculations can be based on the same equations with separated parabolic variables as those used in the calculation of energies (see, e.g., [2, 7, 8]). Although these equations are rather simple, the derivation of Rayleigh–Schrödinger series on their basis is tedious. This is due to the fact that, in addition to the recursive procedure for determining coefficients of the linear combination of the Sturm functions representing the state vector of the atom in the field, one must derive an explicit field dependence from the arguments of those functions and transform their combination to a power series with field-independent coefficients.

The procedure for deriving the power series for the Stark wave functions without solving the system of two coupled equations with separable parabolic variables ξ and η can be developed on the basis of the integral form of the Schrödinger equation for an atom in a field with the help of the closed analytic representation of the reduced Green function in the parabolic coordinates [4].

An analytic representation of the Coulomb Green function makes it possible to obtain expressions for the coefficients of power series in the field strength F for the wave function, matrix elements, and intensities of dipole transitions in the form of polynomials depending on the parabolic quantum numbers of the initial and final levels. In this paper, we present a general method for the calculation of the perturbation theory series for the wave functions of Stark states based on a recurrent relation between the series coefficients. This relation provides a basis for developing a computer code for the analytical and numerical calculation of high-order corrections. The application of this method to the calculation of the first- and second-order corrections yields the same results as those obtained in [4, 5]. In this paper, we derive general symmetry relations, which make it possible to considerably simplify the calculation of coefficients for the expansion of the wave function in the Sturm functions and give analytic expressions for the coefficients of up to the fourth order. These expressions are obtained as functions of the parabolic quantum numbers. Asymptotic properties of corrections for transitions to Rydberg states with large quantum numbers are discussed. Numerical values of the coefficients that determine the corrections of the first four orders to the matrix elements and line intensities of the Lyman and Balmer series are obtained.

2. HIGH-ORDER CORRECTIONS FOR THE WAVE FUNCTIONS OF STARK STATES OF HYDROGEN

The technique of parabolic variable separation, which is used when calculating the Stark effect of hydrogen energy levels [1, 2, 6–8], is inefficient for deriving the wave function in the form of power series with field-independent coefficients (the Rayleigh–Schrödinger series). For this purpose, the use of the Coulomb Green function in the parabolic coordinates [9] seems to be the most convenient. Due to the fact that the interaction operator of the atom with the field is diagonal in the parabolic coordinates, the hydrogen Stark states belonging to the shell with the fixed principal quantum number n are independent in the first order with respect to the field. Hence, the first-order corrections to the wave function and the second-order corrections to the energy can be calculated with the help of perturbation theory for nondegenerate states and the partially reduced Green function [4]. However, already for the second order, the matrix element of the interaction Hamiltonian of the atom with the field is nondiagonal, and the states of the parabolic basis belonging to the given shell are mixed by the field. Thus, the calculation of the second- and higher order corrections must be based on perturbation theory for degenerate states with the completely reduced Green function [5]. Below, we develop a generalization of the method presented in [4, 5] for the case of arbitrary orders of perturbation theory for the wave function that makes it possible to automate the calculations with the help of computer algebra systems. Since the calculations of higher orders use all corrections of the lower orders, the reliability of the results is automatically checked by lower orders. As an example, we present the calculation results up to the fourth order.

2.1. Perturbation Theory for Parabolic States

An unperturbed state of the hydrogen-like ion (in the absence of the external field) is described by the following wave function in the parabolic system of coordinates (a Stark state):[1]

$$\psi_{nn_1n_2m}(\mathbf{r}) = A_{n_1n_2m} f^m_{n_1}\left(\frac{Z\xi}{n}\right) f^m_{n_2}\left(\frac{Z\eta}{n}\right)\frac{e^{im\varphi}}{\sqrt{2\pi}}, \qquad (1)$$

where m is the absolute value of the magnetic quantum number everywhere except for the power of the exponent in which it can be both positive and negative,

$$A_{n_1n_2m} = \frac{1}{n^2}\sqrt{\frac{2Z^3(n_1+m)!(n_2+m)!}{n_1!n_2!}}, \qquad (2)$$

[1] We use the atomic system of units with $e = m = \hbar = 1$.

is the normalizing constant, Z is the charge of the atomic nucleus, and

$$f_k^m(x) = \frac{k!}{(k+m)!}e^{-x/2}x^{m/2}L_k^{(m)}(x) = \frac{1}{m!}e^{-x/2}x^{m/2}{}_1F_1(-k;\, m+1;\, x) \quad (3)$$

is the Sturm function of the Coulomb wave equation represented in terms of generalized Laguerre polynomials $L_k^{(m)}(x)$ and the degenerate hypergeometric function ${}_1F_1(-k;\, m+1;\, x)$ [10].

Parabolic quantum numbers of the states with identical principal, n, and magnetic, m, quantum numbers are reciprocally dependent: $n_1 + n_2 = n - m - 1$. Therefore, along with the parabolic numbers, Stark states can be unambiguously represented with the help of the so-called electric quantum number $q = n_1 - n_2$ such that the set $n_1 n_2 m$ can be replaced by the set nqm as we do below.

The use of the parabolic basis is based on the axial symmetry of the interaction operator of the atom with a homogeneous electric field:

$$\hat{V}(\mathbf{r}) = zF, \quad (4)$$

where z is the projection of the electron's position vector $\mathbf{r}$ on the direction of electric field $\mathbf{F}$. The integral Schrödinger equation for the exact wave function of the state, which transforms into (1) in a weak field, can be represented in the following form taking into account the degeneracy of the states with identical n and m (see [11]):

$$\Psi_{nn_1n_2m}(\mathbf{r}) = \sum_{n_1'=0}^{n-m-1} a_{n_1'}\psi_{nn_1'n_2'm}(\mathbf{r}) - G_E'(\mathbf{r},\mathbf{r}')\hat{V}(\mathbf{r}')|\Psi_{nn_1n_2m}(\mathbf{r}')\rangle = \sum_{n_1'=0}^{n-m-1} a_{n_1'}[1 + G_E'(\mathbf{r},\mathbf{r}')\hat{V}(\mathbf{r}')]^{-1}|\psi_{nn_1'n_2'm}(\mathbf{r}')\rangle, \quad (5)$$

where $n_1' + n_1' = n_1 + n_2 = n - m - 1$. The coefficients of the expansion of the wave function for the atom in a field in the states of the degenerate basis $a_{n_1'} = \langle \psi_{nn_1'n_2'm}|\Psi_{nn_1n_2m}\rangle$ satisfy the initial condition $a_{n_1'} \xrightarrow[F\to 0]{} \delta_{n_1'n_1}$ (where δ is the Kronecker symbol); G_E' is the reduced Green's function

$$G_E'(\mathbf{r},\mathbf{r}') = G_E(\mathbf{r},\mathbf{r}') - \sum_{n_1'=0}^{n-m-1}\frac{\psi_{nn_1'n_2'm}(\mathbf{r})\psi^*_{nn_1'n_2'm}(\mathbf{r}')}{E_n - E}. \quad (6)$$

Here $G_E(\mathbf{r}, \mathbf{r}')$ is the Green's function for the Hamiltonian $\hat{H}_0$ of the free atom and the exact value of the energy E of the atom in the field:

$$(\hat{H}_0 - E)G_E(\mathbf{r},\mathbf{r}') = \delta(\mathbf{r}-\mathbf{r}').$$

The Green's function (6) can be written in terms of the reduced Green's function with the energy E_n of the unperturbed atom:

$$G_{E_n}'(\mathbf{r},\mathbf{r}') = G_m^{(n)}(\xi,\eta;\,\xi',\eta')\frac{e^{im(\varphi-\varphi')}}{2\pi},$$

where

$$G_m^{(n)}(\xi,\eta;\,\xi',\eta') = \frac{2Z}{n}\sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}\frac{(k_1+m)!}{k_1!}\frac{(k_2+m)!}{k_2!} \times \frac{f_{k_1}^m\left(\frac{Z\xi}{n}\right)f_{k_1}^m\left(\frac{Z\xi'}{n}\right)f_{k_2}^m\left(\frac{Z\eta}{n}\right)f_{k_2}^m\left(\frac{Z\eta'}{n}\right)}{k_1+k_2+m+1-n} + \frac{2Z}{n^2}\sum_{\nu_1=0}^{n-m-1}\frac{(\nu_1+m)!}{\nu_1!}\frac{(\nu_2+m)!}{\nu_2!} \times\left(\frac{5}{2} + \xi\frac{\partial}{\partial\xi} + \xi'\frac{\partial}{\partial\xi'} + \eta\frac{\partial}{\partial\eta} + \eta'\frac{\partial}{\partial\eta'}\right) \times f_{\nu_1}^m\left(\frac{Z\xi}{n}\right)f_{\nu_1}^m\left(\frac{Z\xi'}{n}\right)f_{\nu_2}^m\left(\frac{Z\eta}{n}\right)f_{\nu_2}^m\left(\frac{Z\eta'}{n}\right), \quad \nu_2 = n - \nu_1 - m - 1, \quad (7)$$

using the Taylor series expansion

$$G_E'(\mathbf{r},\mathbf{r}') = \sum_{N=0}^{\infty}(G_{E_n}'(\mathbf{r},\mathbf{r}'))^{N+1}(\Delta E)^N = G_{E_n}'(\mathbf{r},\mathbf{r}') + \Delta E G_{E_n}'(\mathbf{r},\mathbf{r}'')G_E'(\mathbf{r}'',\mathbf{r}'). \quad (8)$$

Due to the axial symmetry of perturbation (4), the dependence on the angular variable φ remains the same as for the unperturbed wave function (1). It does not influence on the calculations and is omitted below.

For the coefficients $a_{n_1'}$ in the wave function (5), certain transformations taking into account that the matrix element of operator (4) is diagonal,

$$V_{n_1n_1'} = \langle \psi_{nn_1n_2m}|V(\mathbf{r})|\psi_{nn_1'n_2'm}\rangle = \frac{3Fn}{2Z}q\delta_{n_1n_1'},$$

yield the equation

$$a_{n_1'} = \left[\langle \psi_{nn_1'n_2'm}|\hat{V}(\mathbf{r})G_m^{(n)}(\mathbf{r},\mathbf{r}')(\hat{V}(\mathbf{r}') - \Delta E)|\Psi_{nn_1n_2m}\rangle + (\Delta E - \Delta E^{(1)})a_{n_1'}\right](V_{n_1'n_1'} - \Delta E^{(1)})^{-1}, \quad (9)$$

where $\Delta E^{(1)} = V_{n_1 n_1}$ is the first-order correction to energy and $n_1' \neq n_1$. This equation allows an iterative calculation of those coefficients.

Using (5) and (9), the correction of any order to the wave function can be expressed in terms of corrections of the lower orders in the field strength. This is true for all terms of the sum (5) except for the term with $n_1' = n_1$, which can be determined with the help of the normalization condition

$$\langle \Psi_{nn_1n_2m} | \Psi_{nn_1n_2m} \rangle = \sum_{n_1'} |a_{n_1'}|^2 + \left\langle \Psi_{nn_1n_2m} \middle| \hat{V}(G_E')^2 \hat{V} \middle| \Psi_{nn_1n_2m} \right\rangle = 1.$$

2.2. Sturm Function Expansion of the Corrections to the Wave Functions of Stark States

Orthogonality of generalized Laguerre polynomials allows an analytic computation of integrals over the parabolic variables in (5), (9). Hence, the exact wave function can be written as a linear combination of Sturm functions (3) of the Schrödinger equation for the unperturbed atom:

$$\Psi_{nn_1n_2m}(\mathbf{r}) = A_{n_1n_2m} \times \sum_{i_1} \sum_{i_2} b_{i_1i_2}(n_1n_2m) f^m_{n_1+i_1}\left(\frac{Z\xi}{n}\right) f^m_{n_2+i_2}\left(\frac{Z\eta}{n}\right). \quad (10)$$

This expansion provides a basis for deriving the Rayleigh–Schrödinger series for the wave function of an atom in the field. The main advantage of expansion (10) over the functions that appear in the method of parabolic variable separation is the fact that the arguments of the Sturm functions are independent of the field. The dependence on the field in this expression is hidden in the superposition coefficients *b*, which makes it possible to represent the wave function (10) in the form of a series in powers of the field *F*. It must be noted that, together with Eqs. (5) and (7), expansion (10) follows from the completeness of the Sturm functions of the Coulomb Schrödinger equation with a fixed energy; indeed, this property makes it possible to use the Sturm functions as a basis for resolving any function with the same boundary conditions.

Calculation of the coefficients $b_{i_1i_2}$ can be reduced to a recursive procedure by substituting expansion (10) into (9) and (5). After certain transformations, we obtain the following formula for any coefficient in (10) except for b_{00}:

$$b_{k_1k_2} = \sum_{t_1t_2} X_{t_1t_2k_1k_2} b_{t_1t_2}, \quad (11)$$

where

$$X_{t_1t_2k_1k_2} = \frac{2Z\delta_{k_1+k_2,0}}{3nF(k_1-k_2)} \sum_{N=2}^{\infty} \Delta E^{(N)} \tilde{U}_{n_1+k_1n_2+k_2n_1+t_1n_2+t_2} + \sum_{l_1l_2} (u_{n_1+t_1n_2+t_2l_1l_2} - \Delta E u_{n_1+t_1n_2+t_2l_1l_2}) \times \left(\frac{2Z\delta_{k_1+k_2,0} \tilde{V}_{n_1+k_1n_2+k_2l_1l_2}}{3n(k_1-k_2)F} - \delta_{l_1n_1+k_1}\delta_{l_2n_2+k_2} \right), \quad (12)$$

$$v_{k_1k_2k_1'k_2'} = \frac{n^2}{2Z^2} \times \left[\left(\delta_{k_1'+k_2',n_1+n_2} + \frac{2n(1-\delta_{k_1'+k_2',n_1+n_2})}{k_1'+k_2'+m+1-n} \right) \tilde{V}_{k_1'k_2'k_1k_2} + (\delta_{k_1'+k_2',n_1+n_2+1} - \delta_{k_1'+k_2',n_1+n_2}) \times ((k_1'+m)\tilde{V}_{k_1'-1k_2'k_1k_2} + (k_2'+m)\tilde{V}_{k_1'k_2'-1k_1k_2}) + (\delta_{k_1'+k_2',n_1+n_2} - \delta_{k_1'+k_2',n_1+n_2-1}) \times ((k_1'+1)\tilde{V}_{k_1'+1k_2'k_1k_2} + (k_2'+1)\tilde{V}_{k_1'k_2'+1k_1k_2}) \right]. \quad (13)$$

Here

$$\tilde{V}_{k_1k_2k_1'k_2'} = A^2_{k_1k_2m} \left\langle f^m_{k_1} f^m_{k_2} \middle| \hat{V} \middle| f^m_{k_1'} f^m_{k_2'} \right\rangle.$$

The formula for $u_{k_1k_2k_1'k_2'}$ is obtained from $v_{k_1k_2k_1'k_2'}$ when replacing $\tilde{V}_{k_1k_2k_1'k_2'}$ by

$$\tilde{U}_{k_1k_2k_1'k_2'} = A^2_{k_1k_2m} \left\langle f^m_{k_1} f^m_{k_2} \middle| f^m_{k_1'} f^m_{k_2'} \right\rangle.$$

The tensor (12) is undefined for $k_1 = k_2 = 0$.

Now, we expand every coefficient $b_{i_1i_2}$ in a series in the field strength separating, for convenience, the scale factor, which depends on the principal quantum number:

$$b_{i_1i_2} = \sum_{N=0}^{\infty} F^N \left(\frac{n}{2Z}\right)^{3N} b^{(N)}_{i_1i_2}. \quad (14)$$

A similar expansion for tensor (12) has the form

$$X_{t_1t_2k_1k_2} = \sum_{N=1}^{\infty} F^N \left(\frac{n}{2Z}\right)^{3N} X^{(N)}_{t_1t_2k_1k_2}. \quad (15)$$

Here $b^{(N)}_{i_1i_2}$ and $X^{(N)}_{t_1t_2k_1k_2}$ are independent of the field. The fact should be noted that the expansion for *b* includes the term of the zero order in the field strength, whereas the expansion for *X* begins with the first-order term. This shows that equation (11) reflects a relation between the higher order coefficients of expansion (14)

and the lower order ones. In particular, this explains the fact that Eq. (11) does not hold for $k_1 = k_2 = 0$.

Equations (11)–(13) show that the coefficients $b^{(N)}_{i_1 i_2}$ are nonzero for $|i_1| = 2M$, $2M - 1$ and $|i_2| \le 2N - 2M$ ($M = 0, \ldots, N$). Thus, the total number of nonzero coefficients $b^{(N)}_{i_1 i_2}(n_1 n_2 m)$ is $8N^2 + 1$. Therefore, expansion (10) consists of nine nonzero terms of the first order (one of the indices of $b^{(1)}_{i_1 i_2}(n_1 n_2 m)$ must be zero), 33 terms of the second, 73 terms of the third, and so on.

The technique described makes it possible also to obtain the correction to energy, ΔE of an arbitrary order, representing it in terms of $b_{i_1 i_2}$. Since the energy corrections in an electric field are calculated for almost arbitrarily high orders (see [1]), one can assume them to be known and use them to check the validity of calculations of corrections to the wave functions.

For the first four terms of expansion (14) (except for the zero-order one), Eq. (11) yields

$$b^{(1)}_{k_1 k_2} = \sum_{t_1 t_2} X^{(1)}_{t_1 t_2 k_1 k_2} b^{(0)}_{t_1 t_2} = X^{(1)}_{00 k_1 k_2}, \tag{16}$$

$$b^{(2)}_{k_1 k_2} = \sum_{t_1 t_2} X^{(1)}_{t_1 t_2 k_1 k_2} b^{(1)}_{t_1 t_2}, \tag{17}$$

$$b^{(3)}_{k_1 k_2} = \sum_{t_1 t_2} (X^{(1)}_{t_1 t_2 k_1 k_2} b^{(2)}_{t_1 t_2} + X^{(2)}_{t_1 t_2 k_1 k_2} b^{(1)}_{t_1 t_2}), \tag{18}$$

$$b^{(4)}_{k_1 k_2} = \sum_{t_1 t_2} (X^{(1)}_{t_1 t_2 k_1 k_2} b^{(3)}_{t_1 t_2} + X^{(2)}_{t_1 t_2 k_1 k_2} b^{(2)}_{t_1 t_2} + X^{(3)}_{t_1 t_2 k_1 k_2} b^{(1)}_{t_1 t_2}). \tag{19}$$

It can be demonstrated that, for $t_1 = t_2 = 0$, tensor (12) is nonzero only for $N = 1$. Hence, for $N \ge 2$, the expression for $b^{(N)}$ does not include $X^{(N)}$, which depends on $\Delta E^{(N+1)}$. Thus the energy corrections of order N and lower should be known for determining $b^{(N)}_{k_1 k_2}$ ($X^{(1)}$ and, therefore, $b^{(1)}$ depend on $\Delta E^{(2)}$).

Equations (16)–(19) hold for all values of the indices k_1 and k_2 except for $k_1 = k_2 = 0$. The coefficients $b^{(N)}_{00}$ can be determined from the normalization condition for the wave function. Let us consider the Nth-order term in the equation

$$\langle \Psi_{n n_1 n_2 m} | \Psi_{n n_1 n_2 m} \rangle = 1,$$

take out of it the products $b^{(0)} b^{(N)}$, and solve the equation obtained for $b^{(N)}_{00}$. Then, for $N > 0$, we have

$$b^{(N)}_{00} = -\frac{1}{2} \sum_{k_1 k_2} \frac{A^2_{n_1 n_2 m}}{A^2_{n_1 + k_1 n_2 + k_2 m}} \times \sum_{l_1 l_2} \sum_{i=1}^{N-1} b^{(i)}_{k_1 k_2} b^{(N-i)}_{l_1 l_2} \tilde{U}_{n_1 + k_1 n_2 + k_2 n_1 + l_1 n_2 + l_2} - \sum_{(t_1 t_2) \ne (00)} b^{(N)}_{t_1 t_2} \tilde{U}_{n_1 n_2 n_1 + t_1 n_2 + t_2}. \tag{20}$$

For $N = 1$, the sum over i disappears, and only the sum over t_1, t_2 remains on the right-hand side. Thus, to calculate $b^{(N)}_{00}$ following Eq. (20), one must determine all the coefficients $b^{(N)}_{i_1 i_2}$ with $(i_1 i_2) \ne (00)$, which are determined by similar coefficients of all lower orders.

In the general case, the recurrent relations are rather cumbersome. However, for the "outside" coefficients, general formulas can be obtained of the form

$$b^{(N)}_{2N\,0} = (-1)^N \frac{(n_1 + m + 1)_{2N}}{N!}, \qquad b^{(N)}_{-2N\,0} = \frac{(n_1 - 2N + 1)_{2N}}{N!}, \tag{21}$$

and

$$b^{(N)}_{2N-1\,0} = (-1)^{N-1} \frac{2(n_1 + m + 1)_{2N-1}}{3(N-1)!} \times (6n - 3q + 8N - 2) \tag{22}$$

$$b^{(N)}_{-(2N-1)\,0} = -\frac{2(n_1 - 2N + 2)_{2N-1}}{3(N-1)!} \times (6n - 3q - 8N + 2). \tag{23}$$

Here, we used the conventional notation for the Pochhammer symbol, $(a)_n = a(a+1)\ldots(a+n-1)$.

The symmetry of the perturbation operator (4) is reflected in the symmetry of the coefficients:

$$b^{(N)}_{i_1 i_2}(nqm) = (-1)^N b^{(N)}_{i_2 i_1}(n - qm), \qquad b^{(N)}_{i_1 i_2}(nqm) = (-1)^N b^{(N)}_{-i_1 - i_2}(-n - qm). \tag{24}$$

Here i_1 and i_2 are nonnegative integers. These relations allow a significant simplification of the calculation of coefficients. In general, these calculations are based on recurrent formulas (16)–(19), which express $b^{(N)}_{i_1 i_2}$ in terms of the coefficients of lower orders. For certain i_1 and i_2, only one nonzero term containing $X^{(1)}_{t_1 t_2 k_1 k_2}$ can remain in these formulas; in this case, the relation

between the coefficients of higher and lower orders becomes much simpler:

$$b^{(N)}_{\pm i_1 \pm i_2} = b^{(N-1)}_{\pm i_1\, 0} b^{(1)}_{0 \pm i_2},$$

$$i_1 = 2N-2, 2N-3, \quad i_2 = 1, 2.$$

Coefficients with the indices that do not satisfy the conditions listed above are described by more complex expressions of the general form:

$$\begin{aligned} b^{(N)}_{i_1 i_2}(nqm) &= (n-n_2)_{i_1}(n-n_1)_{i_2} P_K^{(i_1,i_2)}(n,q,m), \\ b^{(N)}_{-i_1 i_2}(nqm) &= (n_1-i_1+1)_{i_1}(n-n_1)_{i_2} R_K^{(i_1,i_2)}(n,q,m), \end{aligned} \tag{25}$$

where the lower index of the polynomials $P_K^{(i_1,i_2)}(n,q,m)$ and $R_K^{(i_1,i_2)}(n,q,m)$ determines their order, $K = 2N - i_1 - i_2$, with respect to each of the three arguments. The magnetic quantum number m occurs in these polynomials only in even powers with an exponent not greater than K. For a zero value of one of the indices i_1 or i_2, there is a simple relation between the polynomials P and R, which follows from the symmetry relations (24):

$$P_K^{(i_1,0)}(n,q,m) = (-1)^{N+i_1} R_K^{(i_1,0)}(-n,-q,m).$$

The coefficients $b^{(N)}_{00}$ possess a specific feature: for even orders $N = 2M$, they are polynomials of degree $2M$ of the squares of quantum numbers; for odd $N = 2M + 1$, they are similar polynomials multiplied by the electric quantum number:

$$\begin{aligned} b^{(2M)}_{00}(nqm) &= Q^e_{2M}(n^2, q^2, m^2), \\ b^{(2M+1)}_{00}(nqm) &= qQ^0_{2M}(n^2, q^2, m^2). \end{aligned} \tag{26}$$

Thus, $Q^e_0 = 1$, $Q^0_0 = 6$,

$$\begin{aligned} Q^e_2(n^2, q^2 m^2) &= -\frac{1}{8}[65n^4 - 6n^2(7q^2 + 11m^2 - 249) \\ &+ 17q^4 - 18q^2m^2 + m^4 - 74q^2 - 650m^2 + 1289], \end{aligned} \tag{27}$$

$$\begin{aligned} Q^0_2(n^2, q^2, m^2) &= \frac{1}{4}[525n^4 \\ &+ 2n^2(31q^2 + 395m^2 + 13191) - 99q^4 + 93m^4 \\ &+ 6q^2m^2 - 818q^2 + 6494m^2 + 31365]. \end{aligned} \tag{28}$$

A similar property holds for the diagonal coefficients $b^{(N)}_{k,k}$ with the indices $k < N$: for odd N, they are proportional to q. In particular, the polynomials that determine $b^{(3)}_{1\,1}$ and $b^{(3)}_{2\,2}$ in accordance with (25) have the form

$$\begin{aligned} P_4^{(1,1)}(n,q,m) &= -4q[4n^2(n+4) \\ &- n(q^2+24) - 13q^2 - 12m^2 + 12], \\ P_2^{(2,2)}(n,q,m) &= -2q(8n+19), \end{aligned} \tag{29}$$

while for the polynomials that determine "antidiagonal" coefficients, for example, $b^{(3)}_{-1\,1}$ and $b^{(3)}_{-2\,2}$, we have

$$\begin{aligned} &R_4^{(1,1)}(n,q,m) \\ &= -2[4n^4 + n^2(3q^2 - 4m^2 - 20q - 352) \\ &\quad - q^4 + 22q^3 + q^2(m^2 + 99) \\ &\quad + q(28m^2 + 180) + 52(m^2 + 3)], \\ &R_2^{(2,2)}(n,q,m) = 2(8n^2 + 2q^2 + 19q + 48). \end{aligned} \tag{30}$$

Equations (21)–(27) supplemented by the polynomials

$$P_2^{(2,0)} = 2[(2n-q)(2n-q+11) + 24] \tag{31}$$

$$\begin{aligned} P_3^{(1,0)} &= \frac{1}{2}[2n^3 + n^2(3q-58) - 2n(m^2 - 33q + 117) \\ &- q^3 + qm^2 - 36q^2 + 75q - 26m^2 - 78] \end{aligned} \tag{32}$$

and the relation between $b^{(1)}_{ik}$ and $b^{(2)}_{ik}$ mentioned above completely determine the analytic expressions for the coefficients of expansion (14) of the first and second order in F. To complete determination of the coefficients of the third order, it is sufficient to write out, in addition to (21)–(30), the polynomials expressing the coefficients $b^{(3)}_{40}$, $b^{(3)}_{30}$, $b^{(3)}_{20}$, $b^{(3)}_{10}$, $b^{(3)}_{21}$, and $b^{(3)}_{-21}$ in accordance with (25):

$$\begin{aligned} &P_2^{(4,0)}(n,q,m) \\ &= -\frac{1}{3}(24n^2 + 196n - 24nq + 6q^2 - 89q + 340), \\ &P_3^{(3,0)}(n,q,m) = \frac{1}{12}[122n^3 - n^2(201q - 2058) \\ &\quad + 2n(48q^2 + 3m^2 - 927q + 4295) - 13q^3 \\ &\quad - 3qm^2 + 504q^2 + 162m^2 - 3161q + 8718], \\ &P_4^{(2,0)}(n,q,m) = \frac{1}{96}[777n^4 - 4n^3(3q + 2362) \\ &\quad - 6n^2(87q^2 - 2508q + 131m^2 + 16511) \\ &\quad - 4n(3q^3 + 2286q^2 - 3qm^2 + 1478m^2 \\ &\quad - 18303q + 70638) + 201q^4 + 9m^4 + 2072q^3 \\ &\quad - 210q^2m^2 + 2920qm^2 - 18810q^2 \\ &\quad - 19738m^2 + 65736q - 162671], \end{aligned}$$

$$P_5^{(1,0)}(n,q,m) = -\frac{1}{12}[294n^5 - n^4(243q - 928) - n^3(300m^2 + 132q^2 - 340q - 688) + 2n^2(69q^3 + 27qm^2 - 246q^2 - 352m^2 + 249q - 14458) + 2n(27q^4 + 3m^4 - 18q^2m^2 + 2q^3 + 166qm^2 - 864q^2 - 2852m^2 + 2710q - 22131) - 39q^5 + 148q^4 + 14q^3(3m^2 + 29) - 2q^2(70m^2 + 892) - q(3m^4 - 5978m^2 - 5985) - 32m^4 - 6052m^2 - 11676], \quad (33)$$

$$P_3^{(2,1)}(n,q,m) = -\frac{1}{2}[66n^3 - n^2(35q - 358) - 2n(8q^2 + m^2 + 41q - 251) + 9q^3 - qm^2 - 84q^2 - 26m^2 - 11q + 306],$$

$$R_3^{(2,1)}(n,q,m) = -\frac{1}{2}[62n^3 - n^2(29q + 230) - 2n(8q^2 - m^2 + 23q - 133) + 7q^3 + qm^2 + 116q^2 + 26m^2 + 395q + 462].$$

All polynomial expressions mentioned above are quadratic functions of the magnetic quantum number m; this is due to the fact that the vectors of the electric field and the dipole moment of the atom are polar. Thus, the corrections to the wave function are independent of the sign of m, as well as the energy corrections. The same property is characteristic of the corrections to the matrix elements and intensities of the radiation transitions, which are discussed in the following section.

3. THE DIPOLE MATRIX ELEMENT AND INTENSITY OF RADIATION TRANSITION

The matrix element for the dipole transition between the Stark levels of the hydrogen atom in a homogeneous field **F**,

$$d_{nn'}(F) = \langle \Psi_{nn_1n_2m}|\mathbf{d}|\Psi_{n'n_1'n_2'm'}\rangle, \quad (34)$$

can be written in the form of a power series in the field strength F with the help of expansion (10) for the wave functions of the initial and final states. In particular, we have for the third order

$$d^{(3)}_{n'\leftarrow n} = \langle \Psi^{(0)}_{n'}|d|\Psi^{(3)}_n\rangle + \langle \Psi^{(1)}_{n'}|d|\Psi^{(2)}_n\rangle + \langle \Psi^{(2)}_{n'}|d|\Psi^{(1)}_n\rangle + \langle \Psi^{(3)}_{n'}|d|\Psi^{(0)}_n\rangle. \quad (35)$$

Using expansion (10) for the wave functions in this equation, we obtain a linear combination of the dipole matrix elements with the Sturm functions (3). The matrix elements can be conveniently expressed in terms of generalized hypergeometric functions of two variables [5], which can be transformed into a combination of the Gauss hypergeometric functions by analogy with the Gordon formula for the matrix element of the radiation transition between the Stark levels [7]. Let us represent the transition frequency, the dipole matrix element, and the intensity in the form of the series

$$\omega_{nn'}(F) = \omega_{nn'}(0)\left(1 + \sum_{N=1}^{\infty} F^N w^{(N)}_{nn'}\right), \quad (36)$$

$$d_{nn'}(F) = d_{nn'}(0)\left(1 + \sum_{N=1}^{\infty} F^N r^{(N)}_{nn'}\right), \quad (37)$$

$$I_{nn'}(F) = I_{nn'}(0)\left(1 + \sum_{N=1}^{\infty} F^N \beta^{(N)}_{nn'}\right), \quad (38)$$

where the quantities corresponding to the unperturbed atom are taken out from the parentheses and the coefficients of the expansion are ratios of the correction terms to the unperturbed ones. Taking into account the fact that

$$I_{n,n'} \sim \omega^4_{nn'}|d_{nn'}|^2, \quad (39)$$

we can write the relation between the coefficients of expansions (36)–(38) up to the third order in the form

$$\beta^{(1)}_{nn'} = 4w^{(1)}_{nn'} + 2r^{(1)}_{nn'}, \quad (40)$$

$$\beta^{(2)}_{nn'} = 4w^{(2)}_{nn'} + 6(w^{(1)}_{nn'})^2 + 8w^{(1)}_{nn'}r^{(1)}_{nn'} + (r^{(1)}_{nn'})^2 + 2r^{(2)}_{nn'}, \quad (41)$$

$$\beta^{(3)}_{nn'} = 4\left[w^{(3)}_{nn'} + 3w^{(1)}_{nn'}w^{(2)}_{nn'} + (w^{(1)}_{nn'})^3\right] + 4r^{(1)}_{nn'}\left[2w^{(2)}_{nn'} + 3(w^{(1)}_{nn'})^2\right] + 4\left[(r^{(1)}_{nn'})^2 + 2r^{(2)}_{nn'}\right]w^{(1)}_{nn'} + 2(r^{(3)}_{nn'} + r^{(1)}_{nn'}r^{(2)}_{nn'}). \quad (42)$$

The symmetry relations (24) transform into the symmetry relations for the coefficients of expansions (37), (38), which we will refer to as the radiation susceptibilities

$$r^{(N)}_{nqm\to n'q'm'} = (-1)^N r^{(N)}_{n-qm\to n'-q'm'}, \quad \beta^{(N)}_{nqm\to n'q'm'} = (-1)^N \beta^{(N)}_{n-qm\to n'-q'm'}. \quad (43)$$

A discussion of properties and numerical values of these quantities in the first and second orders of perturbation theory can be found in [4, 5]. In this paper, we give a thorough analysis of the third- and fourth-order susceptibilities.

3.1. Third-Order Radiation Susceptibilities

Higher order corrections are of specific importance when the lower-order corrections are zero. In particular, this is the case for transitions between dipoleless states ($q = 0$ and $q' = 0$). The symmetry of these transitions is such that the unperturbed dipole matrix element and its even-order corrections are equal to zero for the π-radiation; for the σ-radiation, all odd-order corrections in field strength are equal to zero.

Each of the 73 nonzero coefficients $b^{(3)}_{i_1 i_2}$ describing the cubic (in F) component of expansion (10) for the wave functions of the initial and final states is generally a polynomial of the corresponding parabolic quantum numbers determined by one of Eqs. (21)–(33). After calculating the integrals over the parabolic variables in the matrix elements with corrections to the wave function in (35), we obtain a combination of expressions with hypergeometric functions similar to the Gordon formula for the unperturbed radiation matrix element [7], but with shifted (with respect to the parabolic quantum numbers) indices of the Sturm functions.

Such expressions become significantly simpler when one of the states (as a rule, the lower one) has zero or close to zero parabolic quantum numbers. In particular, for the Lyman series (transitions to the ground state with $n' = 1$ and $q' = m' = 0$), the third-order corrections to the dipole matrix element and to the intensities of π-radiation have the form

$$\begin{aligned} r^{(3)\pi}_{1S\leftarrow n} = &-\frac{n}{384q(n^2-1)^3Z^9}[n^2q^4(171n^{12}-2277n^{10}\\ &+1005n^8+55801n^6-17308n^4+4752n^2-672)\\ &-q^2(n^2-1)(7587n^{14}-1353n^{12}-24815n^{10}\\ &-8571n^8+3880n^6+53768n^4-24144n^2+432)\\ &+n^2(n^2-1)^2(1035n^{12}-3129n^{10}-1391n^8\\ &-491n^6+4792n^4-648n^2+216)], \end{aligned} \tag{44}$$

$$\begin{aligned} \beta^{(3)\pi}_{1S\leftarrow n} = &-\frac{n}{192q(n^2-1)^2Z^9}[n^2q^4(171n^{10}-822n^8\\ &+963n^6+472n^4-2064n^2+1152)\\ &-q^2(7416n^{14}-6576n^{12}-9496n^{10}+5952n^8\\ &+2992n^6+4864n^4-8688n^2+3408)\\ &+n^2(n^2-1)(2385n^{12}-2379n^{10}-757n^8\\ &-265n^6+4376n^4-3024n^2-528)]. \end{aligned} \tag{45}$$

The electric quantum number q in the denominator appears due to the proportionality to this factor of the zero-order matrix element and the intensity, as well as of the even-order corrections. Thus, in the limit of the weak field, the Lyman π_0-lines, which correspond to transitions from dipoleless states with $q = 0$, disappear, whereas the odd-order corrections (in particular, the first- and third-order ones) are nonzero. These corrections cause the appearance of π_0-lines with intensities proportional to the square of the field strength (see Section 3.2 for details). The contribution of these corrections is important for lines with a small value of q as compared to the extremal Stark components, which correspond to $q \sim n$. For large n, the asymptotic dependence is for the central,

$$\beta^{(1)\pi}(q\sim 1)\sim 1.5n^5,\quad \beta^{(3)\pi}(q\sim 1)\sim -10n^{13},$$

and extremal,

$$\beta^{(1)\pi}(q\sim 1)\sim 1.5n^4,\quad \beta^{(3)\pi}(q\sim 1)\sim 20n^{12}.$$

Stark components are also different. The estimate of the critical field at which the third-order correction becomes comparable with the first-order correction does not exceed the estimate of the ionizing field at which the upper level appears above the potential barrier that separates the inner region of the motion with a given energy from the outer one (see, e.g., [12]). This means that perturbation theory is applicable to all practically important cases when the upper level ionization decay in the field can be neglected.

Formulas for the corrections to the matrix elements and intensities of σ-transitions are simpler:

$$\begin{aligned} r^{(3)\sigma}_{1S\leftarrow n} = &-\frac{nq}{384(n^2-1)^3Z^9}[n^2q^2(171n^{12}-2277n^{10}\\ &+1005n^8+55801n^6-17308n^4+4752n^2-672)\\ &-(n^2-1)(7371n^{14}+1998n^{12}-40445n^{10}\\ &+3156n^8+3184n^6+53696n^4-24144n^2+432)], \end{aligned} \tag{46}$$

$$\begin{aligned} \beta^{(3)\sigma}_{1S\leftarrow n} = &-\frac{nq}{192(n^2-1)^2Z^9}[n^2q^2(171n^{10}-822n^8\\ &+963n^6+472n^4-2064n^2+1152)\\ &-7371n^{14}+5550n^{12}+12589n^{10}-9408n^8\\ &-2128n^6-4576n^4+8688n^2-3408]. \end{aligned} \tag{47}$$

These expressions are proportional to q and vanish for the central Stark line ($q = 0$) together with all other odd-order corrections. The line itself has the maximum intensity among the Lyman σ-components, and its first nonvanishing correction is determined by the second-order susceptibility of perturbation theory $\beta^{(2)\sigma}_{1S\leftarrow n}$; the asymptotic dependence of this susceptibility on n has the form (see [5])

$$\beta^{(2)\sigma}(q=0)\sim 11.5n^8.$$

Table 1. Corrections of the first four orders, $\beta^{(N)}$, to the intensity of Stark lines of the Lyman series

nqm	Line	$\beta^{(1)}$	$\beta^{(2)}$	$\beta^{(3)}$	$\beta^{(4)}$
210	$L_\alpha(\pi_2)$	4.80(1)	–2.04(3)	4.00(2)	–1.45(7)
201	$L_\alpha(\sigma_0)$	0	–2.93(3)	0	–1.46(7)
320	$L_\beta(\pi_6)$	1.77(2)	–5.91(4)	5.40(6)	–6.91(9)
311	$L_\beta(\sigma_3)$	–6.0	–7.57(4)	7.71(6)	–7.15(9)
430	$L_\gamma(\pi_{12})$	4.99(2)	–6.08(5)	2.37(8)	–6.10(11)
410	$L_\gamma(\pi_4)$	1.56(3)	–2.23(4)	–6.60(8)	–6.78(11)
421	$L_\gamma(\sigma_8)$	–1.60(1)	7.65(5)	3.45(8)	–6.56(11)
401	$L_\gamma(\sigma_0)$	0	–7.45(5)	0	–5.73(11)

Note: The number in parentheses determines the power of ten; i.e., $a(k) \equiv a \times 10^k$.

The range of validity of this correction can be estimated by evaluating the fourth-order correction.

Similar formulas for other series are the more cumbersome the greater the degeneracy order of the lower level.

3.2. Fourth-Order Susceptibilities

In the fourth order of perturbation theory, linear combination (10) consists of 129 nonzero terms, which are represented as polynomials of the parabolic quantum numbers. Eight of them have already been determined by the general relations (21)–(23). The other ones satisfy Eqs. (25), (26); moreover, the number of such coefficients and the corresponding polynomials P_K and R_K grows approximately by a factor of two in comparison with the third order.

We calculated the analytic expressions for the coefficients $b^{(4)}_{i_1 i_2}$, which are much more complex than those for $b^{(3)}_{i_1 i_2}$. The fourth-order corrections to the matrix element (37) and intensity (38) are also more cumbersome in comparison with the third order. In particular, for the Lyman series, which is the simplest case, the least cumbersome expression is

$$\begin{aligned}\beta^{(4)\sigma}_{1S\leftarrow n} &= -\frac{1}{4608(n^2-1)^3 Z^{12}}\\ &\times[-3n^4q^4(861n^{14}+903n^{12}-13137n^{10}+24021n^8\\ &-7928n^6-34048n^4+44064n^2-15840)\\ &+6n^2q^2(51657n^{18}-27079n^{16}-171975n^{14}\\ &+131355n^{12}+812290n^{10}-20728n^8\\ &-30272n^6-34504n^4+15096n^2+7128)\\ &+478881n^{22}+719823n^{20}-4791577n^{18}\\ &+5154605n^{16}-1962964n^{14}+1895920n^{12}\\ &-2052512n^{10}+272464n^8+3759648n^6\\ &-10519344n^4+10555920n^2-3543264].\end{aligned} \tag{48}$$

The numerical values of the coefficients for the first four orders in F of the intensity expansion (38) for Lyman series transitions are presented in Table 1 and for the Balmer series in Table 2. For the vast majority of the lines, the corrections $r^{(N)}$ have the same sign and the order of magnitude as the corresponding $\beta^{(N)}$.

We did not include the transitions $L_\beta(\pi_0)$ and $H_\beta(\pi_0)$ since the intensity of these lines is zero in the zero approximation, and the formal values of coefficients of expansions (37) and (38) can become infinite. Corrections of the first and third orders for these lines differ from zero; thus, their intensities are quadratic with respect to the field (this is due to squaring the first-order matrix element). The corrections to these dependences determined by the ratio of combinations of the matrix elements and frequencies of the third and first orders (the ratio of the cubic and linear corrections) are also quadratic. In particular, the general expression for the corrections to the intensity of the "forbidden" π_0-lines of the Lyman series corresponding to transitions $(n00) \longrightarrow (100)$ can be written in the form

$$\begin{aligned}I^{(3)}/I^{(1)} &= -F^2(2385n^{12}-2379n^{10}-757n^8-265n^6\\ &+4376n^4-3024n^2-528)\\ &\times[96Z^6(n^2-1)(3n^2+1)]^{-1}.\end{aligned} \tag{49}$$

Numerical values of the first-order corrections show that the intensities of such "forbidden" lines reach as much as several percent of the intensity of the corresponding allowed lines in the field with half the strength of the ionizing field. In this case, the third-order correction does not exceed 10% of the first-order one for the field strength of the ionizing field.

For the line $L_\alpha(\pi_2)$, the value of $\beta^{(3)}$ is less than $\beta^{(2)}$. Thus, we can estimate the applicability of lower orders of perturbation theory for this line in strong fields only by the ratio of the second- and fourth-order corrections,

Table 2. Corrections of the first four orders, $\beta^{(N)}$, to the intensity of Stark lines of the Balmer series

nqm	⟶	*n*'	*q*'	*m*'	Line	$\beta^{(1)}$	$\beta^{(2)}$	$\beta^{(3)}$	$\beta^{(4)}$
320	⟶	2	–1	0	$H_\alpha(\pi_8)$	1.53(3)	9.80(5)	3.63(8)	9.63(11)
320	⟶	2	1	0	$H_\alpha(\pi_4)$	3.80(2)	–1.96(4)	–2.26(6)	–5.38(9)
311	⟶	2	0	1	$H_\alpha(\pi_3)$	2.52(2)	–5.38(4)	–5.60(6)	–5.94(9)
300	⟶	2	–1	0	$H_\alpha(\pi_2)$	1.04(2)	–8.46(4)	–7.08(6)	–6.50(9)
320	⟶	2	0	1	$H_\alpha(\sigma_6)$	8.73(1)	2.48(5)	3.33(7)	1.09(9)
311	⟶	2	–1	0	$H_\alpha(\sigma_5)$	7.56(2)	1.70(5)	1.13(7)	–3.72(9)
311	⟶	2	1	0	$H_\alpha(\sigma_1)$	1.65(2)	–7.01(4)	–1.69(6)	–6.15(9)
300	⟶	2	0	1	$H_\alpha(\sigma_0)$	0	–8.76(4)	0	–6.49(9)
302	⟶	2	0	1	$H_\alpha(\sigma_0)$	0	–6.85(4)	0	–5.77(9)
430	⟶	2	–1	0	$H_\beta(\pi_{14})$	2.44(3)	2.06(6)	1.14(9)	2.85(11)
430	⟶	2	1	0	$H_\beta(\pi_{10})$	8.13(2)	–4.34(5)	1.05(8)	–5.25(11)
421	⟶	2	0	1	$H_\beta(\pi_8)$	4.32(2)	–6.82(5)	8.38(7)	–5.49(11)
410	⟶	2	–1	0	$H_\beta(\pi_6)$	–99.1	–8.86(5)	2.02(8)	–5.64(11)
410	⟶	2	1	0	$H_\beta(\pi_2)$	–3.40(3)	1.42(6)	3.32(9)	–1.81(12)
430	⟶	2	0	1	$H_\beta(\sigma_{12})$	1.54(3)	3.53(5)	9.91(7)	–3.62(11)
421	⟶	2	–1	0	$H_\beta(\sigma_{10})$	1.17(3)	–1.38(5)	–1.68(8)	–4.94(11)
421	⟶	2	1	0	$H_\beta(\sigma_6)$	3.00(2)	–7.39(5)	1.54(8)	–5.70(11)
410	⟶	2	0	1	$H_\beta(\sigma_4)$	–3.84(2)	–8.60(5)	4.03(8)	–6.16(11)
412	⟶	2	0	1	$H_\beta(\sigma_4)$	–1.60(2)	–7.22(5)	2.56(8)	–5.71(11)
401	⟶	2	–1	0	$H_\beta(\sigma_2)$	–9.72(2)	–5.39(5)	5.69(8)	–6.61(11)

Note: The number in parentheses determines the power of ten; i.e., $a(k) \equiv a \times 10^k$.

similar to the lines $L_\alpha(\sigma_0)$ and $L_\gamma(\sigma_0)$, for which the first- and third-order corrections are identically equal to zero.

Abnormally large susceptibilities for $H_\alpha(\pi_8)$ compared to the other lines of this series are caused by a small value of its intensity in the limit of the weak field. It is only 0.04% of the intensity of the strongest line of this series.

For the Lyman series, the second- and fourth-order corrections (both for the matrix elements and intensities) are negative, while the first- and third-order corrections are sign-alternating. It must be noted that there is regularity in the behavior of the first nonzero corrections to the intensity of this series: it is seen from Table 1 that they are positive for π-transitions and negative for σ-transitions. This corresponds to the increase of the π-line intensities in the electric field (which is similar to the increase of the probability of optical transitions from deep impurity centers in semiconductors [13]) and the decrease of the σ-line intensities. For the Balmer series, no such regularity exists, and the influence of the field on individual Stark components both of the π-radiation and σ-radiation becomes selective. The same selectivity of the effect of the field on the intensity of lines is observed for other series of radiation and absorption of a hydrogen-like atom [4, 5].

As in the calculation of the Stark energy, the problem of the convergence of series in powers of F for the matrix elements of dipole transitions is of prime importance for the calculation of radiation characteristics. The analysis of the formulas and numerical results for the Lyman and Balmer series and for certain transitions between Rydberg states shows that the terms of series (37) up to the fourth order form a decreasing (in absolute value) sequence for the field strength not exceeding a certain critical value F_{cr}, which is approximately double the strength of the ionizing field for the upper level. Moreover, the sum of the first terms of the series up to the ionizing field is almost equal to the exact value of the matrix element. Table 3 presents data to compare the square of the matrix element, $|d^{(Int)}(F)|^2$, of the radiation π-transition between the lowest Stark components (($q = -(n-1)$, and $q' = -(n'-1)$) of the Rydberg hydrogen states with $n = 30$ and $n' = 10$ obtained by the numerical integration [3] of the Schrödinger equation for the atom in the field of a strength close to the ionizing one with the results of perturbation theory for the first four orders ($N = 1, 2, 3, 4$). The relative differences

$$\epsilon_N = \frac{\left|d^{(N)}(F)\right|^2 - \left|d^{(Int)}(F)\right|^2}{\left|d^{(Int)}(F)\right|^2},$$

represented on a percentage basis, are given in adjacent columns. It is seen from the table that the results

Table 3. Square values of the matrix elements, $|d|^2$, of the radiation transition between the Rydberg states (30, –29, 0) and (10, –9, 0) obtained by the numerical integration of the Schrödinger equation [3] and the corresponding values calculated by perturbation theory for the N_{th} order (N = 1, 2, 3, 4). The relative differences ϵ_N are given on the percentage basis

F, V/cm	$\lvert d^{(Int)}(F)\rvert^2$	$\lvert d^{(N)}(F)\rvert^2$							
		N = 1	ϵ_1, %	N = 2	ϵ_2, %	N = 3	ϵ_3, %	N = 4	ϵ_4, %
100	0.2901	0.2908	0.24	0.2901	0	0.2901	0	0.2901	0
200	0.2821	0.2851	1.06	0.2824	0.11	0.2822	0.04	0.2821	0
300	0.2722	0.2795	2.68	0.2734	0.44	0.2725	0.11	0.2723	0.04
400	0.2598	0.2738	5.39	0.2630	1.23	0.2609	0.42	0.2603	0.19
500	0.2442	0.2682	9.83	0.2513	2.91	0.2472	1.23	0.2455	0.53
600	0.2235	0.2626	17.49	0.2382	6.58	0.2312	3.45	0.2277	1.88
700	0.1937	0.2569	32.63	0.2237	15.49	0.2126	9.76	0.2062	6.45

obtained by perturbation theory rapidly converge to the exact value such that even in the vicinity of the ionizing field (F = 700 V/cm) the difference of the fourth-order approximation from the exact value does not exceed 7%. Moreover, the terms of the series form a decreasing sequence with a factor of about one-half. As the field strength decreases, the convergence rate sharply increases, and the difference from the exact value tends to zero. At F = 200 V/cm, the value obtained by the fourth-order perturbation theory coincides with the exact value accurate to four decimal places given in the table; at F = 100 V/cm, an accurate value is obtained already in the third order.

A similar comparison was performed for the line next to the extremal one, which corresponds to the transition from the upper level state with $q = -(n - 3)$ (the unperturbed matrix element is less by approximately a factor of two for this level). In this case, the difference between the values obtained by perturbation theory and numerical integration is even less than in the preceding case. For F = 700 V/cm, the difference between the fourth-order and exact values is less than 5%.

The results of experimental measurements and numerical calculation of the ratio of probabilities of the radiation transitions between the lower and adjacent Stark components of the states with n' = 10 and n = 30, n = 44 obtained in [3] are in complete agreement with our data at the field strength F for which fine structure effects are negligible. Higher order corrections yield a significant improvement of the agreement with experimental data obtained in the first (linear) approximation, especially in the vicinity of the upper bound of the range of field strength used. Moreover, the contribution of the fourth and higher order corrections is beyond the accuracy of the graphical representation given in [3] even for the maximal value of F.

Note that the change both in the frequency of transition and the change in the dipole matrix element determine corrections to the intensity; however, they cannot be separated from one another in higher orders (see, e.g., (42)). Nevertheless, numerical calculations show that corrections to the matrix element make a significant, and sometimes a major, contribution to the correction to the line intensity. For transitions from Rydberg states with $n \gg 1$, the contribution of corrections to the matrix elements of the corresponding order for intensities is the major one as is seen from (44)–(47) for the third order and was also noted for the first and second orders in [4, 5].

4. CONCLUSIONS

The main results of this study are relations (11)–(14) for the coefficients of expansion (10) of the wave function of a hydrogen-like atom in the Rayleigh–Schrödinger series in powers of the electric field. Every term of this series is a superposition of the Coulomb Sturm functions in the parabolic coordinates for the Schrödinger equation of the hydrogen-like atom.

Corrections to dipole matrix elements and intensities of the radiation transitions between the Stark sublevels, which are induced by the field, are represented in the form of asymptotic series (37) and (38) in powers of the field F. For the Lyman and Balmer series, analytic expressions are obtained and numerical values of the third and fourth orders are presented. These corrections are polynomials of the quantum numbers of the upper state and sometimes can make a significant contribution to the matrix element of the radiation transition in the vicinity of the ionization threshold.

Notwithstanding the fact that the perturbation theory series for the Stark effect on atoms are asymptotic [1, 2, 6, 12], the data obtained for the matrix elements and intensities of the Lyman and Balmer series do not systematically confirm this fact, since we have for the

majority of the radiation π- and σ-polarized transitions that

$$\frac{r^{(3)}_{nqm\to n'q'm'}}{r^{(2)}_{nqm\to n'q'm'}} \sim \frac{r^{(2)}_{nqm\to n'q'm'}}{r^{(1)}_{nqm\to n'q'm'}},$$
$$\frac{\beta^{(3)}_{nqm\to n'q'm'}}{\beta^{(2)}_{nqm\to n'q'm'}} \sim \frac{\beta^{(2)}_{nqm\to n'q'm'}}{\beta^{(1)}_{nqm\to n'q'm'}}, \quad (50)$$

and large deviations (by an order of magnitude) from these relations to one side or the other occur equally frequently.

In most cases, the third order is sufficient to determine the applicability of perturbation theory. However, if the first- and third-order corrections are zero (e.g., for σ-transitions between dipoleless states), then the fourth order must be invoked to analyze the convergence of the perturbation series.

Calculations presented in this paper are important not only from the academic point of view but can also be used in practice. In particular, quantitative data for the variation of line intensity in a homogeneous electric field can be used to determine the constant component of the field in plasma. The calculation technique proposed in this paper for higher order corrections of perturbation theory for the Stark states allows a generalization for the interaction of an atom with a charged particle or a system of particles in the case when the inverse of the distance from the atomic nucleus is used as the small parameter for perturbation theory.

ACKNOWLEDGMENTS

This work was supported by the Ministry for Education of the Russian Federation, project no. 97-0-5.1-63, and INTAS, project no. 97-0369.

REFERENCES

1. N. Hoe, B. d'Elat, and G. Couland, Phys. Lett. A **85A**, 327 (1981).
2. R. D. Damburg and V. V. Kolosov, in *Rydberg States of Atoms and Molecules*, Ed. by R. F. Stebbings and F. B. Dunning (Cambridge Univ. Press, Cambridge, 1983; Mir, Moscow, 1985).
3. M. Bellermann, T. Bergeman, A. Haffmans, *et al.*, Phys. Rev. A **46**, 5836 (1992).
4. A. A. Kamenski and V. D. Ovsiannikov, J. Phys. B **33**, 491 (2000).
5. A. A. Kamenski and V. D. Ovsiannikov, J. Phys. B **33**, 5543 (2000).
6. S. P. Alliluev and I. A. Malkin, Zh. Éksp. Teor. Fiz. **66**, 1283 (1974) [Sov. Phys. JETP **39**, 627 (1974)].
7. H. A. Bethe and E. E. Salpeter, *Quantum Mechanics of One- and Two-Electron Atoms* (Academic, New York, 1957; Fizmatgiz, Moscow, 1960).
8. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 3: *Quantum Mechanics: Non-Relativistic Theory* (Nauka, Moscow, 1974; Pergamon, New York, 1977).
9. N. L. Manakov and L. P. Rapoport, Opt. Spektrosk. **33**, 988 (1972).
10. *Higher Transcendental Functions (Bateman Manuscript Project)*, Ed. by A. Erdelyi (McGraw-Hill, New York, 1953; Nauka, Moscow, 1966).
11. V. D. Ovsiannikov and S. V. Goossev, Phys. Scr. **57**, 506 (1998).
12. L. A. Bureeva and V. S. Lisitsa, *Perturbed Atom* (IzdAT, Moscow, 1997).
13. S. V. Bulyarskiĭ, N. S. Grushko, and A. V. Zhukov, Zh. Éksp. Teor. Fiz. **118**, 1092 (2000) [JETP **91**, 945 (2000)].

Translated by A. Klimontovich

Journal of Experimental and Theoretical Physics, Vol. 93, No. 1, 2001, pp. 55–62.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 1, 2001, pp. 67–75.

ATOMS, SPECTRA, RADIATION

Three-Wave Interference with Participation of Polaritons

A. V. Burlakov, Yu. B. Mamaeva, A. N. Penin*, and M. V. Chekhova

Moscow State University, Moscow, 119899 Russia
**e-mail: postmast@qopt.phys.msu.su*

Received February 14, 2001

Abstract—Second-order interference with respect to the field strength during spontaneous parametric light scattering was studied under the conditions of absorption at the frequencies of idler (polariton) modes. An expression is derived that describes the scattered light intensity as a function of the scattering angle for an arbitrary shape of the nonlinear interaction region. The interference was experimentally studied for the light scattering on polaritons in a nonlinear crystal (lithium iodate, iodic acid), with a double slit placed into a pumping beam in front of the crystal. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Spontaneous parametric down conversion (SPDC) is a coherent scattering process whereby the pumping photon decays into two photons (biphoton) of lower frequencies [1]. At present, this effect is used in the two fields: quantum optics and spectroscopy.

In quantum optics, the main attention is paid to the transformation of a biphoton field obtained as a result of the SPDC process and to the use of its nonclassical features. A large variety of experiments was reported in which the interference was observed using the second and fourth field moments (see [2] and references therein). These experiments usually reveal a significant difference between the quantum and classical descriptions of the interference. In recent years, biphoton fields have been attracting the attention of researchers engaged in methodological problems of quantum theory [3] and interested in the use of such fields in devices of quantum cryptography [4], quantum teleportation [5], etc.

In the spectroscopic applications, the SPDC phenomenon is used to study the properties of a substance featuring the parametric scattering. For this purpose, the experimental conditions are selected so as to provide that the frequency of one (signal) photon falls within the visible range in the region of transparency of the scattering substance, whereas the region of conjugated (idler) photon occurs in the IR range. The response is measured in the form of the frequency-angular spectra of the scattered signal radiation from which the dispersion characteristics of the medium in the IR range are determined such as dielectric permittivity, second- and third-order nonlinear susceptibility, etc. [6]. The main advantages of the SPDC spectroscopy are the simplicity of experimental implementation and the possibility of obtaining information in a spectral region near the optical phonon frequencies, which is difficult to access by other methods [7]. In this region, it is necessary to take into account a mixed electromagnetic-mechanical nature of the idler waves (polaritons), which does not allow the scattered field to be represented as a pair of free correlated photons [8].

This study was devoted to a situation in which the approaches of quantum optics and SPDC spectroscopy are closely related. The second order interference with respect to the signal field strength, in which case the parametric scattering involves polaritons, was studied by theoretical and experimental methods. An analysis of this phenomenon reveals a relationship between the concepts and methods of quantum optics and SPDC spectroscopy and provides a basis for the development of new methods of the nonlinear spectroscopy [9].

The structure of this paper is as follows. Section 2 briefly outlines the main definitions and approaches used for the description of interference in quantum optics. Section 3 is devoted to the three-wave second-order interference with respect to the field strength under the condition that the scattering substance is nontransparent for the idler (polariton) modes. A general expression is derived for the angular distribution of the scattered signal intensity and the angular profile of the signal line is described under the experimental conditions studied. Section 4 presents the results of experimental investigation of the three-wave interference.

2. THREE-WAVE INTERFERENCE DURING SPONTANEOUS PARAMETRIC LIGHT SCATTERING

The SPDC phenomenon is interpreted as the coherent decay of a pumping photon having the frequency ω_p into two photons, signal and idler with the frequencies ω_s and ω_i, respectively, such that ω_s, $\omega_i < \omega_p$ (it is conventionally adopted that $\omega_i \le \omega_s$). A phenomenological description of the SPDC process is based on the use of

an effective interaction Hamiltonian [10]

$$\hat{H}(t) = \int_V d^3 r \chi_{\text{eff}}(\mathbf{r}) E_p^{(+)}(\mathbf{r}, t) \times \hat{E}_s^{(-)}(\mathbf{r}, t) \hat{E}_i^{(-)}(\mathbf{r}, t) + \text{H.c.}, \quad (1)$$

where E_p is the pumping field (assumed to be a preset classical field); $\hat{E}_s$ and $\hat{E}_i$ are the field operators of the signal and idler waves, respectively (superscripts "+" and "–" denote the positive and negative frequency parts); $\chi_{\text{eff}} = \hat{\chi}^{(2)} \vdots \mathbf{e}_p \mathbf{e}_s \mathbf{e}_i$ ($\mathbf{e}_\sigma$ are the unit polarization vectors of the corresponding waves, $\sigma = p, s, i$, and the sign $\vdots$ denotes tensor convolution). In the first-order perturbation theory, the wavefunction of the scattered field is as follows:

$$|\psi\rangle = |vac\rangle + \sum_{\mathbf{k}_i, \mathbf{k}_s} F(\mathbf{k}_s, \mathbf{k}_i) |1\rangle_s |1\rangle_i, \quad (2)$$

where $|1\rangle_s |1\rangle_i$ is a two-photon Fock state in the s, i modes. The $F(\mathbf{k}_s, \mathbf{k}_i)$ function can be interpreted as a shape of the biphoton wave packet in the eight-dimensional (8D) space $\omega \otimes \mathbf{k}$ ($\mathbf{k}_\sigma$ are the wavevectors of the interacting waves, $|\mathbf{k}_\sigma| = \omega_\sigma n(\omega_\sigma)/c$). The Fourier image determines distribution of the biphoton amplitude in the 8D space $\mathbf{r} \otimes t$. In the most general form, $F(\mathbf{r}, t)$ can be written as follows:

$$F(\mathbf{r}_s, t_s; \mathbf{r}_i, t_i) \propto \int_t^{t+r} dt' \int_V d\mathbf{r} \chi_{\text{eff}}^{(2)}(\mathbf{r}) E_p(\mathbf{r}, t') \times D_s(\mathbf{r}_s, t_s; \mathbf{r}, t') D_i(\mathbf{r}_i, t_i; \mathbf{r}, t'). \quad (3)$$

Here, D_s and D_i are the propagators describing propagation of the signal and idler waves; V and τ are the volume and time of interaction, respectively. It is assumed that the substance in which the scattering takes place is transparent at the frequencies of the pumping radiation and the signal and idler waves. The probability of detecting a biphoton, that is, simultaneous counts in the two ideal detectors tuned to the $\mathbf{k}_s$ and $\mathbf{k}_i$ modes, is determined by the square modulus of the biphoton amplitude [10]:

$$P_c(\mathbf{k}_s, \mathbf{k}_i) = |F(\mathbf{k}_s, \mathbf{k}_i)|^2. \quad (4)$$

The probability that a photoresponse would appear in the detector tuned to the signal radiation can be calculated as

$$P_s(\mathbf{k}_s) = \int d\mathbf{k}_i P_c(\mathbf{k}_s, \mathbf{k}_i). \quad (5)$$

Proceeding from expression (3), we may suggest several ways for influencing the space-time structure of the biphoton field: (i) variation of the spatial distribution of the quadratic susceptibility $\chi_{\text{eff}}^{(2)}(\mathbf{r})$ [11]; (ii) modulation of the spatial distribution of the pumping field amplitude $E_p(\mathbf{r})$ [12]; (iii) using a special form of the propagators $D_{i,s}$.

The latter method, widely employed in quantum optics, provides for the simplest transformation of the biphoton wave packet in the course of its propagation. The scattered radiation frequency usually falls within the visible (or near IR) spectral range in the region of transparency of the nonlinear crystals employed. The propagator transformations are performed with the aid of linear optical (including polarization) devices such as interferometers, phase shifters, delay lines, etc.

The inhomogeneous spatial distributions of nonlinearity $\chi_{\text{eff}}^{(2)}(\mathbf{r})$ and pumping $E_p(\mathbf{r})$ equally affect the angular distribution of the scattered radiation intensity in the diffractionless approximation ($L/k_p a^2 \ll 1$, L is the nonlinear layer thickness in the direction of the pumping wavevector, a is the characteristic size of transverse inhomogeneities). Not specifying the shape of the spatial inhomogeneities in the scattering volume, it is possible to show that the angular profile of the scattering line is determined by the Fourier image of these inhomogeneities [13]:

$$R_s(\mathbf{k}_s) = \int dk_i \left| \int f(\mathbf{r}) \exp(i \Delta\mathbf{k} \cdot \mathbf{r}) d\mathbf{r} \right|^2, \quad (6)$$

where $\Delta_\mathbf{k} = \mathbf{k}_p - \mathbf{k}_s - \mathbf{k}_i$ is the wave detuning, $f(\mathbf{r}) = \chi_{\text{eff}}^{(2)}(\mathbf{r}) E_p(\mathbf{r})$, and $R_s(\mathbf{k}_s) = P_s(\mathbf{k}_s)/\tau$ is the rate of photoresponse counts proportional to the SPDC intensity. For example, in the scattering volume having the shape of a plane infinite layer with a thickness L oriented perpendicularly to the pumping wavevector, the frequency-angular profile of the line is described by the function

$$R_s(\mathbf{k}_s) = \left[\frac{\sin(\Delta\mathbf{k} L/2)}{\Delta\mathbf{k} L/2} \right]^2,$$

which is characteristic of the parametric processes. If the scattering takes place in several regions (e.g., in N crystals), the two-photon probability amplitudes add as

$$F = \sum_{n=1}^{N} F_n \exp(i \Delta\varphi_n),$$

where $\Delta\varphi_n = \varphi_n^s + \varphi_n^i - \varphi_n^p$ is the three-wave phase shift.

The line shape is determined by the phase gained in all three frequencies participating in the scattering process, which allows the process to be considered as the three-wave interference. Use of the interaction volume of a complicated shape leads to modification of the frequency-angular spectra of the scattered field, which acquires an additional fine structure of the interference nature.

It should be noted that the term interference implies here a quantum-mechanical interference of the probability amplitudes of the biphoton field scattering into

preset modes (determined by the boundary conditions), rather than as the classical field superposition in space-time.

3. THREE-WAVE INTERFERENCE IN THE PRESENCE OF ABSORPTION

Consider a scattering substance which is nontransparent only for the idler radiation. For the pumping and signal waves, the absorption is negligibly small. A spatial distribution of the photon count rate in the detector, proportional to the radiation intensity in the signal field mode, is described by the function

$$R_s(\mathbf{k}_s) = \langle E_s^{(-)}(\mathbf{k}_s) E_s^{(+)}(\mathbf{k}_s) \rangle. \quad (7)$$

Proceeding from the Hamiltonian (1) and taking into account that the idler wave field in the substance is not free, we obtain the following expression within the framework of the first-order perturbation theory:

$$R_s(\mathbf{k}_s) \int d\mathbf{k}_i \iint_V d\mathbf{r}' d\mathbf{r}'' f(\mathbf{r}') f^*(\mathbf{r}'') \times G''(\mathbf{k}_i, \omega_p - \omega_s) \exp[i\boldsymbol{\Delta}_\mathbf{k} \cdot (\mathbf{r}' - \mathbf{r}'')]. \quad (8)$$

Here $G''(\mathbf{k}_i, \omega)$ is the anti-Hermitian part of the equilibrium Green function of the nonlinear medium, and $\mathbf{k}_i$ is the wavevector of the idler radiation. In an isotropic medium, the anti-Hermitian part of the Green function of the equilibrium idler wave field in the region of not very strong absorption (where the dielectric permittivity ε obeys the inequality $\mathrm{Im}\varepsilon \ll \mathrm{Re}\varepsilon$) is as follows [8]:

$$G''(k_i, \tilde{\omega}) = \frac{4\pi}{c^2 k(\tilde{\omega})} \tilde{\omega}^2 \frac{\alpha/2}{[k_i - k(\tilde{\omega})]^2 + (\alpha/2)^2}, \quad (9)$$

where α is the absorption coefficient and $\tilde{\omega} = \omega_p - \omega_s$. In the case of negligibly small absorption at the idler wave frequency ($\alpha \longrightarrow 0$),

$$G''(k_i, \tilde{\omega}) \propto \delta(k_i - k(\tilde{\omega})) \quad (10)$$

and expression (8) transforms into relationship (6) valid in the absence of absorption.

Let us proceed from the general formula (8) to the angular distribution of the signal radiation at a fixed frequency ω_s. For convenience, the vectors $\mathbf{k}_p$, $\mathbf{k}_s$, and $\mathbf{k}_i$ are assumed to lie in the same plane (for example, *xy*), with $\mathbf{k}_p$ parallel to the *y*. The angular distribution of the scattered signal intensity is observed also in the *xy* plane (Fig. 1). Substituting (9) into (8) and integrating with respect to $|\mathbf{k}_i|$, we obtain

$$R_s(\theta_s) = \int d\theta_i F_x(\theta_s, \theta_i) F_y(\theta_s, \theta_i),$$

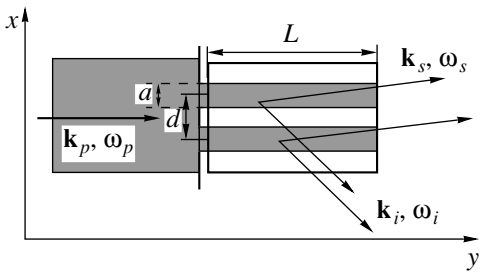

Fig. 1. A schematic diagram showing configuration of the scattering region. A pumping radiation is incident onto a nonlinear crystal of length *L*. A nontransparent screen placed in front of the crystal has two slits with equal widths *a* and a distance *d* between their centers.

where

$$F_x(\theta_s, \theta_i) = \iint_V dx' dx'' f(x') f^*(x'') \times \exp\left[i(x' - x'')\Delta_x(\theta_s, \theta_i) - \frac{\alpha}{2}|x' - x'| \sin\theta_i\right], \quad (11)$$

$$F_y(\theta_s, \theta_i) = \iint_V dy' dy'' f(y') f^*(y'') \times \exp\left[i(y' - y'')\Delta_y(\theta_s, \theta_i) - \frac{\alpha}{2}|y' - y'| \cos\theta_i\right],$$

$\Delta_{x,y}$ is the projection of the wave detuning onto *x* and *y* axes; and θ_s and θ_i are the angles of scattering of the signal and idler waves.

Let *L* and *a* be the characteristic dimensions of inhomogeneities in the interaction region along the *y* and *x* axes. We will consider two typical situations, paying major attention to the transverse (relative to the pumping wavevector) arrangement of inhomogeneities in the scattering volume (immediately related to the experiment).

(a) Strong absorption ($\alpha L \gg 1$, $\alpha a \gg 1$). In this case, F_x and F_y have the form of Lorentz curves (as the functions of θ_i at a fixed θ_s) with comparable widths. Each function has a maximum at θ_i such that the corresponding wave detuning $\Delta_{x,y} = 0$. The angular profile of the signal line also represents a Lorentz curve

$$R_s(\theta) = \frac{\sigma_s^2}{\theta^2 + \sigma_s^2} \quad (12)$$

with the width

$$\sigma_s(\tilde{\omega}) = \frac{\alpha(\tilde{\omega})}{2k_p \sin(\theta_{i0}(\tilde{\omega}))}. \quad (13)$$

Here, $\theta \equiv \theta_s - \theta_{s0}$; θ_{s0} and θ_{i0} are the scattering angles of the signal and idler waves for which the condition of phase synchronism $\boldsymbol{\Delta}_\mathbf{k} = 0$ is strictly fulfilled. Expression (12) is well known in SPDC spectroscopy [14] and

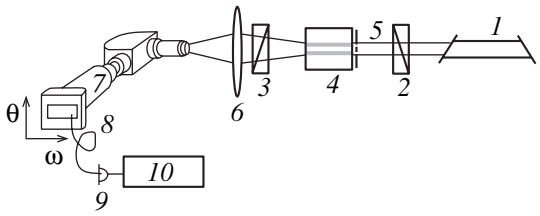

Fig. 2. A schematic diagram of the experimental setup: (*1*) Ar^+ laser ($\lambda = 488$ nm); (*2*, *3*) Glan prisms; (*4*) nonlinear crystal ($LiIO_3$, α-HIO_3); (*5*) mask with two slits ($a = 85$ μm, $d = 195$ μm); (*6*) objective lens; (*7*) spectrograph (ISP-51); (*8*) optical fiber; (*9*) detector; (*10*) photon counter.

is used in this very form in the measurements of the frequency dependence of the absorption coefficient in the spectral region of the upper polariton branch [15]. In this region, the idler radiation frequency $\tilde{\omega}$ is greater than the maximum frequency of the longitudinal optical mode in the crystal lattice ($\tilde{\omega} > \omega_{LO}^{\max}$).

(b) $\alpha L \gg 1$, $\alpha a \approx 1$. In this case, the absorption is sufficiently high for the longitudinal size of the medium be insignificant, but not as high as to ignore the transverse inhomogeneity size.

The $F_y(\theta_i)$ function is still Lorentzian, while $F_x(\theta_i)$ is determined by taking an exact integral along the x axis. For an arbitrary shape of transverse inhomogeneities, expression (11) leads to the following angular profile of the line:

$$R_s(\theta) = \iint dx'dx'' f(x') f^*(x'') \times \exp\left[ik_p(x'-x'')\theta - \frac{\alpha}{2\sin\theta_{i0}}|x'-x''|\right]. \quad (14)$$

Therefore, the effect of transverse inhomogeneities in the interaction volume on the final shape of the signal line decreases with increasing absorption of the idler mode. The quantity $l_x = \alpha^{-1}\sin\theta_{i0} = \tilde{\alpha}_x^{-1}$ can be interpreted as the polariton "free-path length" in the transverse direction.

Let the interaction region represent two macroscopic volumes set as follows (Fig. 1):

$$f(\mathbf{r}) = \chi_{\text{eff}}^{(2)} E_{0p},$$
$$x \in [-(d+a)/2, -(d-a)/2] \cup [(d-a)/2, (d+a)/2], \quad (15)$$
$$y \in [0, L], \quad z \in (-\infty, \infty),$$
$$f(\mathbf{r}) = 0, \text{ in the rest of the space.}$$

This case is convenient both for theoretical analysis and for experimental verification, for example, using a nontransparent screen with two slits placed in front of a nonlinear crystal. Calculating the integral in (14) taking into account conditions (15), we obtain

$$R_s(\theta) = 2R_s^a(\theta, \tilde{\alpha}_x) \times \{1 + \exp(-\tilde{\alpha}_x d)\cos[\Delta_x(\theta)d]\} + \Delta R_s^a. \quad (16)$$

This expression has a simple physical meaning. The function $R_s^a(\theta, \tilde{\alpha}_x)$ describes the angular profile of the scattering line for a layer with the transverse size a (for a single open slit with the width a) and the length L at a arbitrary absorption coefficient $\tilde{\alpha}_x$. In the general case, the form of this profile is intermediate between a Lorentz contour and $\text{sinc}^2(x)$. For a strong absorption, the line shape corresponds to a Lorentz curve, while for $\tilde{\alpha}_x \longrightarrow 0$, the line is described by the function $(\sin x/x)^2 \equiv \text{sinc}^2 x$ [8]:

$$R_s^a(\theta, \tilde{\alpha}_x) = \frac{\tilde{\alpha}_x a}{\tilde{\alpha}_x^2 + \Delta_x^2(\theta)} + \{[\Delta_x^2(\theta) - \tilde{\alpha}_x^2][1 - \exp(-\tilde{\alpha}_x a)\cos(\Delta_x(\theta)a)] - 2\tilde{\alpha}_x a \exp(-\tilde{\alpha}_x a)\sin(\Delta_x(\theta)a)\}[\tilde{\alpha}_x^2 + \Delta_x^2(\theta)]^{-2}. \quad (17)$$

In the expression (16), the factor in braces contains a modulation term proportional to $\cos[\Delta_x(\theta)d]$, which is related to the interference. The amplitude of this term is determined by the exponential factor $\exp(-\tilde{\alpha}_x d)$, which depends on the relationship between the absorption level and the slit spacing. For the classical interference in the Young scheme, the degree of the interference pattern manifestation is determined by the ratio of the radiation coherency radius to the slit spacing [16]. The second term ΔR_s^a exhibits an oscillating character and appears only in the case of distributed absorption. Estimates show that the second term (for the values of realistic parameters) is smaller by two orders of magnitude than the first term amplitude $R_s^a(\theta, \tilde{\alpha}_x)$ and, hence, can be excluded from consideration.

This, expression (16) shows that the idler wave absorption must affect both the envelope shape and the degree of the interference pattern manifestation. The affect of absorption in the second-order interference during SPDC was studied by Wang *et al.* [17], but only in the case when an element semitransparent for the idler wave was situated between two regions emitting the biphoton fields. In the next section, the results reported in [17] are compared to our conclusions.

4. EXPERIMENTAL RESULTS AND DISCUSSION

Figure 2 shows a schematic diagram of the experimental setup representing a traditional SPDC spectrograph. The frequency-angular spectra of the scattered signal field intensity are analyzed by the cross dispersion technique [18], whereby lens *6* (with a focal distance of $F = 85$ mm) is situated so that its focus

would occur in the plane of the entrance slit of the ISP-51 spectrograph. A two-dimensional pattern of intensity distribution in the angle–wavelength (θ_s–λ_s) coordinates formed in the exit slit plane of the spectrograph was recorded using photographic or photoelectric methods. The signal radiation wavelength was selected within the range λ_s = 600–700 nm in order to ensure that the conjugated idler wavelength would fall within the region of λ_i = 2–5 μm (ν_i = 2000–5000 cm^{-1}). Thus, the idler radiation frequency is close to the frequencies of the crystal lattice eigenmodes, which corresponds to the case of scattering on polaritons.

Two macroscopic interaction regions were formed with the aid of a nontransparent screen with two slits situated in front of a lithium iodate ($LiIO_3$) or iodic acid (α-HIO_3) crystal. When the crystal is situated in the near diffraction zone, the spatial distribution of the pumping radiation inside the crystal is determined by the slit parameters. The slits had equal widths of a = 85 μm at a distance between their centers of d = 195 μm. The pumping was provided by an Ar^+ laser operating in a single transverse mode with λ_p = 514.5 nm at a power of up to 4 W. The transverse coherency radius of the pumping radiation was significantly greater than the slit spacing, so that the parametric scattering was coherently excited in both parallel layers (Fig. 1).

It must be noted that, in order to observe the nonclassical interference, it is necessary to satisfy a criterion relating the angular and spatial experimental parameters [12]. For the second-order interference with respect to the field strength, which is studied here, the corresponding criterion in the absence of absorption is as follows:

$$L\tan\theta_{i0} \gg d. \qquad (18)$$

The physical meaning of this criterion consists in that the idler radiation crosses the transverse inhomogeneities in the scattering volume, not going outside of this volume in the longitudinal direction. Owing to this, the SPDC in both emitting regions involves the same idler wave modes.[1] In our experiments, the crystal types, slit dimensions, and scattering angles were determined taking into account condition (18).

We have experimentally studied two cases. In the first (weak absorption), the transverse polariton free path length $l_x \sim \tilde{\alpha}_x^{-1}$ exceeded the size of transverse inhomogeneities in the scattering volume. In the second case, the polariton free path length was on the order of or smaller than the slit spacing: $l_x \leqslant d - a$.

(a) Weak absorption (α < 10 cm^{-1}, L = 18 mm). The nonlinear medium was represented by a $LiIO_3$ crystal. Here, the interference is manifested by a clearly pronounced splitting of the angular profile in the range 1800 < ν_i < 3000 cm^{-1} (the upper frequency was limited by the condition of spatial synchronism of the crystal employed). Figure 3 shows the angular profile measured for a signal wavelength of 588 nm (conjugated polariton wavelength λ_i = 2.869 μm, ν_i = 3485 cm^{-1}) obtained by the electronic scanning technique.[2] The solid curve in Fig. 3 shows the result of a numerical approximation based on Eq. (16). The absorption coefficient was estimated at α < 0.3 cm^{-1}, which does not contradict to the data reported in [19], according to which α < 2 cm^{-1}. As the idler radiation frequency increased, the absorption coefficient remained constant to within the experimental uncertainty.

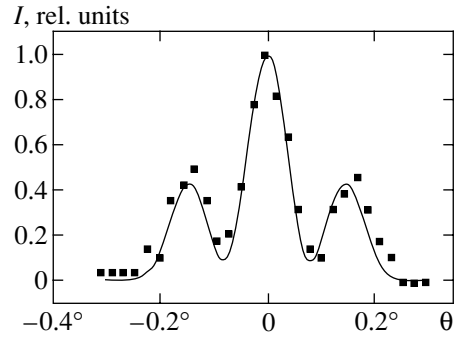

Fig. 3. Angular profile of the signal radiation intensity at the output of a nonlinear Young interferometer (scattering crystal, $LiIO_3$; λ_s = 588 nm; $\theta = \theta_s - \theta_{0s}$).

(b) Not too strong absorption (10 < α < 150 cm^{-1}, L = 26 mm). In this case, the nonlinear medium was represented by an α-HIO_3 crystal. As is well known, the coefficient of absorption of a polariton wave increases when the wave frequency approaches the crystal lattice eigenmode where $\mathrm{Im}\varepsilon(\omega)$ exhibits a significant grows. We have used this fact, selecting a polariton frequency variation range (3300 < ν_i < 5000 cm^{-1}) in the vicinity of the OH group stretching vibration frequency (ν_i^{OH} = 2950 cm^{-1}) for the α-HIO_3 crystal. The measurements were performed in the $Y(X, Z)Y + \Delta X$ scattering geometry.[3] The notation of axes (Fig. 1) coincides with crystallographic; for the α-HIO_3 crystal, the notations are selected so that the refractive indices in the visible range obey the inequality $n_x < n_y < n_z$. For the polariton frequency range selected, the angle between the polariton wavevector and the y axis varied within 30° < θ_i < 45°.

[1] Wang *et al.* [17] interpreted criterion (18) as the induced coherency effect.

[2] The detector (EG&G avalanche photodiode) operating in the photon count mode was moved by a step motor in the focal plane of the spectrograph. Amplified output pulses were fed to a discriminator and then to a counter circuit. Operation of the step motor and the counter was controlled by a computer.

[3] A standard notation used in the Raman scattering spectroscopy: indices outside the parentheses indicate directions of the incident and scattered light, indices in the parentheses indicate the polarization.

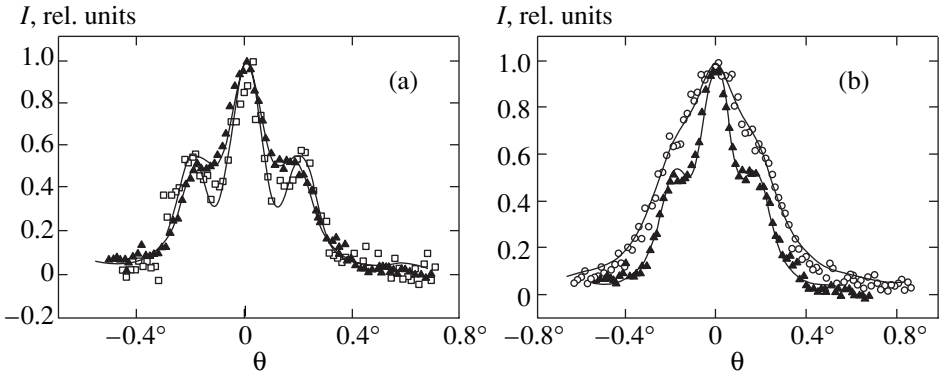

Fig. 4. Angular profiles of scattering in a nonlinear Young interferometer observed in a crystal of iodic acid (α-HIO_3) at various idler radiation frequencies in the vicinity of the resonance frequency ν = 2950 cm^{-1}: (□) ν_i = 4569 cm^{-1}, λ_s = 673 nm (a); (▲) ν_i = 4206 cm^{-1}, λ_s = 657 nm (a, b); (○) ν_i = 3357 cm^{-1}, λ_s = 622 nm (b).

Figure 4 shows three characteristic angular profiles measured for various signal radiation wavelength. As is seen, when the idler radiation frequency approaches a resonance with the crystal lattice eigenmode, the degree of manifestation of the interference pattern decreases. The solid curves in Fig. 4 show the results of theoretical calculations according to Eqs. (16) and (17). In order to study the behavior of the absorption coefficient at the idler wave frequencies, a series of such measurements was performed for various wavelengths and the absorption coefficient α was determined by numerical methods. The results of these calculations are presented in Fig. 5, where a solid curve shows the dispersion of α calculated within the framework of the oscillator model for the scattering geometry employed. The parameters of the OH bond oscillations were taken from [20]. Measured far from the resonance, the absorption corresponds to a "background" level (40–50 cm^{-1}) virtually independent of the scattering geometry [21].

The data presented in Fig. 5 show a good agreement between experiment and calculation, which is evidence of the validity of the employed approach. Figure 6 shows a fragment of the frequency-angular parametric scattering spectrum measured in the vicinity of the frequency of the OH group stretching vibrations.

The fact that the idler wave absorption affects the second- order interference pattern during SPDC was established [17] for a situation in which the absorption is concentrated in the region between scattering volumes. Two nonlinear crystals were arranged so that the idler radiation emitted from the first crystal would pass

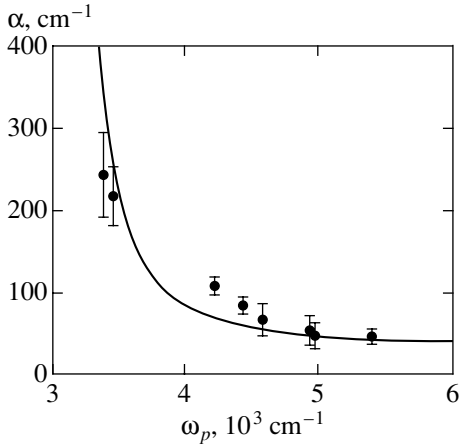

Fig. 5. A plot of the polariton wave absorption versus polariton frequency. The black points with error bars represent experimental data obtained by approximating the angular spectra (see Fig. 4). Sold curve shows the calculated absorption of α-HIO_3.

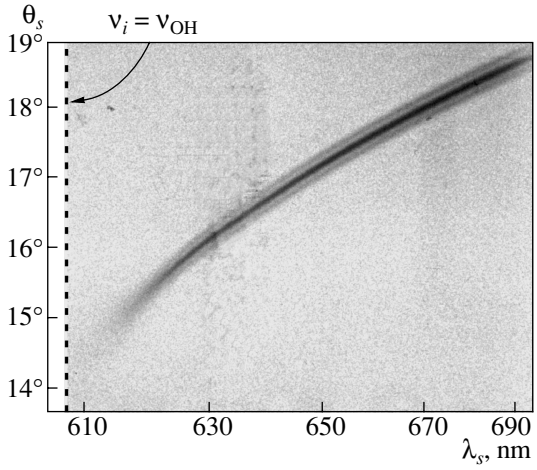

Fig. 6. A photograph of the upper polariton branch of the frequency-angular scattering spectrum of an α-HIO_3 crystal in the region of the OH stretching frequency (ν_{OH} = 2950 cm^{-1}). The measurements were performed using a nontransparent screen with two slits (a = 85 μm, d = 195 μm) placed into the pumping beam (λ_p = 514.5 nm) in front of the scattering crystal.

through the second crystal and system was adjusted so that the idler modes would completely coincide, which ensured that the condition (18) was obeyed. The parametric scattering was excited by a general coherent pumping. A coupling device with a preset transmission coefficient T was placed between the two crystals in the idler mode. The intensity of the conjugated signal radiation emitted from both crystals was studied (upon mixing in the coupler) as a function of the transmission at the idler frequency. It was found that the degree of the interference pattern manifestation varied linearly with the coupler transmission coefficient.

Following the considerations of Section 2, we may say that the experiments in [17] involved variation of the idler and signal field propagators during the parametric scattering. The biphoton field was "prepared" in both crystals and then subject to a certain action (Absorption in the common idler mode). In our experiments, the action took place both in the course of the biphoton field was "preparation" and during its propagation from one scattering region to another (effect of the polariton absorption). In other words, we deal with a distributed absorption, whereas in [17] the absorption was introduced in a localized manner into a free field of the idler wave (i.e., independently of the process of biphoton production). Thus, the approach developed in Section 3 is more general, including (as the limiting case) the conclusions derived in [17]. The exponential dependence of the interference pattern manifestation on the polariton absorption (see Eq. (16)) agrees with the linear dependence of this parameter on the transmission coefficient T, taking into account that (in the absorbing layer)

$$T \equiv \frac{I}{I_0} = \exp(-\tilde{\alpha}_x d).$$

At the same time, according to (17), the envelope of the scattering line observed in the presence of interference also carries information about absorption. This fact serves as basis for the method of polariton scattering spectroscopy [7, 8].

5. CONCLUSION

We have analyzed the effect of the distributed absorption at the idler mode frequency on the second-order interference with resect to the field strength during the spontaneous parametric light scattering. Expressions (16) and (17) were derived that describe the shape of the polariton scattering line for an arbitrary configuration of the scattering volume. An analysis of the dispersion of the absorption coefficient of a nonlinear crystal in the IR range can be performed using any one of the two methods.

1. The first method is a standard polariton scattering spectroscopy. The sample dimensions are selected so as to provide for a maximum intensity of scattering and to ensure the condition $a_0\alpha \gg 1$, where a_0 is the characteristic sample size. The shape of the scattering line is measured and analyzed.

2. The second method is offered by the "polariton interferometry." The size of the transverse inhomogeneity a is selected so that $a\alpha \approx 1$. For example, the scattering region configuration is determined by a nontransparent screen with two slits situated in the pumping beam in front of a sample. Here, both the line envelope and the degree of manifestation of the interference pattern are analyzed. It should be noted that this method can be used to study the dispersion law in substances not possessing a quadratic susceptibility [9, 17].

The effect of absorption in and reflection from the scattering volume boundaries on the shape of the parametric scattering line can be also studied using the method of scattering matrix [22]. However, application of this method to analysis of the parametric scattering line shape in the case of the scattering region of an arbitrary configuration is a very difficult task.

In conclusion, we should like to note an important methodological aspect of the problem considered above. Although more than three decades passed since the discovery of the phenomenon of spontaneous parametric light scattering, the problem of combining the two historically separated branches of this effect, quantum-optical and spectroscopic, is virtually not yet discussed in the literature. We hope that this paper will at least partly fill the gap.

ACKNOWLEDGMENTS

The authors are grateful to S.P. Kulik for numerous fruitful discussions.

The study was supported by the Russian Foundation for Basic Research (project nos. 99-02-16419 and 00-15-96541) within the framework of the "Integration" (Fundamental Optics and Spectroscopy) Program and by the "Fundamental Metrology" Program.

REFERENCES

1. D. N. Klyshko, Pis'ma Zh. Éksp. Teor. Fiz. **6**, 490 (1967) [JETP Lett. **6**, 23 (1967)].
2. L. Mandel, Rev. Mod. Phys. **71**, S274 (1999).
3. D. N. Klyshko, Usp. Fiz. Nauk **164**, 1188 (1994) [Phys. Usp. **37**, 1097 (1994)]; Usp. Fiz. Nauk **166**, 613 (1996) [Phys. Usp. **39**, 573 (1996)]; A. Zeilinger, Rev. Mod. Phys. **71**, S288 (1999).
4. D. S. Naik, C. G. Peterson, A. G. White, *et al.*, Phys. Rev. Lett. **84**, 4733 (2000); W. Tittel, J. Brendel, H. Zbinden, and N. Gizin, Phys. Rev. Lett. **84**, 4737 (2000); T. Jennewein, C. Simon, G. Weihs, *et al.*, Phys. Rev. Lett. **84**, 4729 (2000).
5. C. H. Bennet, G. Brassard, C. Crepeau, *et al.*, Phys. Rev. Lett. **70**, 1895 (1993); D. Bouwnmeester, J.-W. Pan, M. Eibl, *et al.*, Nature **390**, 575 (1997); D. Boshi, S. Branca, F. De Martini, *et al.*, Phys. Rev. Lett. **80**, 1121 (1998); A. Furusawa, J. L. Sorensen, S. L. Braunstain, *et al.*, Science **282**, 706 (1998).

6. O. A. Doil'nitsyna and Yu. N. Polivanov, Tr. Inst. Obshch. Fiz. Akad. Nauk SSSR **2**, 11 (1986).
7. Yu. N. Polivanov, Usp. Fiz. Nauk **126**, 185 (1978) [Sov. Phys. Usp. **21**, 805 (1978)].
8. D. N. Klyshko, *Photons and Nonlinear Optics* (Nauka, Moscow, 1980).
9. A. V. Burlakov, S. P. Kulik, A. N. Penin, and M. V. Chekhova, Zh. Éksp. Teor. Fiz. **113**, 1991 (1998) [JETP **86**, 1090 (1998)]; D. Yu. Korystov, S. P. Kulik, and A. N. Penin, Kvantovaya Élektron. (Moscow) **30**, 922 (2000).
10. A. V. Belinsky and D. N. Klyshko, Laser Phys. **4**, 663 (1994).
11. A. L. Aleksandrovskiĭ, G. Kh. Kitaeva, S. P. Kulik, and A. N. Penin, Zh. Éksp. Teor. Fiz. **106**, 993 (1986) [Sov. Phys. JETP **63**, 613 (1986)].
12. A. V. Burlakov, D. N. Klyshko, S. P. Kulik, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **65**, 20 (1997) [JETP Lett. **65**, 19 (1997)]; E. J. S. Fonseca, C. H. Monken, and S. Padua, Phys. Rev. Lett. **82**, 2868 (1999).
13. A. V. Burlakov, M. V. Chekhova, D. N. Klyshko, *et al.*, Phys. Rev. A **56**, 3214 (1997).
14. D. N. Klyshko, Zh. Éksp. Teor. Fiz. **55**, 1006 (1968) [Sov. Phys. JETP **28**, 522 (1968)].
15. B. N. Mavrin and Kh. E. Sterin, in *Modern Problems of Raman Spectroscopy*, Ed. by M. M. Sushchinskiĭ (Nauka, Moscow, 1978), p. 303.
16. M. Born and E. Wolf, *Principles of Optics* (Pergamon, Oxford, 1969; Nauka, Moscow 1973), p. 547.
17. L. J. Wang, X. Y. Zou, and L. Mandel, Phys. Rev. A **44**, 4614 (1991).
18. D. N. Klyshko, A. N. Penin, and B. F. Polkovnikov, Pis'ma Zh. Éksp. Teor. Fiz. **11**, 11 (1970) [JETP Lett. **11**, 5 (1970)].
19. K. I. Avdienko, S. V. Bogdanov, and S. M. Arkhipov, *Lithium Iodate. Growth of Crystals, Their Properties and Applications* (Nauka, Novosibirsk, 1980), p. 144.
20. T. V. Laptinskaya, Candidate's Dissertation in Mathematical Physics (Mosk. Gos. Univ., Moscow, 1986).
21. T. V. Laptinskaya, A. G. Mikhaĭlovskiĭ, and A. N. Penin, Vestn. Mosk. Univ., Ser. 3: Fiz., Astron. **26**, 62 (1985).
22. G. Kh. Kitaeva, D. N. Klyshko, and I. V. Taubin, Kvantovaya Élektron. (Moscow) **9**, 560 (1982).

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 93, No. 1, 2001, pp. 63–70.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 1, 2001, pp. 76–84.

ATOMS, SPECTRA, RADIATION

Rectification of the Dipole Force in a Monochromatic Field Created by Elliptically Polarized Waves

O. N. Prudnikov*, A. V. Taĭchenachev, A. M. Tumaĭkin, and V. I. Yudin

Novosibirsk State University, ul. Pirogova 2, Novosibirsk, 630090 Russia
**e-mail: llf@admin.nsu.ru*

Received February 19, 2001

Abstract—The rectification of the force of induced light pressure in laser fields formed by elliptically polarized running waves in zero magnetic field is considered. Explicit analytic expressions for the induced and spontaneous forces of light pressure exerted on a stationary atom are obtained for two classes of closed optical transitions: $J_g = J \longrightarrow J_e = J+1$ and $J_g = J \longrightarrow J_e = J$ (J is half-integral), where J_g and J_e are the total angular momenta of the ground and excited energy levels. It is shown that the ellipticity of waves is the necessary condition for the emergence of the rectification of the induced force in a monochromatic field. The optimal parameters of the field and the maximum rectification coefficient are calculated for a number of optical transitions. The dependence of the rectified force on the velocity is investigated analytically and numerically for the simplest $1/2 \longrightarrow 1/2$ transition. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The force of induced light pressure (dipole force) in high-intensity laser fields considerably exceeds the force of spontaneous light pressure, which is limited by the spontaneous decay rate γ:

$$F_{\text{dip}} \sim \hbar k\Omega \gg \hbar k\gamma/2,$$

where Ω is the Rabi frequency and $\hbar k$ is the photon momentum. This makes it possible to use the dipole force for controlling the translational motion of atoms (acceleration, deceleration, and deflection of atomic beams and the formation of deep potential wells) [1, 2]. In many cases, the oscillatory nature of the coordinate dependence of the induced force is the main obstacle in obtaining a noticeable effect. For this reason, it is important to analyze situations in which the effect of rectification of the dipole force takes place; i.e., a force component appears, having a constant sign over spacial scales of the wavelength of light λ.

At present, various methods of rectification of the force in stationary fields have been developed theoretically and verified experimentally. In their pioneering works, Kazantsev and Krasnov [3–5] proved (using the model of a two-level atom) that the induced component of the light pressure force can be rectified in a bichromatic field. In this case, use is made of two standing waves, one of which (having a moderate intensity and detuning) exerts a force on the atom, while the other wave (having a high intensity and considerable detuning from resonance) leads to spatial modulation of effective detuning for the first field (due to optical shifts of atomic levels). For an appropriate choice of parameters, effective detuning oscillates synchronously with the intensity of the first (resonance) field, and the sign reversal of the intensity gradient is compensated by the sign reversal of detuning, leading to rectification which was detected in experiments from the deviation of the atomic beam [6, 7]. Later, Javanainen [8] demonstrated for a three-level Λ-atom that the requirement of a considerable difference in the detuning and intensities of the waves is not necessary. The mechanism of the emergence of a constant-sign component of the dipole force for approximately equal detunings is based on a spatially nonuniform redistribution of atoms over the sublevels of the ground state due to the modulation of the corresponding rates of optical pumping. These two factors (spatial nonuniformity of optical shifts of atomic levels and the rates of optical pumping between these levels) form the basis of various versions of rectification of the dipole force in a bichromatic field in three-level systems with the Λ- [9–11], V- [12], or cascade-type [13, 14] configurations of the levels as well as in a monochromatic nonuniformly polarized field upon the application of a magnetic field leading to the splitting of the Zeeman sublevels [15, 16]. The latter case is equivalent (in the resonance approximation) to the interaction of a multilevel atom with polychromatic radiation; this becomes evident when the quantization axis is chosen along the magnetic field.

It turns out, however, that the rectification effect also emerges in a monochromatic field in zero magnetic field, when atoms with the degenerate ground state interact with the laser field formed by elliptically polarized waves. Namely, we analyzed a one-dimensional field configuration formed by counterpropagating plane waves of the same amplitude and ellipticity ε_0 ($|\tan\varepsilon_0|$ is equal to the ratio of the semiaxes of the polarization ellipse, while the sign of $\sin(2\varepsilon_0)$ determines the direc-

tion of rotation of the electric field vector). The principal axes of the ellipses of counterpropagating waves are oriented at an angle θ relative to each other. For such an ε–θ–ε configuration, the circular components $\sigma_\pm$ are standing waves shifted in space relative to each other by the phase angle θ (see formula (14)). It was shown in [8] for a three-level Λ-atom that the differences in the spatial phases and intensities of standing waves are necessary conditions for the rectification of the dipole force for equal detunings (monochromatic laser field and zero Zeeman splitting).

The necessary conditions for rectification ($\varepsilon_0 \neq 0$, $\pm\pi/4$ and $\theta \neq 0, \pm\pi/2$) can be established from a general symmetry analysis of the force of light pressure. Indeed, it can easily be verified that for the ε–θ–ε configuration, the following relation holds for the force averaged over the spatial period for arbitrary ε_0 and θ:

$$\langle F\rangle(v, \delta) = -\langle F\rangle(v, -\delta), \tag{1}$$

where v is the velocity of an atom and δ is the detuning from resonance. This relation just indicates the possibility of rectification (for $v = 0$) of the dipole force since the force of spontaneous light pressure for $v = 0$ has an even dependence on detuning. However, for special values of the parameters ($\varepsilon_0 = 0, \pi/4$ or for $\theta = 0, \pm\pi/2$), another relation, which also follows from symmetry considerations,

$$\langle F\rangle(v, \delta) = \langle F\rangle(-v, -\delta), \tag{2}$$

indicates, on the contrary, the even dependence on detuning for $v = 0$. Consequently, the average force in this case is equal to zero. Our calculations show that for the optical transitions $J_g \longrightarrow J_e$ (J_g and J_e are the total angular momenta for the ground and excited levels), the necessary conditions for $J_g \geq 1/2$ are also sufficient conditions.

2. STEADY-STATE FORCE OF LIGHT PRESSURE

Let us consider the resonant interaction of a one-dimensional (directed along z) monochromatic field

$$\mathbf{E}(z, t) = \mathbf{E}(z)e^{-i\omega t} + \text{c.c.} \tag{3}$$

with an atom whose degenerate ground and excited states form the dipole- allowed transition $J_g \longrightarrow J_e$. The complex-valued vector amplitude of the field can be written as the product

$$\mathbf{E}(z) = Ee^{i\Phi}\mathbf{e}, \tag{4}$$

where E is the real-valued amplitude, Φ is the phase, and

$$\mathbf{e} = \sum_{q = 0, \pm 1} e^q \mathbf{e}_q$$

is the unit complex-valued vector of elliptic polarization, e^q being its contravariant components in the cyclic basis $\{\mathbf{e}_0 = \mathbf{e}_z;\ \mathbf{e}_{\pm 1} = \mp(\mathbf{e}_x \pm i\mathbf{e}_y)/\sqrt{2}\}$. The explicit dependence of E, Φ, and $\mathbf{e}$ on the coordinate is not specified here to simplify the notation. Without any loss of generality, we assume that the real and imaginary components of $\mathbf{e}$ are orthogonal; i.e., $\text{Im}(\mathbf{e} \cdot \mathbf{e}) = 0$. For such a choice of the phase, the two vectors $\text{Re}\,\mathbf{e}$ and $\text{Im}\,\mathbf{e}$ are directed along the axes of the polarization ellipse. The local ellipticity parameter ε can be defined as an invariant:

$$\cos 2\varepsilon = \mathbf{e} \cdot \mathbf{e}. \tag{5}$$

In this case, the local polarization vector $\mathbf{e}$ can be decomposed in the cyclic unit vectors $\mathbf{e}_{\pm 1}$ as follows:

$$\mathbf{e} = \sin(\varepsilon - \pi/4)e^{i\phi}\mathbf{e}_{-1} + \cos(\varepsilon - \pi/4)e^{-i\phi}\mathbf{e}_{+1}, \tag{6}$$

where the angle φ determines the orientation of the ellipse axes in the xy plane.

It is well known (see, for example, [2]) that the force of light pressure exerted on an atom in a monochromatic field can be presented in the form

$$\begin{aligned} F &= -\overline{\mathbf{D}(z, t) \cdot \nabla_z \mathbf{E}(z, t)} \\ &= -\mathbf{D}^{(-)}(z) \cdot \nabla_z \mathbf{E}(z) + \text{c.c.}, \end{aligned} \tag{7}$$

where $\mathbf{D}^{(-)}(z)$ is the amplitude of the negative-frequency component of the average dipole moment of the atom at point z:

$$\mathbf{D}(z, t) = \mathbf{D}^{(+)}(z)e^{-i\omega t} + \mathbf{D}^{(-)}(z)e^{i\omega t};$$

the bar indicates averaging over time; and $\nabla_z = \partial/\partial z$. Using the steady-state solution obtained by us for the density matrix of atoms [17–19], we single out in the average dipole moment two components, one of which (proportional to A) rotates in the same direction as the external field, while the other (proportional to B) rotates in the opposite direction:

$$\mathbf{D}^{(-)} = \frac{1}{\alpha_0 + 2S\alpha_1}\frac{\Omega^*}{\delta - i\gamma/2}(A\mathbf{e}^* + B\mathbf{e}). \tag{8}$$

Here, γ is the rate of spontaneous decay of the excited level, $\delta = \omega - \omega_0$ is the detuning of the field frequency from the transition frequency ω_0, $S = \Omega^2/(\gamma^2/4 + \delta^2)$ is the saturation parameter of the transition, and the Rabi frequency is defined in terms of the reduced dipole moment d and the real field amplitude as follows: $\Omega = dE/\hbar$. The coefficients α_0, α_1, A, and B are functions of the scalar product $\mathbf{e} \cdot \mathbf{e} = \cos 2\varepsilon$ only. Their explicit form is determined by the type of the optical transition. For example, for the $J \longrightarrow J$ transitions (J is a half-integer), it follows from the results obtained in [17] that

$$\begin{aligned} \alpha_0 &= \frac{4J(J+1)}{\mathbf{e} \cdot \mathbf{e}} \sum_{L = 1, 3, \ldots}^{2J} G_L P_L\left(\frac{1}{\mathbf{e} \cdot \mathbf{e}}\right), \\ \alpha_1 &= 2J + 1, \\ A &= 0, \quad B = \frac{2J+1}{\mathbf{e} \cdot \mathbf{e}}, \end{aligned} \tag{9}$$

where $P_L(x)$ is the Legendre polynomial and

$$C_L = (-1)^{L-1}(2L+1)\left(\frac{(L-1)!!}{L!!}\right)^2 \times \frac{(2J+L)!!(2J-L-1)!!}{(2J-L)!!(2J+L+1)!!}.$$

For another class of transitions ($J \longrightarrow J+1$), the results obtained in [18] lead to

$$\alpha_0 = \frac{1}{(2J+1)(4J+1)!\mathbf{e}\cdot\mathbf{e}} \sum_{L=0(1),2(3),\ldots}^{2J} C_L P_L\left(\frac{1}{\mathbf{e}\cdot\mathbf{e}}\right),$$
$$\alpha_1 = P_{2J+1}\left(\frac{1}{\mathbf{e}\cdot\mathbf{e}}\right),$$
$$A = \frac{1}{(2J+1)\mathbf{e}\cdot\mathbf{e}} P'_{2J+1}\left(\frac{1}{\mathbf{e}\cdot\mathbf{e}}\right), \quad (10)$$
$$B = -\frac{1}{(2J+1)\mathbf{e}\cdot\mathbf{e}} P'_{2J}\left(\frac{1}{\mathbf{e}\cdot\mathbf{e}}\right),$$

where

$$C_L = (2L+1)(2J-L)!(2J+L+1)!.$$

In the relation for α_0 in (10), the summation is carried out over even L (starting from zero) for integral values of momentum J and over odd J (starting from unity) for half-integral values of J, and coefficients A and B are expressed in terms of the derivative $P'_L = dP_L(x)/dx$ of the Legendre polynomials. It should be noted that the coordinate dependence of the force and its dependence on the wave parameters are completely determined by two invariants, $\mathbf{E}\cdot\mathbf{E}$ and $\mathbf{E}\cdot\mathbf{E}^*$ since $S \propto \mathbf{E}\cdot\mathbf{E}^*$, and $(\mathbf{e}\cdot\mathbf{e})^2 = |\mathbf{E}\cdot\mathbf{E}|^2/(\mathbf{E}\cdot\mathbf{E}^*)^2$.

3. DIPOLE FORCE IN THE ε–θ–ε FIELD

In this work, we confine our analysis to the ε–θ–ε configuration described above. It can be proved that in the case of counter-propagating waves with the same intensity, the gradients of phase Φ (see Eq. (4)) and the angle of rotation ϕ (see Eq. (6)) of the local polarization ellipse are connected through the following stringent relation:

$$\nabla_z\Phi = \sin(2\varepsilon)\nabla_z\phi, \quad (11)$$

where ε is the local value of the field ellipticity. In this case, the expression for the force of light pressure assumes the form

$$F = -\mathrm{Re}\left\{\frac{\hbar(\delta + i\gamma/2)S}{\alpha_0 + 2S\alpha_1} \times [(\alpha_1 - A)\nabla_z\ln(\mathbf{E}\cdot\mathbf{E}) + A\nabla_z\ln(\mathbf{E}\cdot\mathbf{E}^*)]\right\}. \quad (12)$$

We can also prove that if the ellipticity parameters of counterpropagating waves are equal, the gradients of the phase and rotation angle and, hence, the force of spontaneous light pressure (which is even in δ and is proportional to γ) vanish. The resultant force is determined by the sum of the forces of induced light pressure associated with the gradients of intensity and ellipticity and is an odd function of the detuning. In the general case, the intensity as well as the ellipticity of the resultant field are spatially inhomogeneous:

$$\mathbf{E}\cdot\mathbf{E}^* = 2E_0^2(1 + \cos\theta\cos 2kz - \sin\theta\sin 2\varepsilon_0\sin 2kz),$$
$$\mathbf{e}\cdot\mathbf{e} = \cos(2\varepsilon) = \frac{\cos 2\varepsilon_0(\cos\theta + \cos 2kz)}{1 + \cos\theta\cos 2kz - \sin\theta\sin 2\varepsilon_0\sin 2kz}, \quad (13)$$

where E_0 denotes the real amplitude of the waves. The ellipticity gradient vanishes in the special case $\theta = 0$ (standing elliptically polarized wave), while the intensity is spatially homogeneous for $\theta = \pi/2$, $\varepsilon_0 = 0$ (*lin* ⊥ *lin* configuration).

The coordinate dependence of the dipole force for some transitions is depicted in Fig. 1. It can be seen that the value of the force averaged over a spatial period,

$$\langle F\rangle = \frac{1}{\lambda}\int_0^\lambda F(z)dz,$$

differs from zero; i.e., the effect of rectification of the force of induced light pressure takes place.

4. RECTIFICATION MECHANISM

In order to determine the mechanism of the emergence of the constant component of the dipole force in the ε–θ–ε field, we consider atoms with the momenta $J_g = 1/2 \longrightarrow J_e = 1/2$ in the limit of small saturations, $S \ll 1$. The specific feature of this case, which considerably simplifies the analysis, is that an atom is insensitive to the interference of the circular field components, σ_+ and σ_- (see the diagram of transitions in Fig. 2). For such a field configuration, these components are standing waves with different amplitudes (for $\varepsilon_0 \neq 0$), which are shifted in space relative to each other by the phase angle θ:

$$I_\pm(z) = |E_\pm(z)|^2 = 2E_0^2\cos^2(kz \pm \theta/2)(1 \pm \sin 2\varepsilon_0). \quad (14)$$

Since the optical pumping is carried out through the independent action of the components $E_\pm(z)$, the situation is close to the rectification in the three-level Λ system under bichromatic excitation considered earlier [8, 9], in which the action of standing waves at adjacent transitions is also independent for a small value of the

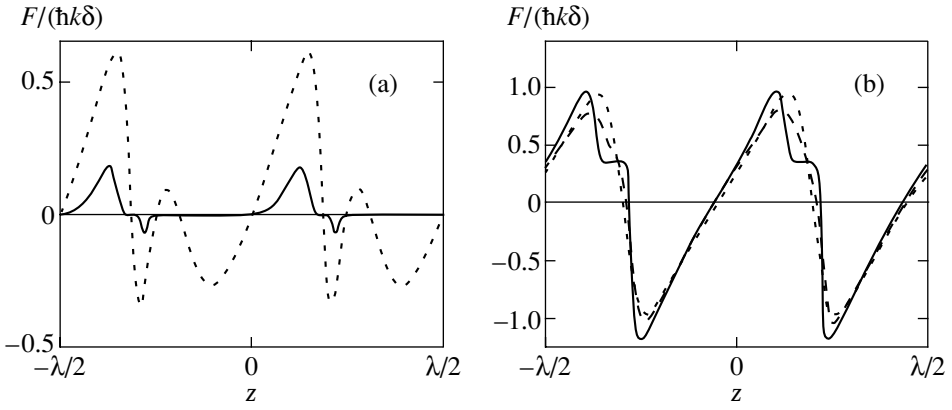

Fig. 1. Coordinate dependence of the force of induced light pressure in the ε–θ–ε field for the following transitions: (a) 1/2 ⟶ 1/2 (dotted curve), 9/2 ⟶ 9/2 (solid curve); (b) 0 ⟶ 1 (dotted curve, rectification is absent), 1/2 ⟶ 3/2 (dashed curve), 4 ⟶ 5 (solid curve). The field parameters are $\varepsilon_0 = \pi/8$, $\theta = \pi/4$, and the saturation parameter per wave $S_0 = 1$.

Rabi frequency (as compared to the two-photon detuning or the coherence relaxation rate in the ground state). In our case of equal detunings, spatial oscillations of forces are shifted in phase through angle θ:

$$F_{\pm}(z) = -\frac{2}{3}\frac{\delta S_0}{E_0^2}\nabla_z I_{\mp}(z), \tag{15}$$

where $S_0 = \Omega_0^2/(\gamma^2/4 + \delta^2)$ is the saturation parameter per wave and Ω_0 is the corresponding Rabi frequency. The coordinate dependence of the populations of sublevels is defined as

$$\pi_{\pm}(z) = \frac{I_{\pm}(z)}{I_+(z) + I_-(z)}. \tag{16}$$

In particular, for equal amplitudes of the circular components (linear polarization of waves, $\varepsilon_0 = 0$), the populations and partial forces satisfy the relations

$$\pi_-(z) = \pi_+(-z), \quad F_-(z) = -F_+(-z),$$

and the resultant force $F = F_+\pi_+ + F_-\pi_-$ vanishes as a result of averaging. In the case of a nonzero ellipticity, the emergence of the constant component of the force is illustrated in Fig. 3, which shows (for $\varepsilon_0 = \pi/8$ and $\theta = \pi/4$) the spatial dependences of populations (Fig. 3a), partial forces (3b), the resultant force and its two components, $F_+\pi_+$ and $F_-\pi_-$ (Fig. 3c). It can be seen that the average value of $F_-\pi_-$ is positive, $F_+\pi_+$ is negative, and the resultant force F is positive on the average due to the difference in their amplitudes.

Thus, the mechanism of the emergence of the constant component of the dipole force in the ε–θ–ε field is associated with a spatially nonuniform optical pumping and with the presence of gradients of intensity and ellipticity. In other words, the rectification effect in a monochromatic field is due to the presence of additional (polarization) degrees of freedom as compared to the model of nondegenerate states.

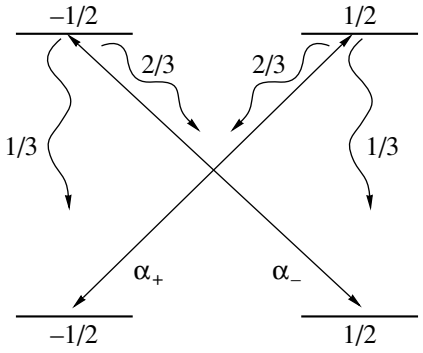

Fig. 2. Diagram of spontaneous (undulating lines) and light-induced (straight lines) transitions $J_g = 1/2 \longrightarrow J_e = 1/2$. The relative amplitudes of light-induced transitions are proportional to the circular field components $\alpha_{\pm} = \mp\cos(\varepsilon \mp \pi/4)$; the numbers indicate the relative probability of spontaneous decay.

5. DEPENDENCE OF THE RECTIFICATION EFFECT ON THE FIELD PARAMETERS

Using expressions (15) and (16) for partial forces and populations, we can easily find the resultant force $F = F_+\pi_+ + F_-\pi_-$ in the first order in the saturation parameter S_0:

$$F(z) = \frac{4\hbar k\delta S_0}{3} \times \frac{\sin 2kz\cos^2 2\varepsilon_0(\cos\theta + \cos 2kz)}{1 + \cos 2kz\cos\theta - \sin 2\varepsilon_0\sin\theta\sin 2kz}. \tag{17}$$

The result of averaging over a spatial period can be written as follows:

$$\langle F\rangle = \frac{2\hbar k\delta S_0}{3} \times \frac{\cos^2 2\varepsilon_0 \sin 2\varepsilon_0 \sin 2\theta (1 - |\sin\theta|\cos 2\varepsilon_0)}{(\cos^2\theta + \sin^2 2\varepsilon_0 \sin^2\theta)^2}. \quad (18)$$

Expressions (17) and (18) demonstrate some features typical of all transitions. In particular, it can be seen that apart from the odd dependence on the detuning of the light field, the force averaged over a spatial period exhibits an odd dependence on angle θ and on the ellipticity ε_0 of light waves, which is valid for all types of optical transitions in the general case of an arbitrary intensity of waves. This can be verified using the symmetry relations for the force. It was noted above that the effect of rectification of the dipole force vanishes for $\varepsilon_0 = 0, \pm\pi/4$ or for $\theta = 0, \pm\pi/2$. Consequently, we can expect that rectification will be maximum for intermediate values of ε_0 and θ. For example, an analysis of formula (18) shows that for the optical transition $1/2 \longrightarrow 1/2$ with small saturation parameters $S_0 \ll 1$, the average force has a smooth peak $\langle F\rangle_{\max} = 0.114\hbar k\delta S_0$ at the point $\varepsilon_0 = 0.35$, $\theta = 0.57$.

For small values of saturation parameter, the dipole force is proportional to S_0. In the opposite limit of large saturations, $S_0 \gg 1$, the average force exerted on a stationary atom decreases in proportion to $1/S_0$, which can be proved using the general expression (12). Thus, the maximum rectification effect takes place for intermediate (not very large or very small) values of the saturation parameter.

In order to find this maximum and the corresponding optimal values of the field parameters, it is convenient to present the average force for Ω_0, $|\delta| \gg \gamma$ in the form

$$\langle F\rangle = \eta\hbar k\Omega_0,$$

where η is the rectification factor depending on the type of optical transition, polarization parameters ε_0 and θ, and the ratio δ/Ω_0. Using numerical integration, we determined the maximum rectification factor η and the optimal values of the parameters ε_0, θ, and δ/Ω_0 for a number of optical transitions. The results are presented in the table for positive detunings and for a positive value of η. All the remaining peaks of $|\eta|$ can be obtained using the antisymmetry of the average force relative to ε_0, θ, and δ.

Another important question is the dependence of the average force $\langle F\rangle$ on velocity. For slow atoms, $kv \ll \gamma$, in the limit of small saturations $S_0 \ll 1$, we can obtain analytic expressions for the average force for transitions with the ground-state momentum $J_g = 1/2$ (i.e., $J_g = 1/2 \longrightarrow J_e = 1/2$ and $J_g = 1/2 \longrightarrow J_e = 3/2$) using a method similar to that used in [20]. For the transition $J_g = 1/2 \longrightarrow J_e = 1/2$, we have

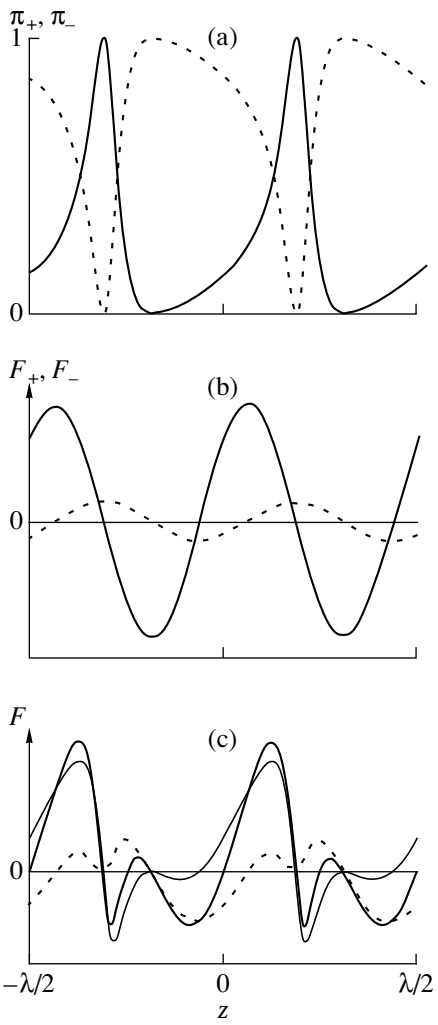

Fig. 3. On the mechanism of dipole force rectification in the ε–θ–ε field. Transition $1/2 \longrightarrow 1/2$ in the limit $S \ll 1$; $\varepsilon_0 = \pi/8$ and $\theta = \pi/4$. Coordinate dependences of (a) populations of magnetic sublevels π_- (solid curve) and π_+ (dotted curve); (b) partial forces F_- (solid curve) and F_+ (dotted curve); (c) products $F_-\pi_-$ (solid curve), $F_+\pi_+$ (dotted curve), and their sum F (bold solid curve).

$$\langle F\rangle(v) = \frac{\hbar k\delta S_0}{3}\int_0^\infty d\tau e^{-\tau}(f_0 + f_1 + f_2). \quad (19)$$

Expressions for f_0, f_1, and f_2 have the form

$$f_0 = -2I_0\left(Q\frac{\sin v\tau}{v}\right)\sin 2v\tau \times (\sin^2\theta + \sin^2 2\varepsilon_0\cos^2\theta),$$

$$f_1 = \frac{4}{Q}I_1\left(Q\frac{\sin v\tau}{v}\right)\sin 2\varepsilon_0 \times \left(\frac{1}{2}\cos^2 2\varepsilon\cos v\tau\sin 2\theta + \sin 2\varepsilon\sin v\tau\right), \quad (20)$$

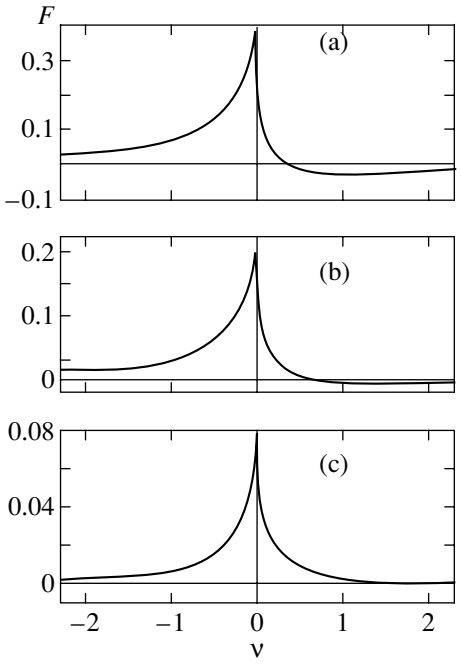

Fig. 4. Dependence of the average force (in the units of $\hbar k\delta S_0/3$) on the dimensionless velocity $\nu = 9kv/4\gamma S_0$ in low-intensity fields, $S_0 \ll 1$, for various parameters of light waves: (a) $\varepsilon_0 = \pi/10$, $\theta = 0.3$; (b) $\varepsilon_0 = \pi/6$, $\theta = 0.3$; (c) $\varepsilon_0 = \pi/6$, $\theta = 0.1$.

$$f_2 = -\frac{2}{Q^2} I_2\left(Q\frac{\sin\nu\tau}{\nu}\right)\sin 2\varepsilon_0 \sin 2\theta \cos^2 2\varepsilon_0,$$

$$Q = \sqrt{\cos^2\theta + \sin^2 2\varepsilon_0 \sin^2\theta},$$

where $I_m(z)$ are modified Bessel's functions and $\nu = 9kv/4\gamma S_0$ is the dimensionless velocity. The dependence of force on velocity for various values of the parameters of light waves is shown in Fig. 4 for $S_0 \ll 1$. It can be seen that for small saturation parameters, the rectification of the dipole force in a monochromatic field of the ε–θ–ε configuration takes place in the velocity range $\Delta v \sim \gamma S_0/k$.

Optimal field parameters and maximum rectification factor for the dipole force

$J_g \longrightarrow J_e$	ε_0	θ	δ/Ω_0	η
1/2 → 1/2	0.46	0.51	0.87	0.057
3/2 → 3/2	0.35	0.39	0.49	0.059
5/2 → 5/2	0.28	0.30	0.39	0.053
7/2 → 7/2	0.23	0.24	0.33	0.048
9/2 → 9/2	0.19	0.19	0.29	0.043
1/2 → 3/2	−0.35	0.43	2.5	0.023
1 → 2	−0.30	0.37	2.4	0.025
3/2 → 5/2	−0.25	0.32	2.4	0.022
2 → 3	−0.22	0.28	2.3	0.019
5/2 → 7.2	−0.20	0.26	2.3	0.017
3 → 4	−0.18	0.24	2.3	0.014
7/2 → 9/2	−0.17	0.22	2.2	0.013
4 → 5	−0.16	0.20	2.2	0.012

The general case of arbitrary velocities and field intensities requires the application of numerical methods. In a one-dimensional periodic field, the approach to the approximate calculation of the atomic density matrix involves the expansion into a Fourier series in the spatial coordinate and the formulation of a system of equations for the corresponding coefficients. Further, we assume that only first n harmonics are excited effectively, while the amplitudes of the remaining harmonics are negligibly small. Discarding harmonics higher than the nth harmonic, we obtain a finite system of algebraic equations, which can be solved numerically for the given values of the parameters. Naturally, the number n is chosen depending on the required accuracy of calculations, this choice being determined by the field parameters and the velocity of the atom. An effective version of such an approach is the method of chain fractions developed by Minogin and Serimaa [21] for a two-level atom in the field of a standing wave and was subsequently generalized to atoms with degenerate levels in fields with polarization gradients (see, for example, [22]). We used this method for calculating the period-averaged force confining the analysis to $n = 20$ (doubling the value of n led only to insignificant changes). The result obtained for the optimal values $\varepsilon_0 = 0.46$, $\theta = -0.51$, and $\delta = -0.87\Omega_0 = -8.7\gamma$ is presented in Fig. 5a. Apart from the rectification effect, the figure shows considerable asymmetry in the dependence of force on velocity in the range $|kv| < 0.1\gamma$. The narrow (with a width much smaller than γ) structure slightly displaced relative to zero is also worth noting.

In order to correctly analyze the change in the velocity distribution of atoms under the action of the rectified force, we must also take into account the fluctuations of the force leading to the diffusion of atoms over momenta. The mechanism considered above suggests that rectification in our case must be accompanied by considerable fluctuations of the force since the effect is associated with transitions between the states in which different forces $F_\pm$ are acting on the atom. The corresponding numerical results (for $n = 20$) for the dependence of the space-averaged diffusion coefficient $\langle D\rangle$ on velocity are presented in Fig. 5b. The main qualitative features are the same as for the force: asymmetry and a sharp peak in the vicinity of zero, which is shifted, however, in the opposite direction.

Simple estimates for the dynamics of variation of the distribution function for an atomic ensemble can be obtained assuming that the localization of atoms at the minima of the optical potential is insignificant and going over to the quantities averaged over a spatial period in the Fokker–Planck equation:

$$\frac{\partial \mathcal{W}(p)}{\partial t} = \left[-\frac{\partial}{\partial p}\langle F\rangle(p) + \frac{\partial^2}{\partial p^2}\langle D\rangle(p)\right]\mathcal{W}(p), \quad (21)$$

where $\mathcal{W}(p)$ is the momentum distribution function for atoms; the momentum and velocity are connected through the conventional relation $p = Mv$. Figure 5c shows the result of numerical integration of Eq. (21) in the case of the Gaussian initial distribution, $\mathcal{W}_0(p) \propto \exp[-(p/p_0)^2]$, with a large (as compared to the photon momentum $\hbar k$) width. The functions $\langle F\rangle(p)$ and $\langle D\rangle(p)$ were chosen in accordance with the results presented in Fig. 5a and 5b, and the following relation was assumed between the recoil frequency and the natural width: $\omega_r = \hbar k^2/2M = 2.5 \times 10^{-3}\gamma$ (which corresponds to ^{23}Na). It can be seen from Fig. 5c that the rectification effect is manifested in the asymmetry of the momentum distribution of the atoms (the amplitude of the right peak is approximately 10% larger than that of the left peak).

The effect of rectification of the dipole force in a field of the ε–θ–ε configuration can be observed experimentally. By way of an example, we consider the interaction of a thermal beam of ^{23}Na atoms with a small angular divergence ($v_\parallel \sim 10^5$ cm/s, $v_\perp \sim 150$ cm/s) with transverse laser beams. Let the radiation frequency be tuned in resonance with the D_1 line (transition $3S_{1/2} \longrightarrow 3P_{1/2}$, $\lambda = 590$ nm). For the beam intensities 5 W/cm^2 (power 40 mW focused to a spot of diameter 1 mm), the Rabi frequency amounts approximately to 10γ, where $\gamma = 2\pi \times 10$ MHz is the spontaneous decay rate for the $3P_{1/2}$ level. For obtaining qualitative estimates, we disregard the hyperfine structure. The remaining field parameters (δ, ε_0, and θ) will be chosen so that they correspond to the maximum rectification (see the table and Fig. 5a); i.e., $\varepsilon_0 = 0.46$, $\theta = -0.51$, $\delta = -0.87\Omega_0$. The initial spread in transverse momenta (measured in the units of the photon momentum), $p_0 \approx 50\hbar k$, and the time of interaction of atoms with the field, $t \approx 60\gamma^{-1}$, correspond to the final momentum distribution depicted in Fig. 5c. Modern experimental technique makes it possible to detect such an asymmetry easily [23, 24].

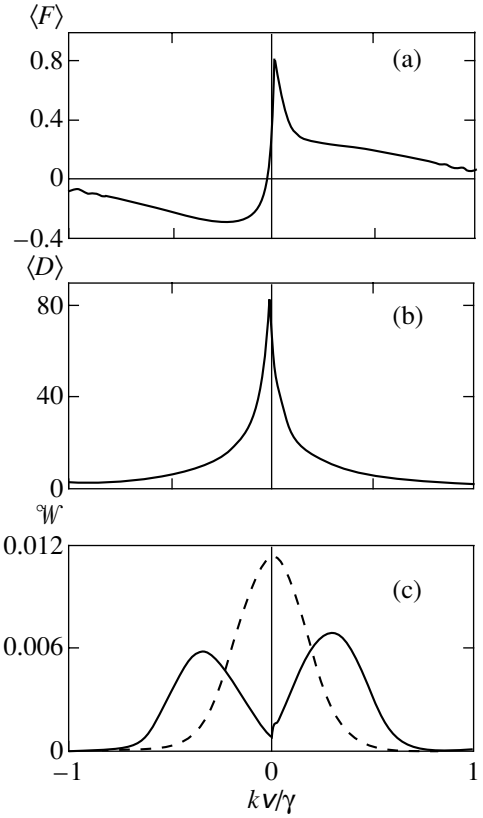

Fig. 5. Results of numerical calculations: (a) dependence of average force (in the units of $\hbar k\gamma$ on velocity; (b) average diffusion coefficient (in the units of $(\hbar k)^2\gamma$) as a function of velocity; (c) initial (dashed curve) and final (solid curve) velocity distributions of atoms. In all the figures, the field parameters are $\varepsilon_0 = 0.46$, $\theta = -0.51$ and $\delta = -0.87\Omega_0 = -8.7\gamma$; the time of interaction in (c) is $t = 60\gamma^{-1}$ and the initial momentum spread is $p_0 = 50\hbar k \approx 0.25M\gamma/k$.

6. CONCLUSIONS

Thus, we have demonstrated the possibility of rectification of the dipole force in a monochromatic field in the absence of a magnetic field. In the one-dimensional case, this requires the application of elliptically polarized waves whose polarization ellipses are oriented at an angle to one another. In the case of the linear or circular polarization of waves (in arbitrary combinations) or in the case of coaxial polarization ellipses, the rectification of the dipole force does not take place, which can be proved rigorously on the basis of symmetry relations. It should be noted that for different ellipticities of counterpropagating waves, both induced (dipole) and spontaneous forces of light pressure are rectified in the general case; i.e., the force contains both odd and even contributions in detuning. The uniqueness of the ε–θ–ε configuration considered here is that the spontaneous component of the force (which is even in detuning) is equal to zero. Moreover, it follows from relation (1) that the average force in the ε–θ–ε configuration is odd in detuning irrespective of the velocity of the atom. This circumstance can be used in some spectroscopic applications.

ACKNOWLEDGMENTS

This work was partly financed by the Russian Foundation for Basic Research (project nos. 01-02-17036 and 01-02-17744) and by the Ministry of Education of the Russian Federation in the field of fundamental education (grant no. E00-3.2-153). The work of one of the authors (O. N. P) was supported by an INTAS grant (YSF grant no. 00-129) for young scientists.

REFERENCES

1. V. G. Minogin and V. S. Letokhov, *The Pressure of Laser Radiation on Atoms* (Nauka, Moscow, 1986).
2. A. P. Kazantsev, G. I. Surdutovich, and V. P. Yakovlev, *The Mechanical Action of Light on Atoms* (Nauka, Moscow, 1991).
3. A. P. Kazantsev and I. V. Krasnov, Pis'ma Zh. Éksp. Teor. Fiz. **46**, 264 (1987) [JETP Lett. **46**, 332 (1987)].
4. A. P. Kazantsev and I. V. Krasnov, Zh. Éksp. Teor. Fiz. **95**, 104 (1989) [Sov. Phys. JETP **68**, 59 (1989)].
5. A. P. Kazantsev and I. V. Krasnov, J. Opt. Soc. Am. B **6**, 2140 (1989).
6. R. Grimm, Y. B. Ovchinnikov, A. I. Sidorov, and V. S. Letokhov, Phys. Rev. Lett. **65**, 1415 (1990).
7. Yu. B. Ovchinnikov, R. Grimm, A. I. Sidorov, and V. S. Letokhov, Opt. Spektrosk. **76**, 188 (1993) [Opt. Spectrosc. **76**, 192 (1993)].
8. J. Javanainen, Phys. Rev. Lett. **64**, 519 (1990).
9. A. I. Sidorov, R. Grimm, and V. S. Letokhov, J. Phys. B **24**, 3733 (1991).
10. P. R. Hemmer, M. S. Shahriar, M. Prentiss, *et al.*, Phys. Rev. Lett. **68**, 3148 (1992).
11. D. V. Kosachiov, Yu. V. Rozhdestvensky, and G. Nienhuis, J. Opt. Soc. Am. B **14**, 535 (1997).
12. T. Cai and N. P. Bigelow, Opt. Commun. **104**, 175 (1993).
13. T. T. Grove and P. L. Gould, Laser Phys. **4**, 957 (1994).
14. H. Pu, T. Cai, N. P. Bigelow, *et al.*, Opt. Commun. **118**, 261 (1995).
15. J. Söding and R. Grimm, Phys. Rev. A **50**, 2517 (1994).
16. A. Haak, W. Ertmer, and H. Wallis, Laser Phys. **4**, 1030 (1994).
17. A. V. Taĭchenachev, A. M. Tumaĭkin, V. I. Yudin, and G. Nienhuis, Zh. Éksp. Teor. Fiz. **108**, 415 (1995) [JETP **81**, 224 (1995)].
18. A. V. Taĭchenachev, A. M. Tumaĭkin, V. I. Yudin, and G. Nienhuis, Zh. Éksp. Teor. Fiz. **114**, 125 (1998) [JETP **87**, 70 (1998)].
19. G. Nienhuis, A. V. Taĭchenachev, A. M. Tumaĭkin, and V. I. Yudin, Europhys. Lett. **44**, 20 (1998).
20. V. Finkelstein, P. R. Berman, and J. Guo, Phys. Rev. A **45**, 1829 (1992).
21. V. G. Minogin and O. T. Serimaa, Opt. Commun. **30**, 373 (1979).
22. O. Emile, R. Kaizer, C. Gerz, *et al.*, J. Phys. II **3**, 1709 (1993).
23. P. L. Gould, G. A. Ruff, and D. E. Pritchard, Phys. Rev. Lett. **56**, 827 (1986).
24. V. A. Grinchuk, E. F. Kuzin, M. L. Nagaeva, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **57**, 524 (1993) [JETP Lett. **57**, 548 (1993)].

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 93, No. 1, 2001, pp. 71–79.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 1, 2001, pp. 85–93.

PLASMA, GASES

On the Criterion for Effectiveness of Wave Linear Transformation in a Smoothly Inhomogeneous Medium and of Nonadiabatic Transitions during Atomic Collisions

K. Yu. Bliokh[a, *] and S. V. Grinyok[b]

[a] *Institute of Radioastronomy, National Academy of Science of Ukraine, Kharkov, 61002 Ukraine*

[b] *Kharkov National University, Kharkov, 61077 Ukraine*

** e-mail: kostya@bliokh.kharkiv.com*

Received February 14, 2001

Abstract—A universal criterion for effectiveness of linear transformation of waves with locally close characteristic exponents in smoothly inhomogeneous media is obtained. The same criterion is applicable for estimating the effectiveness of nonadiabatic transitions in slow atomic collisions. The formalism developed for an analysis of the linear interaction of waves is based of the WKB asymptotic form of the solution of a scalar *n*th order ordinary differential equation. The obtained criterion can be applied in any practical problem for drawing a conclusion about the effectiveness of the linear interaction of modes if only the characteristic equation of waves in a homogeneous medium and the coefficients of the initial differential equation are known. In this case, the solution of the problem is reduced to elementary arithmetic calculations. *© 2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The problem on the linear interaction of waves in smoothly inhomogeneous media has been studied extensively and has many applications (see the review by Zheleznyakov *et al.* [1] and the references cited therein), for example, in the physics of plasma and astrophysics [2–4]. In this paper, we consider the most frequently encountered case when the interaction of waves takes place in the region of local closeness of their dispersion characteristics. The transformation in this case means the conversion of a wave corresponding to a branch of the dispersion curve into a wave corresponding to another branch. A similar situation also takes place for slow atomic collisions [5, 6]. In this case, linear transformations correspond to so-called nonadiabatic transitions between quasi-intersecting terms.

Strictly speaking, the concept of the wave corresponding to a given branch of the dispersion curve is applicable only to homogeneous media. The geometrical-optics approximation or the WKB approximation extend this concept and makes it possible to apply it for waves in smoothly inhomogeneous media. In this approximation, waves in a smoothly inhomogeneous medium are assumed to be locally identical to waves in a homogeneous medium with current characteristics. The waves corresponding to different branches of dispersion curves are regarded as independent. Thus, the linear transformation of waves is essentially a violation of the WKB approximation.

Nevertheless, in order to calculate the effectiveness of a linear transformation, it is sufficient to analyze the equations and their solutions corresponding to the WKB approximation. As a rule, the calculations are made using one of the following two methods. The first method is that of complex phase integrals, which makes it possible to travel around singular points of the WKB approximation along a contour in the complex plane during integration [2–5]. The second method is based on an analysis of the polarization of various types of waves [1]. The latter method possesses a visual general formalism and deals with equations on the real axis.

The first method is more illustrative from the mathematical point of view, while the second reveals the qualitative physical pattern of the phenomenon more clearly. However, both methods are quite cumbersome in calculations when used in various specific applications. In the first case, this is associated with the need to evaluate the integrals of rather complex functions with singularities in the complex plane, while in the second case, the complexity of calculations is due to a cumbersome matrix representation of linear algebra.

It should be noted that in many practical problems, the exact calculation of the coefficients of linear transformation and nonadiabatic transitions is not required, and it is sufficient to obtain only an order-of-magnitude estimate of their effectiveness. In this situation, there are no universal criteria for the effectiveness of the interaction. Obviously, the wave transformation must be the stronger, the closer the dispersion characteristics of interacting modes to each other (resonance conditions). Besides, the effectiveness of the interaction

1063-7761/01/9301-0071$21.00

between the waves must increase with the inhomogeneity gradient since it is precisely inhomogeneity that ensures the coupling between different modes. These two remarks lead to the following intuitive criterion. It is often assumed that linear transformation is significant when the characteristic scale of variation of the system parameters is smaller than or of the order of the period of beats between interacting modes [7]. If the eigenvalues (frequencies or wave vectors) corresponding to these modes are of the same order p and the separation between them is of the order of δp ($|\delta p| \ll |p|$), the effectiveness criterion can be written in the form

$$\left|\frac{p\delta p}{p'}\right| \lesssim 1, \qquad (1)$$

where the prime denotes the derivative with respect to an independent variable. A similar criterion is also introduced in the theory of atomic collisions; in this case, the parameter corresponding to the left-hand side of relation (1) is called the Massey parameter. However, in real situations, criterion (1) is often incorrect. For this reason, the problem of estimating the effectiveness of linear transformations must be solved, as a rule, comprehensively in each specific case.

In this work, we propose a universal criterion for estimating the effectiveness of a linear transformation in each practical problem without resorting to the solution of the problem on linear interaction. We propose an alternative method for analyzing a linear transformation in the region of close values of the refractive indices for different types of waves, which is applicable in problems with a monotonic smooth inhomogeneity and in the absence of reflection of the waves. Our method is based on the WKB formalism of a scalar nth order linear equation on the real axis and makes it possible to obtain result with smaller computational time expenditure.[1] For example, in order to estimate the effectiveness of linear transformation in some cases, it is sufficient to know only the dispersion relation for waves in a homogeneous medium. This is due to the fact that in scalar WKB solutions, the entire information about a wave (including its polarization) is contained in the complex phase; this makes it possible to avoid complex computations of eigenvectors–solutions for waves in the matrix method [1]. Apart from the analysis of the general problem, we also consider several examples of wave interaction and nonadiabatic transitions, which were solved earlier by other methods.

[1] It should be noted that a matrix equation of the general type contains $2n^2$ different coefficients, while a scalar equation has only $2n$ coefficients. The doubling is due to the fact that in the WKB approximation, the terms of different orders of smallness must be separated in each coefficient.

2. GENERAL FORMALISM

2.1. Initial Equations

Let μ be a parameter (or a set of parameters) determining the inhomogeneity of the problem, which is a smooth function of coordinate z:

$$\mu = \mu(\varepsilon z).$$

Here, $\varepsilon \ll 1$ is a small parameter characterizing the smoothness of inhomogeneity. Taking into account the homogeneity of the system under investigation in time t and coordinates $\mathbf{r}_\perp$ transverse to z, we can eliminate the dependence of variables on these quantities from the initial equations with the help of factors $\exp(i\mathbf{k}_\perp\mathbf{r}_\perp - i\omega t)$. As a result, the system of differential equations describing linear waves in an inhomogeneous medium is reduced to a scalar nth order ordinary linear differential equation. In the geometrical optics approximation, the solutions asymptotic in ε are considered with an accuracy of ε^1. In this case, the differential equation under investigation can be linearized in $\mu' \sim \varepsilon$ (the prime indicates the derivative with respect to z); discarding higher-order terms, we can write this equation in the general form:

$$\sum_{j=0}^{n} [a_j(\varepsilon z) + \mu'(\varepsilon z) b_j(\varepsilon z)] y^{(n-j)}(z) = 0, \qquad (2)$$

where $y(z)$ is the sought function and a_j and b_j are the coefficients of the equation, which are inhomogeneous in view of their dependence on μ. Without any loss of generality, we can put $a_0 \equiv 1$ and $b_0 \equiv 0$. The first terms in the sum in Eq. (2) are of the order of ε^0 and the second, of the order of ε^1.

Equation (2) corresponds to the characteristic (dispersion) equation

$$l(p, \varepsilon z) \equiv \sum_{j=0}^{n} [a_j(\varepsilon z) + \mu'(\varepsilon z) b_j(\varepsilon z)] p^{n-j} = 0. \quad (3)$$

Its roots (characteristic values) are smooth functions of z: $p = p(\varepsilon z)$. The asymptotic WKB solution of Eq. (2) for the kth mode can be written in the form [8]

$$y_k \approx \exp\left[\int_{z_0}^{z} \Phi_k(\varepsilon\xi) d\xi\right],$$
$$\Phi_k(\varepsilon\xi) = p_k(\varepsilon\xi) - \frac{1}{2} p_k'(\varepsilon\xi) \frac{l_{pp}(p, \varepsilon\xi)}{l_p(p, \varepsilon\xi)}\bigg|_{p = p_k(\varepsilon z)}. \qquad (4)$$

Here and below, the subscripts p indicate derivatives with respect to the corresponding argument of the characteristic equation.

In Eq. (3), its roots, and expression (4), we single out the terms corresponding to the zeroth and first orders of smallness in ε:

$$\begin{aligned} l &= l^{(0)} + l^{(1)}, \\ p &= p^{(0)} + p^{(1)}, \end{aligned} \tag{5}$$

where $l^{(i)} \sim p^{(i)} \sim \varepsilon^i$

$$l^{(0)}(p, \varepsilon z) \equiv \sum_{j=0}^{n} a_j(\varepsilon z) p^{n-j} = \prod_{j=1}^{n} (p - p_j^{(0)}(\varepsilon z)), \tag{6}$$

$$l^{(1)}(p, \varepsilon z) \equiv \mu'(\varepsilon z) \sum_{j=1}^{n} b_j(\varepsilon z) p^{n-j}. \tag{7}$$

Using the perturbation method, we can easily obtain the following relation from Eqs. (3), (6), and (7):

$$p_k^{(1)} = -\frac{l^{(1)}}{l_p^{(0)}}\bigg|_{p = p_k^{(0)}} = -\frac{\mu' \sum_{j=1}^{n} b_j (p_k^{(0)})^{n-j}}{\prod_{j \neq k} (p_k^{(0)} - p_j^{(0)})}. \tag{8}$$

Here and below, the arguments of the known functions are omitted for the sake of brevity. Expression (4) for the WKB phase assumes the form

$$\begin{aligned} \Phi_k &= \Phi_k^{(0)} + \Phi_{k1}^{(1)} + \Phi_{k2}^{(1)} \\ &= p_k^{(0)} + p_k^{(1)} - \frac{1}{2} p_k^{(0)'} \frac{l_{pp}^{(0)}}{l_p^{(0)}}\bigg|_{p = p_k^{(0)}}, \end{aligned} \tag{9}$$

where the three quantities $\Phi_k^{(i)} \sim \varepsilon^i$ are equal, respectively, to the three terms on the right-hand side of relation (9). Substituting relations (6) and (8) into (9), we obtain the final expression for the WKB phase:

$$\begin{aligned} \Phi_k &= \Phi_k^{(0)} + \Phi_{k1}^{(1)} + \Phi_{k2}^{(1)} \\ &= p_k^{(0)} - \frac{\mu' \sum_{j=1}^{n} b_j (p_k^{(0)})^{n-j}}{\prod_{j \neq k} (p_k^{(0)} - p_j^{(0)})} - \sum_{j \neq k} \frac{p_k^{(0)'}}{(p_k^{(0)} - p_j^{(0)})}. \end{aligned} \tag{10}$$

The quantity $\Phi_k^{(0)} = p_k^{(0)}$ forms the ordinary dynamic phase in the WKB solution (4). The quantity

$$\Phi_{k1}^{(1)} = -\frac{\mu' \sum_{j=1}^{n} b_j (p_k^{(0)})^{n-j}}{\prod_{j \neq k} (p_k^{(0)} - p_j^{(0)})}$$

forms the ε-order correction in the WKB phase, which is due to the presence of derivatives of the inhomogeneous parameter μ in the initial equation (2). The quantity

$$\Phi_{k2}^{(1)} = -\sum_{j \neq k} \frac{p_k^{(0)'}}{p_k^{(0)} - p_j^{(0)}}$$

forms the ε-order correction in the WKB phase, which is associated with inhomogeneity of running characteristic numbers $p_k^{(0)}$.

2.2. Criterion of the Effectiveness of Interaction between Waves

The construction of asymptotic WKB solutions presumes that all the terms proportional to ε^1 are much smaller than the terms of the order of ε^0; i.e.,

$$\left|\Phi_{k1}^{(1)} + \Phi_{k2}^{(1)}\right| \ll \left|\Phi_k^{(0)}\right|. \tag{11}$$

The violation of the WKB approximation (violation of condition (11)) can be interpreted as the violation of independence of different solutions, i.e., interaction and transformation of waves.

Let us first consider the case when the term $\Phi_{k1}^{(1)}$ can be neglected: $\left|\Phi_{k1}^{(1)}\right| \ll \left|\Phi_{k2}^{(1)}\right|$. In this case, the waves are no longer independent for $\left|\Phi_k^{(0)}\right| \gtrsim \left|\Phi_{k2}^{(1)}\right|$ or

$$\left|\sum_{j \neq k} \frac{p_k^{(0)'}}{p_k^{(0)} - p_j^{(0)}}\right| \gtrsim \left|p_k^{(0)}\right|. \tag{12}$$

If the distance between the roots of the characteristic equation is not small ($\left|p_k^{(0)} - p_j^{(0)}\right| \sim \left|p_k^{(0)}\right|$), relation (12) leads us to the well-known estimate

$$\left|\frac{p_k^{(0)'}}{p_k^{(0)}}\right| \gtrsim \left|p_k^{(0)}\right|. \tag{13}$$

If, however, two roots of the characteristic equation approach each other to a small distance,

$$\left|p_k^{(0)} - p_m^{(0)}\right| = 2\delta p \ll \left|p_k^{(0)}\right|,$$

the WKB approximation is violated in this region under a less stringent condition imposed on $p_k^{(0)'}$:

$$\left|\frac{p_k^{(0)'}}{\delta p}\right| \gtrsim \left|p_k^{(0)}\right|. \tag{14}$$

This condition is exactly the criterion (1) given in the Introduction. It may appear that inequality (14) is just the condition for the effective transformation of waves in the region of convergence of their characteristic exponents. However, the following important circumstance is disregarded in this case. As a matter of fact, the interaction between two modes with close

characteristic values is of the resonance nature, and the effective energy exchange between them takes place even in the case of a weak coupling. In order to take this circumstance into account, we proceed in the following way.

In Eq. (2), we carry out the substitution

$$y = \tilde{y}\exp(p_{\mathrm{mid}}z), \quad p_{\mathrm{mid}} = \left.\frac{p_k^{(0)} + p_m^{(0)}}{2}\right|_{z = z_{\min}}, \quad (15)$$

where $z_{\min}$ is the point at which the difference between the converging roots $p_k^{(0)}$ and $p_m^{(0)}$ is minimal. After such a substitution, the current roots of the new characteristic equation are

$$\tilde{p}_j^{(0)}(\varepsilon z) = p_j^{(0)}(\varepsilon z) - p_{\mathrm{mid}}. \quad (16)$$

In the new variables, the interaction between the modes under investigation is described in a nonresonance way since the two converging roots for $z = z_{\min}$ are arranged symmetrically and are of the same order of magnitude as their separation:

$$\tilde{p}_{k,m}^{(0)} = \pm\delta p. \quad (17)$$

Considering the zeroth approximation in the small parameter $\delta p/p_{\mathrm{mid}}$, we can easily see that the converging roots satisfy the characteristic equation (3) written in new coordinates:

$$\tilde{l}(p, \varepsilon z) \equiv \prod_{j=1}^{n}(p - \tilde{p}_j^{(0)}(\varepsilon z)) + \mu'(\varepsilon z)\tilde{b}_n(\varepsilon z), \quad (18)$$

where

$$\tilde{b}_n(\varepsilon z) \approx \sum_{j=1}^{n} b_j(\varepsilon z)p_{\mathrm{mid}}^{n-j}(\varepsilon z). \quad (19)$$

Comparing Eq. (18) with Eqs. (5)–(7) and (10) and taking into account relation (19), we obtain the expression for the WKB phase in the new variables:

$$\tilde{\Phi}_k = \tilde{\Phi}_k^{(0)} + \tilde{\Phi}_{k1}^{(1)} + \tilde{\Phi}_{k2}^{(1)}$$
$$\approx \tilde{p}_k^{(0)} - \frac{\mu'\sum_{j=1}^{n} b_j p_{\mathrm{mid}}^{n-j}}{\prod_{j\neq k}(\tilde{p}_k^{(0)} - \tilde{p}_j^{(0)})} - \sum_{j\neq k}\frac{\tilde{p}_k^{(0)'}}{\tilde{p}_k^{(0)} - \tilde{p}_j^{(0)}}. \quad (20)$$

Now, the condition for the effectiveness of a linear interaction between waves has the form

$$\left|\tilde{\Phi}_{k1}^{(1)} + \tilde{\Phi}_{k2}^{(1)}\right| \gtrsim \left|\tilde{\Phi}_k^{(0)}\right|. \quad (21)$$

Expression (20) combined with (17) leads to

$$\left|\tilde{\Phi}_k^{(0)}\right| = \delta p,$$

$$\left|\tilde{\Phi}_{k1}^{(1)}\right| \approx \left|\frac{\mu'\sum_{j=1}^{n} b_j p_{\mathrm{mid}}^{n-j}}{2\delta p\prod_{j\neq k,m}(\tilde{p}_k^{(0)} - \tilde{p}_j^{(0)})}\right|, \quad (22)$$

$$\left|\tilde{\Phi}_{k2}^{(1)}\right| \approx \left|\frac{p_k^{(0)'}}{2\delta p}\right|.$$

In the case when $\left|\tilde{\Phi}_{k1}^{(1)}\right| \ll \left|\tilde{\Phi}_{k2}^{(1)}\right|$, the following estimate holds:

$$\left|\frac{p_k^{(0)'}}{\delta p}\right| \gtrsim |\delta p|. \quad (23)$$

This condition indicates that owing to the resonance nature of interaction, the transformation of waves becomes effective for smoother inhomogeneities than in relations (1) and (14). For $\left|\tilde{\Phi}_{k2}^{(1)}\right| \ll \left|\tilde{\Phi}_{k1}^{(1)}\right|$, the effectiveness criterion for a transformation is

$$\left|\frac{\mu'\sum_{j=1}^{n} b_j p_{\mathrm{mid}}^{n-j}}{\delta p\prod_{j\neq k,m}(\tilde{p}_k^{(0)} - \tilde{p}_j^{(0)})}\right| \gtrsim |\delta p|. \quad (24)$$

In the general case, condition (21) can be written in the form

$$\left|\frac{1}{\delta p}\left(p_k^{(0)'} + \frac{\mu'\sum_{j=1}^{n} b_j p_{\mathrm{mid}}^{n-j}}{\delta p\prod_{j\neq k,m}(\tilde{p}_k^{(0)} - \tilde{p}_j^{(0)})}\right)\right| \gtrsim |\delta p|. \quad (25)$$

Thus, we obtained criteria (23)–(25) of the effectiveness for the linear interaction of waves in the general case. The estimation of the effectiveness of the linear transformation of waves in each specific case is now reduced to elementary computations using only the coefficients of the initial differential equation and the roots of the dispersion equation.

3. EXAMPLES

3.1. The Case $\left|\tilde{\Phi}_{k1}^{(1)}\right| \ll \left|\tilde{\Phi}_{k2}^{(1)}\right|$. Linear Transformation of Waves in Magnetospheres of Pulsars

Let us consider the problem [9] on the interaction of waves in the inhomogeneous magnetized ultrarelativistic electron–positron plasma of the magnetosphere of a

pulsar. The basic equations in this system can be written in the form

$$\begin{aligned} y_1' &= iy_2, \\ y_2' &= i(1-\vartheta^2)y_1 + i\vartheta y_3, \\ y_3' &= iy_4, \\ y_4' &= -i\vartheta\frac{\mu}{4\gamma^4}y_1 - \left[i\left(\beta^2 - \frac{\mu}{4\gamma^4}\right) + \beta\frac{\mu'}{\mu}\right]y_3 \\ &\quad + \left(2i\beta + \frac{\mu'}{\mu}\right)y_4, \end{aligned} \tag{26}$$

where y_j are the field variables, the coordinate is reduced to the dimensionless form with the help of the characteristic wavelength $l = c/\omega$, $\vartheta \ll 1$ is the angle between the wave vector and the external magnetic field, $\gamma \gg 1$ is the Lorentz factor, β^{-1} is the ratio of the plasma flow velocity to the velocity of light, and $\mu = \mu(\varepsilon z)$ is the parameter characterizing the inhomogeneity of the plasma and proportional to its density.

System (26) can be reduced to the following fourth-order scalar differential equation:

$$\begin{aligned} &y_1^{(4)} - \left(2i\beta + \frac{\mu'}{\mu}\right)y_1^{(3)} \\ &+ \left(1 - \vartheta^2 + \frac{\mu}{4\gamma^4} - \beta^2 + i\beta\frac{\mu'}{\mu}\right)y_1^{(2)} \\ &- \left(2i\beta + \frac{\mu'}{\mu}\right)(1-\vartheta^2)y_1^{(1)} \\ &+ \left\{(1-\vartheta^2)\left[\frac{\mu}{4\gamma^4} - \beta^2 + i\beta\frac{\mu'}{\mu}\right] + \frac{\vartheta^2\mu}{4\gamma^4}\right\}y_1 = 0. \end{aligned} \tag{27}$$

Comparing Eq. (27) with Eq. (2), we see that in the former equation,

$$\begin{aligned} b_1 &= -\frac{1}{\mu}, \quad b_2 = i\frac{\beta}{\mu}, \\ b_3 &= -\frac{1-\vartheta^2}{\mu}, \quad b_4 = i\frac{\beta(1-\vartheta^2)}{\mu}. \end{aligned} \tag{28}$$

In the zeroth approximation in ε, Eq. (27) corresponds to the dispersion equation

$$(\lambda - \beta)^2(\lambda^2 + \vartheta^2 - 1) = \frac{\mu}{4\gamma^4}(\lambda^2 - 1), \tag{29}$$

where $i\lambda \equiv p^{(0)}$.

The dependence $\lambda(\mu)$ is shown in Fig. 1. It can be clearly seen that in the vicinity of $\mu = 1$, the two roots whose values are close to unity converge (encircled region). When $\vartheta = 0$, the two roots coincide at point $\mu = 1$ and the reconnection of the dispersion branches takes place: in this region, they have the form of two intersecting straight lines. During passage through the region of $\mu \approx 1$, the representative point of the wave with a finite ϑ moves along the curve in Fig. 1 and goes over from one asymptote to the other, while the representation point of the wave with $\vartheta = 0$ moves along the straight line and remains on the same asymptote. For small values of $\vartheta \ll 1$, the representation point of the wave goes over from one curve to the other, remaining on the same asymptote as in the case of $\vartheta = 0$, which ensures the continuity of the limiting transition $\vartheta \longrightarrow 0$. This is just the linear transformation of the wave. In the case of effective transformation (when stringent inequalities (23)–(25) hold), the wave corresponding to one curve is converted almost completely into the wave corresponding to another curve, but on the same asymptote as that for the initial wave. In this case and in the examples considered below, as in most real cases, the waves corresponding to the same asymptote have close polarizations, while those corresponding to different asymptotes have essentially different polarizations. This is exactly the physical reason behind the linear transformation of waves.

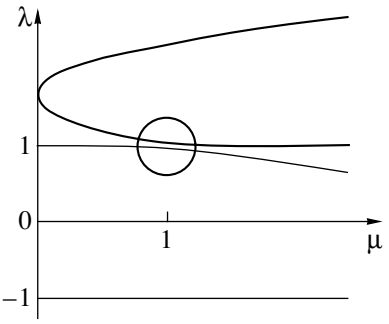

Fig. 1. Dependence of the refractive indices on the density of plasma for quasi-longitudinal waves in an electron–positron magnetized plasma. Here and in the following figures, the region of possible interaction between the modes is encircled.

Let us estimate the value of angle ϑ for which the linear transformation is quite effective. The following relations approximately describe the behavior of converging roots of Eq. (29) in the vicinity of the interaction region:

$$\lambda_{1,2} \approx 1 + \frac{1-\mu}{8\gamma^2} \pm \sqrt{\left(\frac{1-\mu}{8\gamma^2}\right)^2 + \frac{\vartheta^2}{8\gamma^2}}.$$

These relation holds when

$$\vartheta \ll \frac{1}{\gamma}. \tag{30}$$

Carrying out the transition to new characteristic exponents ($\lambda_{\text{mid}} = 1$) as in relations (15)–(17), we obtain

$$\tilde{\lambda}_{1,2} \approx \frac{1-\mu}{8\gamma^2} \pm \sqrt{\left(\frac{1-\mu}{8\gamma^2}\right)^2 + \frac{\vartheta^2}{8\gamma^2}}. \tag{31}$$

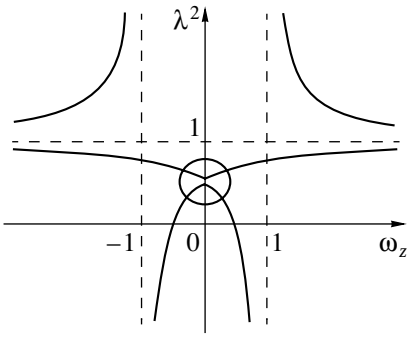

Fig. 2. Dependence of the squared refractive indices on the longitudinal component of the external magnetic field for quasi-transverse waves in the solar corona plasma.

At the point of maximum convergence $\mu = 1$, we have

$$\tilde{\lambda}_{1,2} = \pm\delta\lambda = \pm\frac{\vartheta}{2\sqrt{2}\gamma}. \quad (32)$$

Equation (29) also readily gives the values of the other two roots for $\mu = 1$ in the zeroth approximation in ϑ:

$$\lambda_3 \approx -1, \quad \lambda_4 \approx \beta + \frac{1}{2\gamma^2} \approx 1 + \frac{1}{\gamma^2}$$

or

$$\tilde{\lambda}_3 \approx -2, \quad \tilde{\lambda}_4 \approx \frac{1}{\gamma^2}. \quad (33)$$

Substituting relations (28) and (31)–(33) into formulas (22) (assuming that $i\tilde{\lambda} \equiv \tilde{p}^{(0)}$), we calculate the WKB phase for the waves under investigation for $\mu = 1$:

$$\left|\tilde{\Phi}_k^{(0)}\right| \approx \delta\lambda = \frac{\vartheta}{2\sqrt{2}\gamma}, \quad \left|\tilde{\Phi}_{k1}^{(1)}\right| \approx \frac{\vartheta\gamma\mu'}{2\sqrt{2}},$$

$$\left|\tilde{\Phi}_{k2}^{(1)}\right| \approx \frac{\mu'}{4\sqrt{2}\vartheta\gamma}.$$

By virtue of condition (30), we have $\left|\tilde{\Phi}_{k1}^{(1)}\right| \ll \left|\tilde{\Phi}_{k2}^{(1)}\right|$. Using estimate (25) or (23), we obtain

$$\frac{\vartheta}{\gamma} \gtrsim \frac{\mu'}{\vartheta\gamma}.$$

Considering that $\mu' \sim l/L$ (where L is the inhomogeneity characteristic scale), we find that the linear transformation is effective for waves propagating at angles

$$\vartheta \lesssim \sqrt{\frac{l}{L}}. \quad (34)$$

This estimate coincides with that obtained in [9] by the Zwan method.

3.2. The Case $\left|\tilde{\Phi}_{k2}^{(1)}\right| \ll \left|\tilde{\Phi}_{k1}^{(1)}\right|$. Transformation of Waves in the Solar Corona

Following Zheleznyakov [3, Section 24], we consider the transformation of quasi-transverse waves in a magnetically active plasma in a nonuniform external magnetic field. This situation takes place in the solar corona. The basic equations are

$$\begin{aligned} F_1'' + (C + B)F_1 &= -iAF_2, \\ F_2'' + (C - B)F_2 &= iAF_1, \end{aligned} \quad (35)$$

where F_i are complex field variables, the dimensionless variable $\zeta = \omega z/c$ is chosen as an independent variable, and

$$\begin{aligned} A &= -\frac{\nu\omega_y^2}{2\omega_y^2 - (1-\nu)(1-\omega_z^2)}, \\ B &= -\frac{\nu(1-\nu)\omega_z}{2\omega_y^2 - (1-\nu)(1-\omega_z^2)}, \\ C &= 1 - \frac{\nu(\omega_y^2 - 1 + \nu)}{2\omega_y^2 - (1-\nu)(1-\omega_z^2)}. \end{aligned} \quad (36)$$

Here, the quantity $\omega_z = \omega_z(\varepsilon\zeta)$ is assumed to vary monotonically and to pass through zero.

The system of equations (35) can be reduced to the following equation:

$$F_1^{(4)} - \frac{2A'}{A}F_1^{(3)} + 2CF_1^{(2)} + 2\left(C' + B' - (C + B)\frac{A'}{A}\right)F_1^{(1)} + (C^2 - B^2 - A^2)F_1 = 0. \quad (37)$$

Comparing this equation with Eq. (2), we observe that

$$b_1\mu' = -\frac{2A'}{A}, \quad b_3\mu' = 2\left(C' + B' - (C + B)\frac{A'}{A}\right), \quad b_2 = b_4 = 0. \quad (38)$$

In the zeroth approximation in ε, Eq. (37) corresponds to the dispersion equation

$$\lambda^4 - 2C\lambda^2 + (C^2 - B^2 - A^2) = 0, \quad (39)$$

where $i\lambda \equiv p^{(0)}$ as before.

The dependence $\lambda^2(\omega_z)$ is presented in Fig. 2. In the encircled neighborhood of $\omega_z = 0$, two dispersion curves converge and the transformation of the waves becomes possible. The small parameter responsible for the convergence of the two curves is

$$\omega_y^2 \ll 1. \quad (40)$$

Besides, we assume that

$$\nu \ll 1. \tag{41}$$

For $\omega_y^2 \longrightarrow 0$, a situation similar to that considered in the previous subsection takes place: the dispersion curves are reconnected, and the linear transformation of waves passing through the region $\omega_z \approx 0$ for small ω_y^2 ensures the continuity of this limiting transition.

It should be noted that the slope of the dispersion curves at the point of their maximum convergence is equal to zero ($\lambda' = 0$) and the third term in the WKB phase (20) for these curves vanishes: $\tilde{\Phi}_{k2}^{(1)} = 0$. Thus, the WKB approximation may be violated in the vicinity of $\omega_z = 0$ due to the second term $\tilde{\Phi}_{k1}^{(1)}$ in relation (20). Let us estimate the value of this term.

For $\omega_z = 0$, we have $B = 0$, and the values of the converging roots of the dispersion equation (39) are given by

$$\lambda_{1,2} = \sqrt{C \pm A}.$$

A transition to the variables (15)–(17) symmetric relative to the mean value gives

$$\tilde{\lambda}_{1,2} = \pm\delta\lambda = \pm(\sqrt{C+A} - \sqrt{C-A}). \tag{42}$$

Substituting into these expressions relations (36), we obtain in the first approximation in ν,

$$\tilde{\lambda}_{1,2} = \pm\frac{\nu\omega_y^2}{1-2\omega_y^2}. \tag{43}$$

For two distant roots of the dispersion equation (39) in the zeroth approximation in ν, we put $A = B = 0$, $C = 1$ (see relations (36)), and obtain $\lambda_{3,4} = -1$ or

$$\tilde{\lambda}_{3,4} = -2. \tag{44}$$

Substituting now relations (38) and (42)–(44) into the first two expressions from (22) and taking into account the fact that $i\lambda \equiv p^{(0)}$, we obtain the values of the WKB phases at point $\omega_z = 0$:

$$\left|\tilde{\Phi}_k^{(0)}\right| \approx \delta\lambda = \frac{\nu\omega_y^2}{1-2\omega_y^2}, \quad \left|\tilde{\Phi}_{k1}^{(1)}\right| \approx \frac{|\omega_z'|}{4\omega_y^2}.$$

Using relation (40) and considering that $\omega_z' = (c/\omega)d\omega_z/dz$, we obtain the following estimate for the ratio of the parameters, for which the transformation is effective:

$$\frac{\omega\nu\omega_y^4}{c\left|\dfrac{d\omega_z}{dz}\right|} \lesssim 1. \tag{45}$$

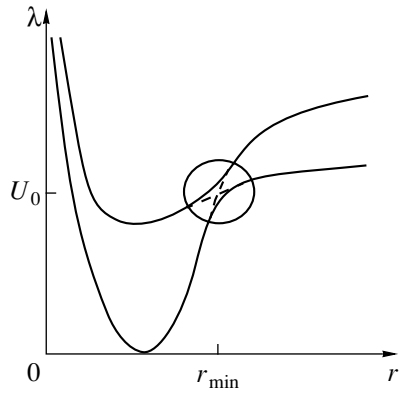

Fig. 3. Dependence of the terms of a quasimolecule formed by two colliding atoms on the distance between them.

This estimate coincides with that obtained in [3] by the method of phase integrals to within factors of the order of unity.

3.3. General Case $\left|\tilde{\Phi}_{k1}^{(1)}\right| \sim \left|\tilde{\Phi}_{k2}^{(1)}\right|$. Nonadiabatic Landau–Zener transitions

Following [10, Section 90], we consider a quantum system of two colliding atoms with quasi-intersecting terms (Fig. 3). In the vicinity of point $r = r_{\min}$, the system is described by the equations

$$\begin{aligned} i\hbar\frac{dy_1}{dt} &= U_1 y_1 + V y_2, \\ i\hbar\frac{dy_2}{dt} &= U_2 y_2 + V y_1, \end{aligned} \tag{46}$$

where $U_{1,2} = U_{1,2}(r)$, $r = \varepsilon t$ is the distance between the atoms, ε is the velocity of their relative motion, which is a small parameter of slowness, and $|V| \ll |U_{1,2}|$. Equations (46) can be reduced to the form

$$y_1'' + i(U_1 + U_2)y_1' + (V^2 - U_1U_2 + iU_1')y_1 = 0. \tag{47}$$

Here, the prime indicates the differentiation with respect to variable $\hbar^{-1}t$. A comparison of Eqs. (47) and (2) shows that in this case,

$$b_1 = 0, \quad b_2\mu' = iU_1'. \tag{48}$$

In the zeroth approximation in ε, Eq. (47) corresponds to the dispersion equation

$$\lambda^2 - (U_1 + U_2)\lambda + (V^2 - U_1U_2) = 0, \tag{49}$$

where $-i\lambda \equiv p^{(0)}$.

The roots of Eq. (49) are given by

$$\lambda_{1,2} = \frac{U_1 + U_2}{2} \pm \sqrt{\frac{(U_1 - U_2)^2}{4} + V^2}. \tag{50}$$

It can easily be seen that at the point $r_{\min}$ corresponding to the minimum distance between the terms, the condition $U_1(r_{\min}) = U_2(r_{\min}) = U_0$ is satisfied. In this case,

$$\lambda_{\text{mid}} = U_0, \quad \delta\lambda = V. \tag{51}$$

A transition to the variables (15)–(17) symmetric relative to the mean value gives

$$\tilde{\lambda}_{1,2} = \frac{(U_1 + U_2) - 2U_0}{2} \pm \sqrt{\frac{(U_1 - U_2)^2}{4} + V^2}. \tag{52}$$

Substituting now expressions (48), (51), and (52) into relations (20) and (22) and considering that $-i\lambda \equiv p^{(0)}$, we obtain the values of the WKB phases at point $r_{\min}$:

$$\tilde{\Phi}_k^{(0)} = \pm V, \quad \tilde{\Phi}_{k1}^{(1)} \approx \pm\frac{U_1'}{2V}, \quad \tilde{\Phi}_{k2}^{(1)} \approx \mp\frac{U_1' + U_2'}{4V}.$$

It can be seen that the phases $\tilde{\Phi}_{k1}^{(1)}$ and $\tilde{\Phi}_{k2}^{(1)}$ here are of the same order of magnitude and the general estimate (25) should be used. As a result, going over to the derivatives with respect to r, we find that nonadiabatic transitions between the terms under investigation are effective for

$$\frac{V^2}{\hbar\varepsilon\left|\frac{dU_1}{dr} - \frac{dU_2}{dr}\right|_{r = r_{\min}}} \lesssim 1. \tag{53}$$

This formula describes the nonadiabatic transitions in the Landau–Zener model to within a factor of the order of unity. If the estimate (53) is valid, the probability of the nonadiabatic transition between the terms is close to unity for the passage through point $r = r_{\min}$ and is exponentially small in the opposite case.

4. CONCLUSIONS

We have proposed a method for estimating the effectiveness of linear transformation of waves in a smoothly inhomogeneous medium in the case when their characteristic indices (in fact, refractive index) locally converge. The same method is suitable for estimating the effectiveness of nonadiabatic transitions in slow atomic collisions. It is based on the WKB formalism for a scalar nth order ordinary linear differential equation [8]. As compared to the known methods of complex phase integrals [2–5] and the vector method [1], our method has the following advantages and drawbacks.

1. The estimation of the effectiveness of wave interaction on the basis of the proposed method virtually does not require calculations. In each specific case, it is only necessary to know the dispersion of the waves in the region of interaction and the values of the coefficients in the equation in order to use condition (25) for the effectiveness of the interaction between waves. In simpler cases which are often encountered, estimates (23) and (24) can be used. If, for example, the equation does not contain the terms proportional to the derivatives of the parameters ($b_j \equiv 0$, $\tilde{\Phi}_{k1}^{(1)} = 0$), the estimation of the effectiveness of wave transformation requires only the knowledge of the dispersion of the interacting waves (an example of such a problem can be found in [11]). If the dependences of the refractive indices on the coordinate at the point of their maximum convergence have horizontal tangents, then $\tilde{\Phi}_{k2}^{(1)} = 0$, and estimate (24) can be used (see example in Subsection 3.2).

2. The proposed method can be extended virtually without modification to the case of three and more interacting waves (see [11]).

3. Besides, this method makes it possible to estimate easily the interaction of waves for which several parameters vary smoothly. All the formulas in this case remain unchanged, but μ now stands for a set of s smoothly varying parameters: $\mu \equiv (\mu_1, \ldots, \mu_s) \in \mathbf{R}^s$; the quantities b_j are also vectors in $\mathbf{R}^s$, which experience the convolution in the scalar product with μ'.

4. A drawback of this method is that while providing an estimate for the effectiveness of interaction, it does not directly indicate the energy redistribution between the interacting waves. However, first, in some problems, it is sufficient to only estimate the effectiveness of the interaction (as, for example, in [9]), and second, the energy redistribution can be easily determined from other physical considerations (e.g., the form of wave polarization). For instance, if the stringent inequalities opposite to (23)–(25) are satisfied, the interaction between the waves can be neglected and the geometrical optics approximation can be used. If the stringent inequality in one of the conditions (23)–(25) holds, the approximation of abrupt jump of the parameters can be used, and solutions with close polarizations at the opposite sides of the jump can be joined (see [11]).

5. It should be noted that in some cases, this method is inapplicable, while the method of complex phase integrals holds. Such a situation is observed when the interaction between waves takes place near reflection points at which the refractive indices vanish or become infinitely large (see, for example, the interaction of quasi-longitudinal waves in a magnetically active plasma [2, 3]).

The general formalism of the proposed method might probably appear as more complex and less visual as compared, for example, with the method in [1]. This is natural since in other methods, all calculations must be made anew each time, starting from the initial equations, while in the method proposed by us all main calculations were made by obtaining the universal estimates (23)–(25). The examples considered above make it possible to speak of the complete coincidence of the results obtained by the given method with the results obtained by the classical methods.

Apart from the applications considered here, the results obtained can also be used in mechanics and in the theory of dynamic systems. Conditions (23)–(25) are the conditions of violation of the adiabatic approximation in the case of close eigenvalues of a system of the general form (2).

REFERENCES

1. V. V. Zheleznyakov, V. V. Kocharovskiĭ, and Vl. V. Kocharovskiĭ, Usp. Fiz. Nauk **141** (2), 257 (1983) [Sov. Phys. Usp. **26**, 877 (1983)].
2. V. L. Ginzburg, *The Propagation of Electromagnetic Waves in Plasmas* (Nauka, Moscow, 1967; Pergamon, Oxford, 1970).
3. V. V. Zheleznyakov, *Radio Emission from the Sun and Planets* (Nauka, Moscow, 1964).
4. N. S. Erokhin and S. S. Moiseev, in *Reviews of Plasma Physics*, Ed. by M. A. Leontovich (Atomizdat, Moscow, 1973; Consultants Bureau, New York, 1979), Vol. 7.
5. E. A. Solov'ev, Usp. Fiz. Nauk **157** (3), 437 (1989) [Sov. Phys. Usp. **32**, 228 (1989)].
6. E. E. Nikitin and B. M. Smirnov, *Slow Atomic Collisions* (Énergoatomizdat, Moscow, 1990).
7. V. V. Kocharovskiĭ and Vl. V. Kocharovskiĭ, *Physical Encyclopedia* (Sovetskaya Éntsiklopediya, Moscow, 1990), Vol. 2.
8. M. V. Fedoryuk, *Asymptotic Methods for Linear Ordinary Differential Equations* (Nauka, Moscow, 1983).
9. K. Yu. Bliokh and Yu. É. Lyubarskiĭ, Pis'ma Astron. Zh. **22** (7), 539 (1996) [Astron. Lett. **22**, 482 (1996)].
10. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 3: *Quantum Mechanics: Non-Relativistic Theory* (Fizmatgiz, Moscow, 1963; Pergamon, New York, 1977).
11. K. Yu. Bliokh, Radiofiz. Radioastron. **3**, 49 (1998).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 93, No. 1, 2001, pp. 80–88.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 1, 2001, pp. 94–103.

FLUIDS

Rheological Properties of Polydisperse Magnetic Fluids. Effect of Chain Aggregates

A. Yu. Zubarev
Ural State University, Yekaterinburg, 620083 Russia
e-mail: andrey.zubarev@usu.ru
Received August 9, 2000

Abstract—A theoretical model of medium-density polydisperse magnetic fluids is proposed. The model takes into account that the major fraction of particles in typical ferrofluids is characterized by a magnetic core diameter of about 10 nm. In addition, there is a certain proportion of large particles with a core diameter of about 16 nm. As a result of the magnetic dipole interaction, the large particles form chain aggregates. Small particles, for which the magnetic dipole interaction energy (both with each other and with large particles) is smaller than the thermal energy, remain in the individual nonaggregated state. The distribution of chains with respect to the number of (large) particles and some rheological characteristics of the ferrofluids are determined. The proposed model is capable of explaining, in principle, the giant magnetoviscosity effect and a strong dependence of the rheological properties of ferrofluids on the shear rate observed in some recent experiments. © *2001 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Magnetic fluids (ferrofluids, ferrocolloids) are stable colloidal suspensions of monodomain ferromagnetic particles in a carrying liquid medium. A small size of particles (which are typically on the order of 10 nm in diameter) accounts for their intensive Brownian movements and explains a high sedimentation stability of these colloids. In order to prevent the coalescence of particles under the action of van der Waals forces, the particles are usually coated with stabilizing layers of a polymeric or an ion-electrostatic nature. As a result, the particles in ferrofluids interact with each other by means of magnetic dipole, sterical, and hydrodynamic forces. The thickness of stabilizing polymer layers is usually on the order of 2–3 nm, so that the hydrodynamic radius of particles may significantly differ from the magnetic core radius.

One of the most interesting and important properties of ferrofluids is the possibility of controlling their hydrodynamic flows and rheological properties by applying external magnetic fields. The magnetic interactions of particles may significantly change their hydrodynamic behavior and the entire system rheology.

Previously [1], the dynamic characteristics of medium-density ferrofluids were theoretically studied for the systems composed of individual particles interacting with each other. It was demonstrated that the magnetic and hydrodynamic interactions of particles may lead to a significant (reaching several tens per cent) increase in the effective viscosity of real magnetic fluids. It should be noted that the well-known models of dilute ferrocolloids completely ignoring the interactions between particles (see, e.g., [2–4]) predict a field-induced increase in viscosity by only a few percent under otherwise equal conditions.

Recently, Odenbach *et al.* [5–7] experimentally observed a giant increase in the viscosity of typical ferrofluids in moderate and weak magnetic fields and a strong dependence of the viscosity on the shear rate. In particular [6], the viscosity observed at a shear rate of 0.1 s^{-1} and a field strength in a sample on the order of 20 kA/m was greater by a factor of 14–16 as compared to the value obtained with the field switched off. At a shear rate of about 1 s^{-1}, the relative increase in viscosity in the same field dropped to 2.5 [5, 6], and an increase in the shear rate up to 500 s^{-1} reduced the viscosity growth down to about 5% [7]. It should be noted that the experimental results reported in [7] satisfactorily agree with the theoretical estimates obtained in [1]. Neither the theories of moderately concentrated homogeneous ferrocolloids composed of interacting individual particles (see, e.g., [1]) nor the models of dilute ferrofluids [2–4] can explain this strong dependence of the effective viscosity on the field strength and shear rate. A very large (reaching tens and hundreds times) discrepancy between the experimental data [5–7] and the theories of homogeneous ferrocolloids composed of individual particles indicates that real magnetic fluids feature a qualitatively new physical situation not taken into account by the aforementioned models.

The results of experiments reported in [5–7] can be rationalized by taking into account that real ferrofluids always represent polydisperse systems. The energy of the magnetic dipole interaction between the largest particles in typical ferrocolloids (such as those used in the experiments described in [5–7]) are considerably

1063-7761/01/9301-0080$21.00

greater than their thermal energy at room temperature. Under these conditions, large particles may combine with each other to form chainlike aggregates, the influence of which on the rheological properties of ferrofluids can be very significant [8–10]. This effect can be especially pronounced when (as in the experiments described in [5–7]) the external field is oriented along the flow velocity gradient.

The content of large particles in typical ferrofluids is usually rather small: on the order of 1% (a typical particle size distribution histogram from [7] is reproduced in Fig. 1). The main fraction represents small-size particles. The interactions of these particles with each other and with large ones are characterized by small energies (so that the particles occur in the individual state) and weakly influence the magnetic and rheological properties of the colloidal system. The monodisperse models of (actually) polydisperse magnetic fluids naturally operate with particles possessing an system-averaged size, that is, with small particles. Therefore, these models cannot provide for a satisfactory description of the properties of polydisperse systems, except the case of ultimately diluted solutions where no chain aggregates or other heterostructures can appear.

In previous papers [9, 10], statistical models of monodisperse ferrofluids containing chain aggregates were proposed and rheological characteristics of such systems were estimated (references to alternative models of ferrocolloids with particle chains were also given). In this study, the approach developed in [9, 10] is expanded so as to include polydisperse media as well. It is demonstrated that a strong field dependence of the rheological properties of ferrocolloids are determined by a small fraction of large particles capable of combining into chainlike clusters. As the shear rate increases, the chains break, this leading to a rapid decrease in the effective viscosity.

2. CHAIN SIZE DISTRIBUTION

As noted above, the models of ferrofluids composed of individual particles cannot explain large magnetoviscosity effects observed in [5–7]. An analysis of the rheological properties of suspensions indicates that a strong increase in the effective viscosity under the action of an orienting field is possible, provided that a suspension contains nonspherical, sufficiently elongated objects (particles, aggregates, etc.) [9–11]. In a conventional ferrocolloid, such objects can be represented only by chainlike or bulky (droplike) aggregates. The latter acquire elongated shape under the action of an applied magnetic field. The appearance of droplike aggregates in polydisperse ferrocolloids characterized by a particle size histogram such as that determined in [7] (Fig. 1) is hardly probable. Indeed, simple estimates show that typical magnetite particles (see the characteristics in [5– 7]) possess a magnetic core with a diameter of 10 nm. This core size is so small that the energy of magnetic dipole interactions is considerably smaller

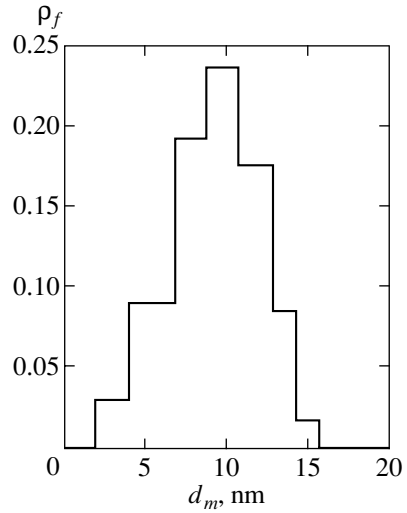

Fig. 1. A histogram of the particle distribution with respect to the magnetic core diameter d_m in a ferrofluid experimentally studied in [5–7, 14] (ρ_f is the volume concentration of a given particle fraction).

than the thermal energy of the system. Therefore, only the largest particles are capable of forming linear of bulky aggregates. However, the volume fraction of small particles and, hence, the osmotic pressure produced by this fraction, is large. When a bulky aggregate of large particles (present in a very small concentration) appears, small particles (driven by the osmotic pressure) tend inside the aggregate. As a result, the small particles (neutral with respect to the magnetic dipole interactions) occupy a part of the aggregate volume. As a result, the concentration of large particles (involved into the magnetic dipole interaction) in the aggregate decreases and the structure breaks. Thus, it is much more probable that large particles form linear (chainlike) structures. It must be emphasized that the principal possibility of the existence of bulky aggregates in polydisperse ferrocolloids is by no means rejected. This question is certainly worth special investigation. At the same time, it is hardly possible to take into account the effects due to all probable heterostructures within the framework of a single model. For this reason, the consecration below is concentrated on the chainlike clusters that seem more likely to form. Note that the effect of droplike aggregates on the rheological properties of monodisperse metastable ferrocolloids featuring the process of separation into dense and dilute phases was estimated in [12].

The well-known difficulties encountered in the statistical physics of dense systems hinder the formulation of a rigorous theory of ferrofluids containing even a single type of linear aggregates. In order to obtain constructive results, we will use the following strong approximations.

First, instead of real polydisperse systems with a more or less smooth particle size distribution function,

we will consider a model bidisperse medium composed of "small" and "large" particles. It will be assumed that the size of small particles virtually coincides with an average particle size in the given ferrocolloid (about 9–10 nm). The size of large particles and their volume concentration (assumed to be small) will be estimated below.

Second, by analogy with models [9, 10], it is assumed that the chains formed by large particles appear as straight rigid (rodlike) aggregates The limits of this approximation were also assessed in [9, 10].

Third, the chains are assumed to include only large particles, whereas small particles occur only in the nonaggregated individual state. Inside each chain, only the energy of the magnetic dipole interaction between adjacent (nearest neighbor) particles is taken into account. The interaction between chains (including monomers) is ignored. In a real system, the interaction between chains can be significant, even leading to their condensation. However, it seems reasonable to restrict the consideration initially to the "ideal chain gas." The problem of interchain interaction requires special consideration.

Fourth, also by analogy with models [9, 10], the consideration is restricted to the systems in which the energy of the magnetic dipole interaction between adjacent particles in the chain is greater than the energy of the particle interaction with the external field. This assumption is valid under usual experimental conditions [5–7].

In the proposed model, the bidisperse medium is "formed" as follows. In the first step, we take into account only the small particles. The ferrocolloid of small particles is considered as a homogeneous ferrofluid, the rheological properties of which can be described, for example, using the theory developed in [1] (if the magnetic moments of small particles are rigidly bound to their bodies) or using the theory of "neutral" suspensions (if the particles behave as superparamagnets). Then large particles are placed in the model "base" ferrofluid and the chains formed of such particles are taken into consideration. Simple estimates show that the energy of interaction between small and large particles in this ferrofluid (e.g., see the histogram in Fig. 1 and the particle characteristics in [5, 7]) is small as compared to the characteristic thermal energy of the system. Therefore, we may neglect the interaction of large particles with the surrounding weakly paramagnetic base ferrofluid and describe the chain formation as a process proceeding in a magnetically neutral medium.

It is obvious that, employing the above set of strong assumptions, the proposed model can be used for estimating values by the order of magnitude, rather than for obtaining precise quantitative results. Nevertheless, the results of calculations performed within the framework of this model coincide with the experimental values reported in [5–7] to within several tens percent. Taking into account that the other existing theories deviate from these experimental data by one–two orders of magnitude, the observed coincidence can be considered as evidence of the qualitative adequacy of the proposed model.

Using the approach proposed in [9, 10], let us estimate the equilibrium distribution function g_n of chains with respect to the number of (large) particles (or the number of n-member chains per unit volume of the magnetic fluid under consideration). Within the framework of the approximations adopted, the free energy F per unit volume of the ferrocolloid containing large-particle chains can be written as [9, 10]

$$F = T\sum_{n=1}^{\infty}\left(g_n\ln\frac{g_n v_l}{e} + g_n f_n\right),$$
$$v_l = \frac{4\pi}{3}a_l^3, \quad e = 2.7\ldots. \tag{1}$$

Here, T is the absolute temperature (expressed in energy units), a_l is the hydrodynamic radius of a large particle (that is, the magnetic core radius plus stabilizing layer thickness). The first term in parentheses (1) represents the entropy of the ideal gas of n-particle chains determined by their translational motion; f_n is a dimensionless "internal" free energy on an n-particle chain determined by the interaction of particles with each other and with the external magnetic field and by the thermal rotational motion of chains. It should be recalled that the particle aggregates are modeled by straight rigid rods, rather than by flexible polymeric chains; this approximation is justified if the average chain length is shorter than the persistent length. The corresponding estimates showing the validity of this assumption were obtained in [9, 10], where it was also demonstrated that, within the framework of the approximations adopted, the following relationship can be used:

$$f_n = -[\varepsilon(n-1) - f_{nH}(\kappa)], \tag{2}$$

where

$$\boldsymbol{\kappa} = \frac{m_l\mathbf{H}}{T}, \quad \varepsilon = \frac{m_l^2}{4a_l^3 T}. \tag{3}$$

$\mathbf{H}$ is the macroscopic magnetic field inside the sample volume, m_l is the absolute value of the magnetic moment of a large particle, f_{nH} is the free energy of the interaction between an n-particle chain and the magnetic field. The last value can be expressed as

$$f_{nH} = \int\phi_n\ln\frac{\phi_n}{2.72\ldots}d\mathbf{e} - \int(\mathbf{e}\kappa)\phi_n d\mathbf{e}, \tag{4}$$

where $\mathbf{e}$ is the unit vector directed along the magnetic moment of a particle (i.e., along the chain containing

this particle) and $\phi_n(\mathbf{e})$ is the orientation distribution function for an n-particle chain (normalized to unity).

In the equilibrium state, the system obeys the following relationships [9, 10]:

$$\phi_n = \phi_n^\circ = \frac{1}{4\pi}\frac{\kappa n}{\sinh \kappa n}\exp(\mathbf{e}\cdot\boldsymbol{\kappa} n),$$
$$f_{nH}^\circ = -\ln\left(\frac{\sinh \kappa n}{\kappa n}\right) \quad (5)$$

(here and below, the circle superscript refers to quantities determined in the equilibrium state).

For nonequilibrium systems, such as those in the state of a stationary shear flow, a possible method of determining a stable internal structure is based on the search for a minimum free energy with an allowance for forces inducing deviations of the chain orientations from equilibrium direction (along the magnetic field).

Minimization of the free energy functional (1) is performed with an allowance for the obvious normalization condition

$$\sum_{n=1}^{\infty} n g_n = \frac{\varphi}{v_l}, \quad (6)$$

where φ is the volume concentration of large particles to be determined. Minimizing (1) and taking into account (6), we obtain

$$g_n = \frac{x^n}{v_l}\exp(-f_{nH})\exp(-\varepsilon). \quad (7)$$

In order to determine the parameter x, formula (7) has to be substituted into the normalization condition (6). In the general case, this leads to a transcendental equation that can be readily solved by numerical methods. However, the x value for the equilibrium system can be determined analytically. Indeed, substituting the second relationship (5) into (7), then (7) into (6), and accomplishing the necessary transformations (see [9]), we eventually obtain

$$g_n^0 = \frac{(x^0)^n}{v_l}\frac{\sinh \kappa n}{\kappa n}\exp(-\varepsilon), \quad (8)$$

where

$$x^0 = \frac{2y\cosh\kappa - \sinh\kappa - \sqrt{(2y\cosh\kappa - \sinh\kappa)^2 - 4y^2}}{2y},$$
$$y = \kappa\varphi\exp\varepsilon.$$

The average number of particles in the chain for both the equilibrium and nonequilibrium state can be determined as the ratio of the total number of particles per unit volume of the colloid to the total number of chains in this volume:

$$\langle n\rangle = \frac{\sum_n n g_n}{\sum_n g_n} = \frac{\varphi}{v_l\sum_n g_n}. \quad (9)$$

3. RHEOLOGICAL CHARACTERISTICS OF FERROCOLLOIDS AT A NEGLIGIBLY SMALL SHEAR RATE

Let us consider a magnetic fluid occurring in the state of shear flow. In this section, it will be assumed that the shear rate is sufficiently small, so that we may neglect the deformation of chains in the flow and a deviation of the distribution function g_n from the equilibrium value g_n^0. It was previously demonstrated [9] that the chain deformation can be ignored, provided that the flow velocity gradient ν obeys the following strong inequality:

$$\nu \ll \frac{T\varepsilon}{\eta^0 d_l^3\langle n\rangle},$$

where $d_l = 2a_l$ and η^0 is the "base" medium viscosity. In our model, this medium is represented by the effective magnetic fluid composed of small particles. Simple estimates show that the above inequality is valid in most of the real situations.

Since the appearance of Einstein's classical paper, it is well known that the rheological properties of suspensions are determined by the perturbations introduced by particles into the hydrodynamic flow. It is virtually impossible to calculate rigorously such a perturbation introduced by a particle chain, even modeled by a rodlike aggregate. In order to obtain constructive estimates, let us model (as in [9, 10]) an n-particle chain by an ellipsoid of revolution with a semiminor axis a_l and semimajor axis $a_l n$. It is important that the volume of such a spheroidal body is equal to the sum of volumes of the component particles.

Using well-known results of the statistical hydromechanics of dilute suspensions composed of solid spheroids (see, e.g., [11]), expressions for the Cartesian components of the mean stress tensor $\boldsymbol{\sigma}$ can be written in the following form:

$$\sigma_{ik} = \sigma_{ik}^s + \sigma_{ik}^a,$$
$$\sigma_{ik}^s = 2\eta^0\gamma_{ik} + \eta^0\Big\langle\Big\langle [(2\alpha_n\gamma_{ik} - \rho_n\langle e_j e_s\rangle_n\delta_{ik}\gamma_{js})$$
$$+ (\zeta_n + \beta_n\lambda_n)(\langle e_i e_j\rangle_n\gamma_{jk} + \langle e_k e_j\rangle_n\gamma_{ji})$$
$$+ \beta_n(\omega_{ij}\langle e_j e_k\rangle_n + \omega_{kj}\langle e_j e_i\rangle_n) + (\chi_n - 2\lambda_n\beta_n) \quad (10)$$

$$\times \langle e_i e_k e_j e_s \rangle_n \gamma_{js} - \beta_n \frac{d}{dt} \langle e_i e_k \rangle_n] \Big\rangle\Big\rangle,$$

$$\sigma_{ik}^a = \frac{kT}{2\nu_l} \langle\langle \langle e_l \rangle h_k - \langle e_k \rangle h_i \rangle\rangle, \quad h_i = \frac{H_i}{H},$$

$$\gamma_{ik} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_k} + \frac{\partial u_k}{\partial x_i}\right), \quad \omega_{ik} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_k} - \frac{\partial u_k}{\partial x_i}\right).$$

Here and in what follows,

$$\langle\langle \ldots \rangle\rangle = \sum_n \ldots n\nu_l g_n,$$

$$\langle \ldots \rangle = \int \ldots \mathbf{e} \phi_n(\mathbf{r}) d\mathbf{e}, \quad \langle \ldots \rangle^0 = \int \ldots \mathbf{e} \phi_n^0(\mathbf{e}) d\mathbf{e},$$

$$i, \ldots, k = x, y, z.$$

Vector **u** is the average flow velocity in the suspension; $\alpha_n \ldots \zeta_n$ are the parameters determined in [13], and tensors $\boldsymbol{\sigma}^s$ and $\boldsymbol{\sigma}^a$ represent the symmetric and antisymmetric parts of the total stress tensor $\boldsymbol{\sigma}$.[1]

In order to calculate the moments entering into expression (10), it is necessary to determine the nonequilibrium orientation distribution function ϕ_n. This function can be obtained as a solution to the corresponding Fokker–Planck equation for a spheroidal particle (see, e.g., [11, 13]):

$$\frac{\partial \phi_n}{\partial t} = -\mathbf{I}\left(\phi_n \frac{d\mathbf{e}}{dt}\right) + D_n \mathbf{I}^2 \phi_n, \tag{11}$$

where

$$\frac{d\mathbf{e}}{dt} = -\frac{D_n}{T}\mathbf{I}U + \boldsymbol{\omega}\mathbf{e} + \lambda_n(\boldsymbol{\gamma}\mathbf{e} - (\mathbf{e}\boldsymbol{\gamma}\mathbf{e})\mathbf{e}),$$

$$\mathbf{I} = \mathbf{e} \times \frac{\partial}{\partial \mathbf{e}}, \quad U = -m_l(\mathbf{e} \cdot \mathbf{H}) = -T(\mathbf{e}\boldsymbol{\kappa}), \tag{12}$$

$$\nu_{ij} = \frac{\partial u_i}{\partial x_j},$$

$$D_n = \frac{D}{n\delta_n}, \quad D = \frac{T}{6\eta^0 \nu_l}.$$

D is the rotational diffusion coefficient of a single large particle in the "base" fluid, D_n is the rotational diffusion coefficient of a spheroid modeling the n-particle chain, δ_n is a parameter determined in [11], U is the potential energy of an ellipsoid in a magnetic field, $\boldsymbol{\nu}$ is the tensor of the average flow velocity gradient, and $\boldsymbol{\omega}$ and $\boldsymbol{\gamma}$ are the tensors determined by relationships (10).

An exact solution to Eq. (11) is unknown. In order to find an approximate solution, we will use an approach developed in [2, 9] and represent the distribution function ϕ_n as the set-up function

$$\phi_n = \frac{1}{Z}\exp((\boldsymbol{\kappa} n + \mathbf{a}) \cdot \mathbf{e}) + (\mathbf{e} \cdot \mathbf{b} \cdot \mathbf{e})),$$
$$Z = \int \exp(((\boldsymbol{\kappa} n + \mathbf{a}) \cdot \mathbf{e}) + (\mathbf{e} \cdot \mathbf{b} \cdot \mathbf{e}))d\mathbf{e}, \tag{13}$$

where **a** and **b** are the unknown vector and second-rank tensor.

In this section, we study the flows with very small velocity gradients. This implies that deviations of the ϕ_n function from the equilibrium distribution ϕ_n^0 are small. Therefore, the parameters **a** and **b** are small as well. In the linear approximation with respect to these parameters, the first relationship (13) can be written as

$$\phi_n = \phi_n^0(1 + a_i(e_i - \langle e_i \rangle_n^0) + b_{ik}(e_i e_k - \langle e_i e_k \rangle_n^0)). \tag{14}$$

According to the approach developed in [2, 9], a_i and b_{ij} are determined from equations for the first and second moments of the orientation distribution function. Multiplying Eq. (11) by e_k, then by $e_i e_k - (1/3)\delta_{ik}$, and integrating with respect to **e**, we obtain (see, e.g., [11]):

$$\frac{d\langle e_k \rangle_n}{dt} = -\frac{1}{\tau_{1n}}\langle e_k \rangle_n + \lambda_n(\langle e_j \rangle_n \gamma_{jk} + \langle e_k e_j e_s \rangle_n \gamma_{js}) + \omega_{kj}\langle e_j \rangle_n + D_n \kappa n(h_k - \langle e_k e_j \rangle_n h_j) \tag{15}$$

and

$$\frac{d\langle e_i e_k \rangle_n}{dt} = -\frac{1}{\tau_{2n}}\left(\langle e_i e_k \rangle_n - \frac{1}{3}\delta_{ik}\right) + \lambda_n(\langle e_i e_s \rangle_n \gamma_{sk} + \langle e_k e_s \rangle_n \gamma_{si}) + \omega_{ij}\langle e_j e_k \rangle_n + \omega_{kj}\langle e_j e_i \rangle_n - 2\lambda_n \langle e_i e_k e_s e_j \rangle_n \gamma_{sj} + D_n \kappa n(\langle e_k \rangle_n h_i - 2\langle e_j e_i e_k \rangle_n h_j + \langle e_i \rangle_n h_k), \tag{16}$$

where

$$\tau_{1n} = \frac{1}{2D_n}, \quad \tau_{2n} = \frac{1}{6D_n}.$$

Substituting (14) into (15) and (16), we arrive in the linear approximation with respect to **a**, **b**, $\boldsymbol{\gamma}$, and $\boldsymbol{\omega}$ at a system of differential equations for the functions $a_i(t)$ and $b_{ik}(t)$ corresponding to a given n. Upon solving this system, we may use the function (14) for determining the nonequilibrium moments in (10). Note that the linear approximation in γ_{ij}, ω_{ij} corresponds to the linear approximation in a_i, b_{ij}. Within the framework of this

[1] The antisymmetric part of the stress tensor in an anisotropic medium exposed to an external orienting field arises due to the fact that the rotation speed of particles (molecules) differs (as a result of the blocking action of the field) from the rotation speed of the whole fluid (or its element). This leads to an additional (as compared to the isotropic medium) energy dissipation, which is reflected by a new rheological characteristic usually referred to as the rotational viscosity.

approximation, the moments $\langle\ldots\rangle_n$ multiplied by ω_{ij} or γ_{ij} must be replaced in (10), (15), and (16) by the corresponding equilibrium values $\langle\ldots\rangle_n^0$.

In the general case, the problem of calculating **a** and **b** is formally simple, but rather cumbersome. For definiteness, let us consider an oscillating flow of a ferrofluid in which the constant magnetic field is directed along the velocity gradient. In the Cartesian coordinate system (x, y, z), the flow velocity is $\mathbf{u} = (vz, 0, 0)$ and the field vector is $\mathbf{H} = (0, 0, H)$.

Using the Fourier transform with respect to time in (10), (15), and (16) and denoting the Fourier images by the same symbols as their prototypes, we obtain the following system of equations for a_x and b_{xz},

$$\left[\left(i\Omega + \frac{1}{\tau_{1n}}\right)\langle e_x^0\rangle_n^0 + D_n\kappa n\langle e_x^2 e_z\rangle_n^0\right]a_x$$

$$+\left[\left(i\Omega + \frac{1}{\tau_{1n}}\right)\langle e_x^2 e_z\rangle_n^0 + D_n\kappa n\langle e_x^2 e_z^2\rangle_n^0\right]2b_{xz} \qquad (17)$$

$$= [\lambda_n(\langle e_z^0\rangle_n^0 - 2\langle e_x^2 e_z\rangle_n^0) + \langle e_z\rangle_n^0]\frac{\nu}{2},$$

$$\left[\left(i\Omega + \frac{1}{\tau_{2n}}\right)\langle e_x^2 e_z\rangle_n^0 + D_n\kappa n(2\langle e_x^2 e_z^2\rangle_n^0 - \langle e_x^2\rangle_n^0)\right]a_x$$

$$+\left[\left(i\Omega + \frac{1}{\tau_{2n}}\right)\langle e_x^2 e_z^2\rangle_n^0\right. \qquad (18)$$

$$\left. + D_n\kappa n(2\langle e_x^2 e_z^3\rangle_n^0 - \langle e_x^2 e_z\rangle_n^0)\right]2b_{xz}$$

$$= [\lambda_n(\langle e_x^2\rangle_n^0 - 4\langle e_x^2 e_z^2\rangle_n^0 + \langle e_z^0\rangle_n^0)(\langle e_z^2\rangle_n^0 - \langle e_x^2\rangle_n^0)]\frac{\nu}{2},$$

and an expression for the hydrodynamic stress:

$$\sigma_{xz} = \eta_\Omega\nu,$$

$$\eta_\Omega = \eta^0\left[1 + \left\langle\left\langle \alpha_n + \frac{1}{2}((\zeta_n + \beta_n\lambda_n)(\langle e_x^0\rangle_n^0 + \langle e_z^2\rangle_n^0)\right.\right.\right.$$

$$+ \beta_n(\langle e_z^2\rangle_n^0 - \langle e_x^0\rangle_n^0) + 2(\chi_n - 2\lambda_n\beta_n)\langle e_x^2 e_z^2\rangle_n^0)$$

$$+ \frac{1}{2\nu_l}\frac{T\kappa}{\eta^0}(A_1\langle e_x^2\rangle_n^0 + B_1\langle e_x^2 e_z\rangle_n^0) \qquad (19)$$

$$\left.\left.\left. - i\Omega\beta_n(A_1\langle e_x^2 e_z\rangle_n^0 + B_1\langle e_x^2 e_z^2\rangle_n^0)\right\rangle\right\rangle\right],$$

$$A_1 = \frac{a_x}{2\nu}, \quad B_1 = 2\frac{b_{xz}}{2\nu}.$$

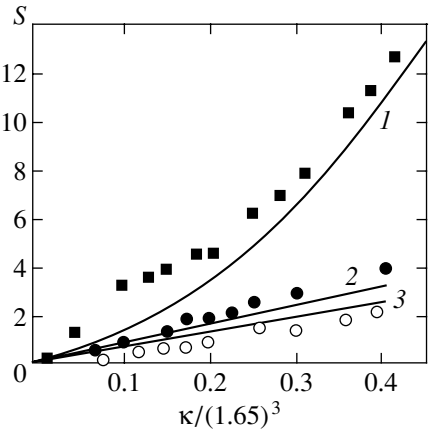

Fig. 2. Plots of the magnetoviscosity parameter S versus dimensionless magnetic field strength $\kappa = m_l H/T$. Symbols represent the experimental data taken from [6] for $\nu = 0.1$ (squares), 0.5 (filled circles), and 0.9 s^{-1} (open circles; curves *1–3* show the results of model calculations for $\nu \longrightarrow 0$, 0.5, and 0.9 s^{-1}, respectively.

Here, Ω is the Fourier frequency and η_Ω is the complex effective viscosity of the ferrofluid. Since a characteristic hydrodynamic relaxation time for the "base" magnetic medium containing only small particles is much smaller than the value for the "total" fluid with large particles and their chains, we may replace (without introducing large error) the complex effective viscosity of the "base" medium by its stationary vale η^0. An estimate of η^0 for a medium-concentration ferrocolloid with magnetically hard particles can be found, for example, in [1].

For the following calculations, it is convenient to introduce diameters d_{ml} and d_{ms} of the large and small particles, respectively. The magnetic moments m_l and m_s of these particles are related to their diameters by the obvious formulas

$$m_{l,s} = M_p\frac{\pi}{6}d_{ml,s}^3,$$

where M_p is the saturation magnetization of the particle material.

A relationship between the hydrodynamic radii $a_{l,s}$ and core diameters of the particles is also obvious:

$$2a_{l,s} = d_{ml,s} + 2s,$$

where s is the thickness of the stabilizing layer (typical values for real ferrofluids are s = 2–3 nm).

Figure 2 shows the experimental data (taken from [6]) for a ferrofluid with a particles size distribution function represented by a histogram in Fig. 1 (the physical characteristics can be found in [5, 7]). The experiments were performed using shear flows with various small velocity gradients. For comparison, Fig. 2 shows the results obtained by using the proposed model for

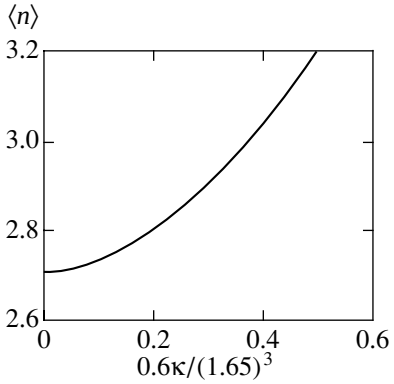

Fig. 3. The plot of average number of large particles in the chain n versus dimensionless magnetic field strength κ.

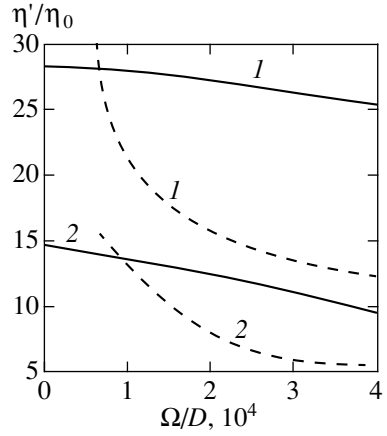

Fig. 4. The ratio of the real part of the complex viscosity to the viscosity of a solvent medium as a function of the flow oscillation frequency Ω for $\kappa = 1.73$ (*1*) and 1.15 (*2*). Solid curves show the results of calculations, dashed curves represent the experimental data from [6].

calculating the stationary ($\Omega = 0$) magnetoviscosity effect as characterized by the parameter

$$S = \frac{\eta(\kappa) - \eta(0)}{\eta(0)},$$

where η is a stationary value of the effective viscosity η_Ω. For this calculation, the diameter d_{ml} and the volume hydrodynamic fraction φ of large particles were selected so as to provide for the best fit of the values calculated for a negligibly small ν/D ratio to the experimental values for a minimum (in this study) shear rate ($\nu = 0.1\ s^{-1}$). The calculation yields $d_{ml} \approx 1.65$ nm and $\varphi \approx 0.017$. These values agree well with the independent estimates obtained in [14], where (in our notation) $d_{ml} \approx 1.6$ nm and $\varphi \approx 1.56\%$ (for $s = 2$ nm) and 2.07% ($s = 3$ nm). It should be noted that the results of calculations are highly sensitive with respect to the magnetic core diameter d_{ml}. This is explained by the fact that relationships (8) and (9) lead to exponential dependence of the average number $\langle n \rangle$ of particles in the chain on the ε value determined by (3) (for more detail, see [9]). In turn, the ε value is proportional to d_{ml}^3. For this reason, the characteristic chain length and, hence, the effective viscosity of the ferrocolloid strongly depend on the magnetic core diameter in large particles.

In the calculations, it was assumed that (in accordance with the data reported in [5, 7]) the total hydrodynamic concentration of particles in the ferrocolloid is 0.27. Therefore, the volume fraction of small particles is $\rho = 0.27–0.017$. For the small particles, the magnetic core diameter can be estimated using an obvious relationship

$$d_{ml}\varphi + d_{ml}\rho = \langle d \rangle \approx 9–10\ \text{nm},$$

which yields $d_{ms} \approx 8.5$ nm. As is known, in particles with the magnetic core diameter exceeding a certain threshold d_c, the magnetic moment is rigidly bound to the particle body. The effect of a magnetic field on the viscosity of such ferrofluids may be significant. If the core diameter is smaller than d_c, the particle behaves as a superparamagnet, that is, the magnetic moment may rotate freely with respect to the body. In this case, the viscosity of the ferrofluid is virtually not affected by the field. For magnetite particles used in the experiments described in [5–7, 14], the critical core size is $d_c \approx 12–14$ nm. Therefore, the "base" ferrofluid composed of small particles exhibit no magnetoviscosity effect; the effective viscosity of this medium can be estimated, for example, using the Batchelor–Green theory [15].

Figure 3 shows the results of calculations of the average number of particles in the chain, which elucidate the internal structure of ferrocolloids in the case of negligibly small velocity gradients ν.

Figure 4 presents the results of measurements of the real part η'_Ω of the complex viscosity for the aforementioned ferrofluid (studied in [6]) in comparison with the vales calculated using relationship (19). As is seen, the experimental and theoretical values are close to within the order of magnitude. A more rapid drop in the experimental η'_Ω values with increasing frequency Ω is explained by some experimental features. The measurements in [6] were performed in such way that the amplitude of the oscillating velocity gradient was proportional to the frequency Ω. As will be demonstrated in the next section, the breakage of chains by a hydrodynamic flow leads to a rapid drop in viscosity with increasing flow velocity gradient; this circumstance apparently affected the results obtained in [6].

4. RHEOLOGICAL PROPERTIES OF FERROCOLLOIDS AT A FINITE VELOCITY GRADIENT

In this section, the consideration is restricted to stationary flows. Within the framework of the proposed model, the effective viscosity is calculated by the following scheme.

1. Substituting (13) into (15) and (16) and calculating the corresponding moments, we arrive at a system of transcendental equations with respect to **a** and **b**. This system can be solved by numerical methods.

2. Substituting the obtained solutions into (13), we calculate the moments of the distribution function in (10) and the dimensionless free energy (4). Substituting this energy into (6) and (7), we determine the nonequilibrium orientation distribution function g_n. The calculations show that (for finite ν) the g_n corresponding to large n is smaller than the equilibrium function g_n^0; for small n, the nonequilibrium function g_n is greater than g_n^0. This is explained by the fact that long chains are more strongly deflected by the flow from the field direction. As a result, the free energy f_{nm} increases and, hence, the g_n value decreases. The breakage of long chains leads to an increase in the number of short ones.

3. Odenbach and Stork [5] obtained an estimate in the order of magnitude for the critical number n_c of particles in the chain, such that the hydrodynamic flow breaks the chains with $n > n_c$. Therefore, the summation in (1), (6), and (10) should be truncated at $n = n_c$. Using the notation adopted here, the estimate obtained in [5] can be expressed as

$$n_c \sim \frac{1}{3}\sqrt{\varepsilon\frac{D}{\nu}}.$$

Figures 2 and 5 show a comparison of the experimental data obtained in [5, 6] for various finite ν values to the results of calculations performed according to the proposed model. As is seen, the theoretical and experimental results show a quite reasonable agreement.

Figure 6 presents the experimental results obtained in [7] for $\nu = 500\ s^{-1}$. The estimates show that, for such a high shear rate, virtually all chainlike aggregates are broken and the average number of $\langle n\rangle$ particles in the chain is almost equal to unity. For the comparison, Fig. 6 shows the results of calculations of the effective viscosity in the aforementioned model bidisperse ferrofluid with all chains broken and the large particles occurring in the individual state. The contribution of small superparamagnetic particles into the magnetoviscosity effect was still ignored. As is seen, experimental and theoretical values agree in the order of magnitude. The fact that the experimental values are higher than the calculated ones is related, first of all, to the simplicity of the bidisperse model. Particles with the magnetic core diameter only slightly exceeding $d_c \sim 12$–14 nm are still too small to form chains. At small shear rates, when the effective viscosity is related predominantly to clusters, such particles virtually do not contribute to the rheological characteristics of the ferrofluid and do not influence the procedure of fitting to the bidisperse medium parameters. Therefore, these particles "fall out" of the bidisperse model. At large shear rates, when the chains are broken, these particles play a more significant role because their concentration is relatively large.

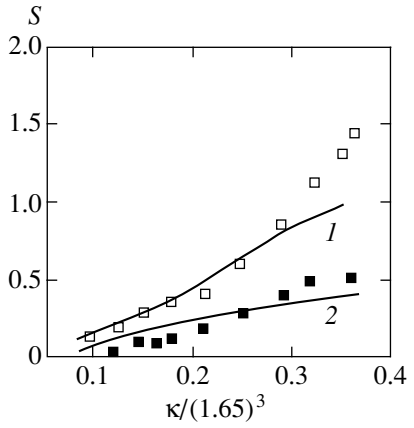

Fig. 5. Plots of the magnetoviscosity parameter S versus dimensionless magnetic field strength $\kappa = m_l H/T$. Symbols represent the experimental data taken from [5] for $\nu = 1.05$ (open squares) and $5.23\ s^{-1}$ (filled squares); curves *1* and *2* show the results of model calculations for $\nu = 1.05$ and $5.23\ s^{-1}$, respectively.

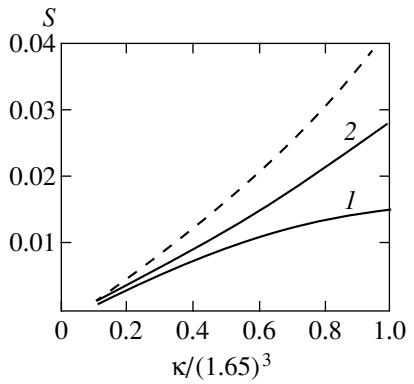

Fig. 6. The plots of effective viscosity of a ferrocolloid at large shear rates, when all chains are broken. Solid curves show the results of calculations (*1*) according to the proposed model and (*2*) using the model of noninteracting individual particles with the size distribution depicted in Fig. 1 [7]; dashed curve show the experimental for $\nu = 500\ s^{-1}$ [7].

This explanation is confirmed by the fact that calculations [7] performed within the framework of the model of individual particles with an allowance of the histogram in Fig. 1 agree with the experimental data [7] better than do the values calculated using the bidisperse model (see Fig. 6). The fact that the experimental data reported in [7] are greater than the results of calculations performed in the same study can be explained as follows. First, the calculations [7] did not took into account the interactions between individual particles. In order to assess these interactions, it is necessary to develop a theory of the rheological properties of polydisperse undiluted ferrocolloids containing interacting

individual particles. Second, the difference between experiment and theory in [7] can be related to the presence of unbreakable aggregates (dimers) linked by van der Waals forces due to the presence of defects in the stabilizing surface layers. The contribution to the effective viscosity due to dimers oriented along the flow velocity gradient is greater than the contribution of a spherical particle with the same volume. The existence of hard dimers in ferrocolloids, as well as the influence of these dimers on the system rheology was discussed, for example, in [5, 14, 16].

5. CONCLUSIONS

The results of our investigations lead to the following conclusions. The polydispersity is a property inherent in modern magnetic fluids. The rheological properties of typical ferrofluids as small flow velocity gradients are determined by the largest particles, although their concentration is usually much lower than that of small particles. The influence of large particles on the macroscopic dynamic properties of ferrofluids is related to the formation of internal heterogeneous structures as a result of the magnetic dipole interaction with each other. Depending on the fluid type, particle concentration, etc., these heterogeneous structures may appear as linear (chainlike) or bulky (droplike) aggregates. The aggregates tend to orient along the magnetic field lines, which leads to a strong magnetoviscosity effect. The value of this effect in the ferrofluids with aggregates exceeds by one–two orders of magnitude the analogous effect in the systems of individual particles. For this reason, the standard theories based on monodisperse models of magnetic fluids, operating with average (small) particles and ignoring the particles aggregation effects, fail to adequately describe the experimental data. As the flow velocity gradient in a ferrocolloid increases, the aggregates break, this leading to a rapid drop in the effective viscosity. Under conditions leading to complete breakage of the chains, the models based on individual particles may provide for a quite acceptable result.

The model assuming the chain cluster formation explained the experimental results obtained in [5–7, 14] for the ferrofluids of a certain type. It is not excluded that the properties of other ferrofluids can be significantly affected by the formation of droplike or other internal heterostructures. We may expect that a decisive role of large, albeit not numerous, particles usually not taken into account can be important in the polar dispersions of some other types, such as magneto- and electrorheological suspensions and analogous systems.

It must be noted that a model neglecting the interaction of chains, despite their large effect on the rheological characteristics of colloids, is certainly a rather crude approximation. Taking into account the hydrodynamic interaction of chains will lead to an additional increase in the viscosity values predicted by the theory and to some renormalization of the bidisperse model characteristics. However, the theory of hydrodynamic properties of the suspensions of hydrodynamically interacting nonspherical particles (in particular, chains) is not yet developed. However, even the simple proposed model (linear with respect to the chain concentration) predicts a strong field-induced increase in the viscosity and provides for the estimates close to the experimental data for both the rheological characteristics of the medium and for the size and concentration of large particles.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research (project nos. 98-01-00031 and 00-02-17731), by the Ministry of Education of the Russian Federation (project no. E00-3.2-210), and by the CRDF Project RUS 005.

REFERENCES

1. A. Yu. Zubarev and A. V. Yushkov, Zh. Éksp. Teor. Fiz. **114**, 892 (1998) [JETP **87**, 484 (1998)].
2. M. A. Martsenyuk, Y. A. Raikher, and M. I. Shliomis, Zh. Éksp. Teor. Fiz. **65**, 834 (1973) [Sov. Phys. JETP **38**, 413 (1974)].
3. M. C. Miguel and J. M. Rubi, Physica A (Amsterdam) **231**, 288 (1996).
4. J. P. Shen and M. Doi, J. Phys. Soc. Jpn. **59**, 111 (1990).
5. S. Odenbach and H. Stork, J. Magn. Magn. Mater. **183**, 188 (1998).
6. S. Odenbach and J. Fletcher, private communication.
7. S. Odenbach and H. Gilly, J. Magn. Magn. Mater. **152**, 123 (1996).
8. S. Kamiyama and A. Satoh, J. Colloid Interface Sci. **127**, 173 (1989).
9. A. Yu. Zubarev and L. Yu. Iskakova, Zh. Éksp. Teor. Fiz. **107**, 1534 (1995) [JETP **80**, 857 (1995)].
10. A. Yu. Zubarev and L. Yu. Iskakova, Phys. Rev. E **61**, 5415 (2000).
11. V. N. Pokrovskii, *Statistical Mechanics of Dilute Suspension* (Nauka, Moscow, 1978).
12. A. O. Ivanov and A. Yu. Zubarev, Physica A (Amsterdam) **251**, 348 (1998).
13. M. Doi and S. F. Edwards, J. Chem. Soc., Faraday Trans. 2 **74**, 560 (1978).
14. O. Ambacher, S. Odenbach, and K. Stierstadt, Z. Phys. B **86**, 29 (1992); S. Odenbach and K. Ray, submitted to Phys. Fluids.
15. G. K. Batchelor and J. T. Green, J. Fluid Mech. **56**, 375 (1972).
16. K. O'Grady, H. K. Stewardson, R. W. Chantrell, *et al.*, IEEE Trans. Magn. **22**, 1134 (1986).

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 93, No. 1, 2001, pp. 89–93.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 1, 2001, pp. 104–108.

FLUIDS

Critical Behavior of Dilute Electrolyte Solutions¶

A. R. Muratov
Institute for Oil and Gas Research, Russian Academy of Sciences, Moscow, 117971 Russia
e-mail: muratov@ogri.ru
Received November 24, 2000

Abstract—A theory of the critical behavior of a dilute ionic solution is constructed. An expression for the susceptibility in a wide temperature range is obtained. It is shown that ionic solutions belong to the universality class of the Ising model. The Ginzburg parameter of the ionic solutions decreases with the increase of the solvent concentration. In the general case, the critical exponent of susceptibility nonmonotonically depends on the temperature in the crossover region from the Ising-like to the mean-field behavior. In the vicinity of the transition point, the Debye–Hückel screening radius is proportional to the correlation length. As $T \longrightarrow T_c$, the screening radius tends to infinity and the screening disappears. The voltage between the two phases of the ionic solution is proportional to the order parameter and changes as $|T/T_c - 1|^\beta$ in the vicinity of the phase transition point.

1. INTRODUCTION

Critical phenomena in liquids and liquid mixtures are an extensively studied region of condensed matter physics [1, 2]. In general, the behavior of physical characteristics near the critical solution point can be described sufficiently well as a crossover between the Ising-like asymptotic behavior and the mean-field behavior [3, 4, 5]. The Ising-like behavior occurs in a narrow vicinity of the critical point for

$$|T/T_c - 1| \ll G,$$

where T is the temperature, T_c is the critical temperature, and G is the so-called Ginzburg parameter. The mean-field behavior occurs for

$$|T/T_c - 1| \gg G.$$

For liquids, the parameter G is of the order 0.01–0.1.

The critical behavior of ionic solutions was discussed in a number of works (see [6, 7, 8, 9, 10]). Experimental studies show that the critical behavior of electrolyte solutions significantly differs from that of ordinary solutions. This difference occurs for both nonaqueous [11, 12, 13] and aqueous solutions [14]. The main feature of the electrolyte solutions is their mean-field behavior in the region

$$|T/T_c - 1| > t_x,$$

where t_x varies from 1 to 10^{-4} for different solutions [15]. Numerous attempts to understand the ionic criticality have been made recently (see, e.g., [15, 16, 17]). In a recent experimental work [14], it was stated that the crossover behavior is not monotonic and the effective susceptibility exponent γ has its maximum value in the crossover region exceeding the asymptotic value $\gamma \sim 1.24$.

2. THEORY

We consider the electrolyte solution near the critical solution point of the solvent. We let $\varphi(\mathbf{r})$ be a field proportional to the order parameter (the solution density for the critical point or the concentration for the critical solution point), $\phi(\mathbf{r})$ be the electric potential field, and $\rho_1(\mathbf{r})$ and $\rho_2(\mathbf{r})$ be the respective volume densities of the ion numbers with positive and negative charges. For simplicity, we assume that these charges are equal to e and $-e$ respectively.

Near the transition point, the ionic solution consists of two subsystems: the "fast" subsystem that depends on the ion densities ρ_1 and ρ_2 and the "slow" subsystem that depends on the order parameter φ and the electric potential ϕ. As in the standard Debye–Hückel (Hartree) approach, ions can be considered in the average long-wavelength field $\phi(\mathbf{r})$ and, in our case, $\varphi(\mathbf{r})$. We first evaluate the fluctuation corrections to the Hamiltonian of the "slow" part of the system that are caused by the "fast" subsystem and then consider only the part of the Hamiltonian containing the density field φ and the electric potential field ϕ.

The main interaction terms of the fields ϕ, φ and ρ_1, ρ_2 can be written as

$$H_{\text{int}} = \int d\mathbf{r}((e\rho_1(\mathbf{r}) - e\rho_2(\mathbf{r}))\phi(\mathbf{r}) + (\gamma_1\rho_1(\mathbf{r}) - \gamma_2\rho_2(\mathbf{r}))\varphi(\mathbf{r})). \quad (1)$$

The first interaction term in Eq. (1) is the ordinary electrostatic interaction, the second term is some phenomenological interaction between the order parameter

¶This article was submitted by the author in English.

and the ion densities, and γ_1 and γ_2 are the interaction constants.

It is easy to see from Eq. (1) that the values

$$\begin{aligned}\Delta\mu_1(\mathbf{r}) &= e\phi(\mathbf{r}) + \gamma_1\varphi(\mathbf{r}),\\ \Delta\mu_2(\mathbf{r}) &= -e\phi(\mathbf{r}) - \gamma_2\varphi(\mathbf{r})\end{aligned} \tag{2}$$

can be considered as local additional corrections to the chemical potentials of the respective positive (μ_1) and negative (μ_2) ions. The Gauss distribution of the thermodynamic fluctuations of the chemical potential is given by (see [18])

$$w \propto \exp\left(-\frac{\Delta\mu\Delta\rho}{T}\right) = \exp\left(-\frac{\bar{\rho}}{2T^2}(\Delta\mu)^2\right), \tag{3}$$

where T is the absolute temperature in energy units and $\bar{\rho} = N/V$ is the average volume density of particles.

Inserting expressions (2) in Eq. (3), we obtain the fluctuation correction to the Hamiltonian of the "slow" part of the system,

$$\begin{aligned}\Delta H &= \int d\mathbf{r}\\ &\times\left(\frac{\bar{\rho}_1}{2T}(e\phi(\mathbf{r}) + \gamma_1\varphi(\mathbf{r}))^2 + \frac{\bar{\rho}_2}{2T}(e\phi(\mathbf{r}) + \gamma_2\varphi(\mathbf{r}))^2\right).\end{aligned} \tag{4}$$

For the ionic solution, we must add expression (4) and the expression for the electrostatic energy

$$\int d\mathbf{r}(\mathbf{E}\cdot\mathbf{D}/8\pi)$$

(where $\mathbf{D} = \varepsilon\mathbf{E}$) to the standard Landau Hamiltonian of the system near the critical point. Taking into account that

$$\bar{\rho}_1 = \bar{\rho}_2 = \rho,$$

we obtain the effective Hamiltonian

$$\begin{aligned}H_{\text{eff}} &= \int d\mathbf{r}\left(\frac{1}{2}\left(a + \frac{\rho}{T}(\gamma_1^2 + \gamma_2^2)\right)\varphi^2 + \frac{1}{2}\alpha(\nabla\varphi)^2\right.\\ &\left. + \frac{\rho}{T}e^2\phi^2 + \frac{\varepsilon}{8\pi}(\nabla\phi)^2 + \frac{\rho}{T}e(\gamma_1 + \gamma_2)\phi\varphi + \lambda\varphi^4\right).\end{aligned} \tag{5}$$

There are three different fluctuation corrections in Hamiltonian (5). The correction proportional to ϕ^2 describes the Debye-Hückel screening. The correction proportional to φ^2 shifts the bare temperature of the phase transition. The correction proportional to $\phi\varphi$ couples the electric potential field to the order parameter. Far from the phase transition point, the parameter a is not small and the last two corrections to the Hamiltonian are not relevant. In the near-critical region, the parameter

$$a \propto |T/T_c - 1|$$

is small and we must take these corrections and the φ^4 term in the Hamiltonian into account.

In what follows, the angular brackets denote averaging over effective Hamiltonian (5). The quadratic part of effective Hamiltonian (5) determines the bare values of the correlation functions. The inverse correlation function is equal to

$$\begin{aligned}&\begin{pmatrix}\langle\varphi(\mathbf{k})\varphi(-\mathbf{k})\rangle & \langle\varphi(\mathbf{k})\phi(-\mathbf{k})\rangle\\ \langle\phi(\mathbf{k})\varphi(-\mathbf{k})\rangle & \langle\phi(\mathbf{k})\phi(-\mathbf{k})\rangle\end{pmatrix}^{-1} = \frac{1}{T}\\ &\times\begin{pmatrix}a + \alpha k^2 + \rho(\gamma_1^2 + \gamma_2^2)/T & e\rho(\gamma_1 + \gamma_2)/T\\ e\rho(\gamma_1 + \gamma_2)/T & \varepsilon(k^2 + \kappa^2)/4\pi\end{pmatrix},\end{aligned} \tag{6}$$

where $\kappa = \sqrt{8\pi\rho e^2/T\varepsilon}$ is the inverse Debye–Hückel radius. Calculating the inverse matrix, we obtain

$$\begin{aligned}\langle\varphi(\mathbf{k})\varphi(-\mathbf{k})\rangle &= \frac{T}{\tilde{a} + \alpha k^2(1 + b\kappa^2/(k^2 + \kappa^2))},\\ \langle\varphi(\mathbf{k})\phi(-\mathbf{k})\rangle &= -\frac{T(\gamma_1 + \gamma_2)/2e}{\tilde{a} + (k/\kappa)^2(\tilde{a} + b\alpha\kappa^2 + \alpha(k^2 + \kappa^2))},\\ \langle\phi(\mathbf{k})\phi(-\mathbf{k})\rangle &= \frac{4\pi T/\varepsilon}{k^2 + \kappa^2(\tilde{a} + \alpha k^2)/(\tilde{a} + b\alpha\kappa^2 + \alpha k^2)},\end{aligned} \tag{7}$$

$$\tilde{a} = a + \frac{\rho}{2T}(\gamma_1 - \gamma_2)^2,$$

$$b = \frac{\varepsilon}{16\pi\alpha e^2}(\gamma_1 + \gamma_2)^2 = \frac{\rho}{2T\alpha\kappa^2}(\gamma_1 + \gamma_2)^2.$$

The correlation function of the electric potential field in Eqs. (7) describes the Coulomb screening near the critical point of the ionic solution. It is easy to see that the screening has the ordinary form far from the critical point, where $a \gg \rho\gamma^2/T$. In the vicinity of the phase transition point $\tilde{a} = 0$ in the mean-field approximation, the screening radius behaves as

$$r_s = \frac{1}{\kappa}\sqrt{\frac{\tilde{a} + b\alpha\kappa^2}{\tilde{a}}}.$$

As $\tilde{a} \longrightarrow 0$, the screening radius tends to infinity and the screening disappears.

The density correlation function in Eqs. (7) has the asymptotic behavior

$$\begin{aligned}\langle\varphi(k)\varphi(-k)\rangle &\approx \frac{T}{\tilde{a} + \alpha(1 + b)k^2}, \quad k \ll \kappa,\\ \langle\varphi(k)\varphi(-k)\rangle &\approx \frac{T}{\tilde{a} + \frac{\rho}{2T}(\gamma_1 + \gamma_2)^2 + \alpha k^2}, \quad k \gg \kappa.\end{aligned} \tag{8}$$

The interaction between the mass density field (order parameter) and the charge density field redefines

bare correlation function (4) in the long-wavelength region as:

$$a \longrightarrow \tilde{a}, \quad \alpha \longrightarrow \alpha(1+b). \tag{9}$$

The asymptotic forms of correlation function (8) in the long and short wavelength regions are analogous to the ordinary form of the correlation function near the critical point. This means that the behavior of physical characteristics in the corresponding limits is the same as in the nonionic case. The bare critical point

$$a = -\rho(\gamma_1^2 + \gamma_2^2)/T$$

and the real singularity point

$$a = -\rho(\gamma_1 - \gamma_2)^2/2T$$

are now different, both of them are less than the pure solvent transition point $a = 0$.

In the mean-field approximation, the correlation length in vicinity of the transition point is given by

$$r_c = \sqrt{\frac{\alpha(1+b)}{\tilde{a}}}.$$

In the vicinity of the transition point, the screening length is proportional to the correlation length and their ratio is

$$\frac{r_s}{r_c} = \sqrt{\frac{\tilde{a}/\alpha\kappa^2 + b}{1+b}}.$$

The mean-field approximation does not correctly describe the behavior of the system near the transition point. It can be used in the region far from the transition point, where $\tilde{a} \gg G$. In close vicinity of the transition point, the fluctuation effects must be taken into account.

3. RENORMALIZATION

We now calculate the expression for the susceptibility, which is an experimentally measured value [13, 14]. We must apply the standard renormalization procedure to effective Hamiltonian (5). The leading corrections to the coefficients $\tilde{a}$ and λ due to fluctuation effects can be represented by the diagrams in Fig. 1. The renormalization equations for these coefficients are given by

$$\frac{d\tilde{a}}{d\Lambda} = \frac{1}{3}K\frac{\tilde{a}(\Lambda)\lambda(\Lambda)\Lambda^{d-5}}{(1 + b\kappa^2/(\Lambda^2+\kappa^2))^2},$$
$$\frac{d\lambda}{d\Lambda} = K\frac{\lambda^2(\Lambda)\Lambda^{d-5}}{(1 + b\kappa^2/(\Lambda^2+\kappa^2))^2}, \tag{10}$$

where Λ is the renormalization equation parameter ($[\Lambda] = k$) and

$$K = \frac{18TS_d}{(2\pi)^d\alpha^2},$$

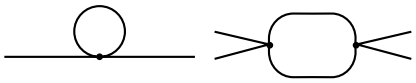

Fig. 1. One-loop corrections to the effective Hamiltonian parameters a and λ. The solid line corresponds to the field φ. The circle in the vertex is equal to λ.

with d being the space dimension and S_d the unit sphere surface. Equations (10) are written in the one-loop approximation, which corresponds to the first approximation in the parameter $\epsilon = 4 - d$. For simplicity, we do not use the ϵ-expansion but consider the one-loop renormalization equations directly for $d = 3$.

It is easy to see from Eqs. (10) that

$$\frac{d\tilde{a}}{d\lambda} = \frac{\tilde{a}}{3\lambda}, \quad \frac{\tilde{a}^3(\Lambda)}{\lambda(\Lambda)} = \text{const}. \tag{11}$$

The solution of the equation for the interaction vertex is given by

$$\frac{1}{\lambda(\Lambda)} - \frac{1}{\tilde{\lambda}} = K\int_0^{1/\Lambda} dx\left(\frac{1+\kappa^2x^2}{1+(1+b)\kappa^2x^2}\right)^2. \tag{12}$$

Calculating the integral in (12) and taking Eq. (11) into account, we find

$$\lambda(\Lambda) = \frac{\lambda(\infty)}{1+F(\Lambda)}, \quad \tilde{a}(\Lambda) = \frac{\tilde{a}(\infty)}{(1+F(\Lambda))^{1/3}}, \tag{13}$$

where

$$F(\Lambda) = \frac{K\lambda(\infty)}{2\kappa(1+b)^{5/2}}$$
$$\times\left(t\left(2 + \frac{b^2}{1+t^2}\right) + b(b+4)\arctan t\right), \tag{14}$$
$$t = \frac{\kappa}{\Lambda}\sqrt{1+b}.$$

To obtain the expressions for physical quantities valid in a wide region near the transition point, it is necessary to choose a matching point Λ^* for solutions (13). The matching point can be chosen in the same way as in [3] (see also [19]),

$$\Lambda^* = \sqrt{\frac{\tilde{a}(\Lambda^*)}{\alpha}}. \tag{15}$$

Inserting this in Eq. (13), we obtain the equation for the susceptibility,

$$\tilde{a}(\Lambda^*)\left(1 + \frac{K\lambda(\infty)}{2\kappa(1+b)^{5/2}}\left(t^*\left(2 + \frac{b^2}{1+t^{*2}}\right)\right.\right.$$
$$\left.\left. + b(b+4)\arctan t^*\right)\right)^{1/3} = \tilde{a}(\infty), \tag{16}$$

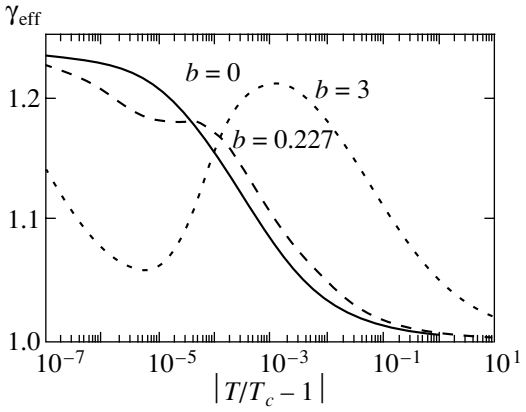

Fig. 2. The dependence of the effective susceptibility critical exponent γ_{eff} on the value of b for the model system. The curve where $b = 0$ corresponds to the non-ionic solution.

where

$$t^* = \sqrt{\frac{(1+b)\alpha\kappa^2}{\tilde{a}(\Lambda^*)}}.$$

Equation (16) determines the susceptibility in a wide temperature region in the one-loop approximation. The critical exponent γ is equal to 6/5 in this approximation. It is necessary to modify this expression in order to apply it to experimental data. The simplest way is to replace the exponent 1/3 in (16) with $2 - 2/\gamma$, where $\gamma \approx 1.24$.

4. DISCUSSION

The expressions obtained can be used for electrolyte solutions (strong or weak) if the ion density ρ is small. As $\rho \longrightarrow 0$, all the expressions transform to the ordinary form valid for nonionic liquids. The pure solvent limit can also be obtained as $b \longrightarrow 0$, i.e., if the density is weakly coupled to the charge density.

The Debye–Hückel screening radius in the vicinity of the transition point is proportional to the correlation length. As $T \longrightarrow T_c$, the screening radius tends to infinity and the screening disappears.

The asymptotic forms of the density correlation function for $k \gg \kappa$ and $k \ll \kappa$ are the same as for the nonionic liquid, and therefore, the corresponding (mean-field or Ising-like) asymptotic behavior is the same. The value of $F(\Lambda)\sqrt{\tilde{a}^*}$ in the ionic solution is similar to the Ginzburg parameter G for the ordinary liquids. It decreases as the salt concentration increases, and the Ising-like region of ionic solutions therefore decreases as the concentration increases, which agrees with the experimental data.

Equation (16) determines behavior of the susceptibility in the crossover region. Dependence of the effective exponent γ on the temperature for the model system is presented in Fig. 2. For small values of b, this dependence has the usual shape, but it becomes nonmonotonic for $b \sim 1$. For the real experimental situation in [14], this specific behavior occurs in a close vicinity of the transition point and it is difficult to reveal it. The authors of [14] state that they see this peculiarity.

The shape of the crossover curve depends on the dimensionless parameter b (Eqs. (7)), i.e., on $\gamma_1 + \gamma_2$. This quantity can be estimated from a simple electric measurement. In the mean-field approximation, Eq. (5) gives

$$\langle\phi\rangle = -\frac{\gamma_1 + \gamma_2}{2e}\langle\varphi\rangle = -\frac{\gamma_1 + \gamma_2}{4e}\sqrt{\frac{-\tilde{a}}{\lambda}} \qquad (17)$$

for the two-phase state. Therefore, the voltage between the two phases of the ionic solution is determined by $\gamma_1 + \gamma_2$. Near the transition point, this voltage is proportional to the order parameter and behaves as $(T/T_c - 1)^{\beta}$.

A detailed comparison of the obtained expressions to the experimental data is not a simple task. The data presented in [11–13] have significant error bars. The ionic solutions in [14, 20] were ternary and their behavior was studied near the critical solution points. These solutions had a high concentration (10%), and there were additional problems due to clustering in the vicinity of double critical and tricritical points. The data for a fixed concentration can be easily fitted in accordance with Eq. (16). Unfortunately, this fit is not informative and further experimental studies are necessary.

ACKNOWLEDGMENTS

The author is deeply grateful to V. Lebedev and E. Gorodetskii for valuable discussions.

REFERENCES

1. C. Domb and M. S. Green, *Phase Transitions and Critical Phenomena* (Academic, New York, 1974), Vol. 3.
2. A. Z. Patashinskii and V. L. Pokrovskii, *Fluctuation Theory of Phase Transitions* (Nauka, Moscow, 1982, 2nd ed.; Pergamon, Oxford, 1979).
3. J. Rudnick and D. R. Nelson, Phys. Rev. B **13**, 2208 (1976).
4. J. F. Nicoll and P. C. Albright, Phys. Rev. B **31**, 4576 (1985).
5. S. Tang, J. V. Sengers, and Z. Y. Chen, Physica A (Amsterdam) **179**, 344 (1991).
6. M. E. Fisher, J. Stat. Phys. **75**, 1 (1994).
7. G. Stell, J. Stat. Phys. **78**, 197 (1995).
8. J. M. H. Levelt Sengers and J. A. Given, J. Mol. Phys. **80**, 899 (1993).
9. K. S. Pitzer, Acc. Chem. Res. **22**, 333 (1990).
10. H. Weingartner, M. Kleemeier, S. Wiegand, and W. Schroer, J. Stat. Phys. **78**, 169 (1995).

11. T. Narayan and K. S. Pitzer, Phys. Rev. Lett. **73**, 3002 (1994).

12. T. Narayan and K. S. Pitzer, J. Phys. Chem. **98**, 9170 (1994).

13. T. Narayan and K. S. Pitzer, J. Chem. Phys. **102**, 8118 (1995).

14. J. Jacob, A. Kumar, M. Anisimov, *et al.*, Phys. Rev. E **58**, 2188 (1998).

15. B. P. Lee and M. E. Fisher, Phys. Rev. Lett. **76**, 2906 (1996).

16. M. E. Fisher and B. P. Lee, Phys. Rev. Lett. **77**, 3561 (1996).

17. B. P. Lee and M. E. Fisher, Europhys. Lett. **39**, 611 (1997).

18. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 5: *Statistical Physics* (Nauka, Moscow, 1976; Pergamon, Oxford, 1980).

19. M. Yu. Belyakov, S. B. Kiselev, and A. R. Muratov, Zh. Éksp. Teor. Fiz. **104**, 2785 (1993) [JETP **77**, 279 (1993)].

20. M. A. Anisimov, J. Jacob, A. Kumar, *et al.*, Phys. Rev. Lett. **85**, 2336 (2000).

Journal of Experimental and Theoretical Physics, Vol. 93, No. 1, 2001, pp. 94–102.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 120, No. 1, 2001, pp. 109–118.

FLUIDS

Linear Electromechanical Effect in Free-Standing Ferroelectric Liquid-Crystalline Films

S. V. Yablonskii[a, b, *], A. S. Mikhailov[a], K. Nakano[b], M. Ozaki[b], and K. Yoshino[b]

[a]*Shubnikov Institute of Crystallography, Russian Academy of Sciences, Moscow, 117333 Russia*

[b]*Department of Electronic Engineering, Faculty of Engineering, Osaka University, 2-1 Ymada-Oka, Suita, Osaka 565, Japan*

**e-mail: palto@online.ru*

Received January 25, 2001

Abstract—Mechanical oscillations of free-standing films based on ferroelectric liquid crystals were studied by optical methods. The intrinsic oscillation modes in the samples were excited by applying an alternating electric field parallel to the film surface. The surface viscosity of the films determined using the electromechanical effect was $\eta_s = 8.8 \times 10^{-3}$ g/s under normal pressure and $\eta_s = 1.5 \times 10^{-3}$ g/s in vacuum. The surface tension measured in a special experiment was $\sigma = 35.3$ din/cm. It was established that the spectrum of mechanical oscillations in the system studied is affected by the vapors of volatile organic solvents such as kerosene, toluene, and ethyl alcohol. The linear electromechanical effect in the free-standing films was used to observe inversion of the sign of spontaneous polarization in a ferroelectric liquid crystal. © *2001 MAIK "Nauka/Interperiodica".*

Free-standing films can be obtained using both lyotropic [1] and thermotropic smectic liquid crystals [1, 2]. The former type includes, for example, the films of a usual soap. Both classes of film-forming liquid-crystalline (LC) materials represent layered systems featuring a liquid order inside layers and a long-range crystalline order in the direction of the smectic normal. The principal possibility that thin layers may exist in the form of free-standing films is related to the presence of internal forces holding the centers of mass of the molecules within a single layer. When the liquid crystal transforms into a nematic or isotropic phase, the layered structure disappears and the free-standing films lose stability [3].

Necessary factors always accompanying a free-standing film are a hard frame and a transition region between a homogeneous part of the film and the hard frame. The transition region, called meniscus, possesses a complicated structure. The mass of the meniscus is usually considerably (one or even two orders of magnitude) greater than the mass of the film as such. The meniscus plays an important role in the process of attaining thermodynamic equilibrium in the system. Involved in the mass exchange with the homogeneous part of the film, the meniscus maintains a constant chemical potential of the system [4].

From the standpoint of mechanics, a free-standing film is usually considered as an elastic membrane, the motion of which is described by the wave equation [5, 6]. The intrinsic frequencies of such a membrane with a given geometry depend (provided that the effects of meniscus can be ignored) only on the isotropic surface tension σ and a homogeneous two-dimensional density $\rho_s = h\rho_{lc}$, where ρ_{lc} is the liquid crystal density and h is the film thickness (the latter value may vary from hundreds of molecular lengths to a single bimolecular layer) [7]. In this approximation, the surface tension of both liquidlike smectic A (sm*A*) and more ordered smectic B (sm*B*) free-standing films can be calculated using the spectrum of mechanical oscillations [8].

In this simplified model, the properties of a free-standing film are independent of its thickness. At the same time, there are indications that the film structure may play an important role. For example, Kraus *et al.* [9] demonstrated the effect of a three-dimensional elasticity on the spectrum of mechanical oscillations of a free-standing film. Cladis *et al.* [10] established that topological defects appearing in a free-standing film correspond to the symmetry of the LC phase employed. Brazovskaya *et al.* [11] observed a nonlinear effect in a free-standing film oscillating with large amplitude, which was manifested by the appearance of a meniscus instability threshold. Boudaoud *et al.* [12] showed that the thickness of an oscillating thick soap film does not remain homogeneous in the course of its motion and that the spectrum of mechanical oscillations of the film differs significantly from the spectrum of a linearly elastic membrane.

Thus, there is a sufficiently large number of experimental facts indicative of a more complicated (as compared to the Rayleigh assumptions) organization of free-standing LC films.

The purpose of this work was to study the features of mechanical oscillations of free-standing films based on ferroelectric liquid crystals, depending on various external factors such as the frequency and amplitude of

the applied electric field, gas pressure, and temperature. Since all ferroelectric materials possess piezoelectric properties as well, a natural method of exciting mechanical oscillations in the free-standing films [13–16] is offered by the inverse piezoelectric effect or, using the term suggested by Jakli *et al.* [17, 18] (having originally observed this phenomenon in a standard geometry), the linear electromechanical effect.

1. THE EQUATION OF MOTION OF A FREE-STANDING FILM WITH AN ALLOWANCE FOR THE ENERGY DISSIPATION

Consider a free-standing film of a ferroelectric liquid crystal in a rectangular frame (Figs. 1a and 1b). Let us analyze how an electric field **E** interacting with a spontaneous polarization **p** (per unit area) would disturb the film surface. For $b \gg a$, we may solve the problem in a one-dimensional approximation. In the first step, we will demonstrate that the competition of only electric and elastic forces is insufficient to violate equilibrium of a flat elastic film in the small-perturbation approximation,

The free energy density f of the free-standing film under consideration can be presented as a sum of the electrostatic f_{el} and elastic f_s contributions:

$$f = f_{el} + f_s. \tag{1}$$

A free energy increment per unit length is

$$\Delta F = \Delta F_{el} + \Delta F_s, \tag{2}$$

where

$$\Delta F_{el} = f_{el} dx = -Ep\cos(\pi - \varphi)dx,$$

$\Delta F_s = f_s dx = 2\sigma dl$, and dl is the film area increment. The integration in Eq. (2) can be performed with an allowance for smallness of the angle φ describing deviation of the normal from the vertical direction, which implies that

$$\tan\varphi \approx \sin\varphi \approx \varphi \approx \frac{dz}{dx},$$

$$\cos\varphi \approx 1 - \frac{\varphi^2}{2} = 1 - \frac{1}{2}\left(\frac{dz}{dx}\right)^2$$

and

$$dl = \sqrt{1 + \left(\frac{dz}{dx}\right)^2}\,dx,$$

where $z(x)$ is the film displacement. The free energy per unit length can be written as

$$F = \int_{-a/2}^{a/2} (f_{el} + f_s)dx \tag{3}$$

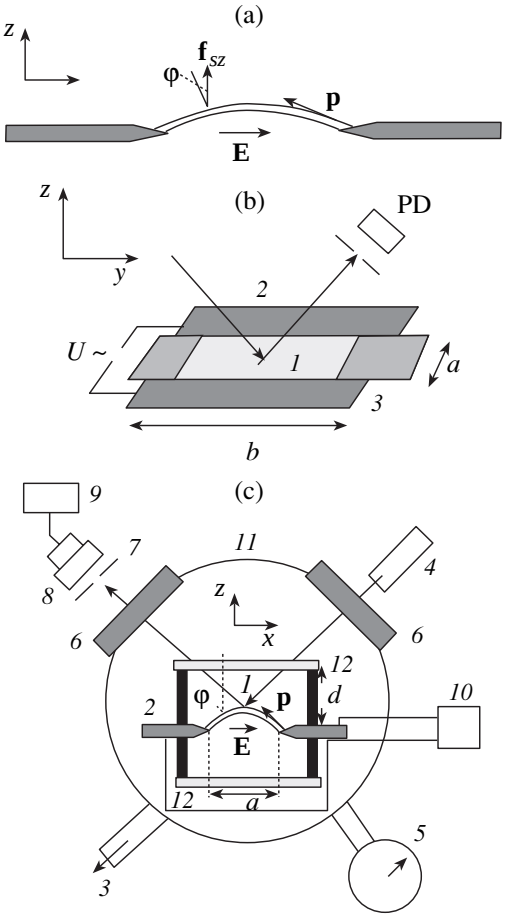

Fig. 1. Schematic diagrams of the (a) free-standing film sample configuration, (b) framework geometry, and (c) experimental setup. (a) Ferroelectric LC film deformed by external electric field **E** parallel to smectic layers in the undisturbed sample, **p** is the two-dimensional spontaneous polarization density, $\mathbf{f}_{sz}$ is the density of external forces, and φ is the angle between normals to the disturbed and undisturbed films. (b) Framework geometry showing photodiode (PD) viewing LC film *1* framed between fixed electrodes *2* and movable thin (100 μm) mylar barriers *3*. (c) Experimental setup for studying the effect of atmospheric pressure upon the free-standing ferroelectric LC film: (*1*) sample film; (*2*) aluminum electrodes deposited onto glass; (*3*) vacuum pump outlet; (*4*) He–Ne laser; (*5*) pressure gauge; (*6*) fused quartz windows; (*7*) slit diaphragm; (*8*) silicon photodiode; (*9*) Fourier spectrometer or lock-in detector; (*10*) sound generator with amplifier; (*11*) vacuum chamber; (*12*) two glass plates at a variable distance d from the film surface.

or

$$F = \int_{-a/2}^{a/2} \left[(pE + 2\sigma) + \frac{1}{2}(2\sigma - pE)\left(\frac{dz}{dx}\right)^2\right]dx. \tag{4}$$

Minimization of the free energy functional (4) leads to the following boundary problem:

$$\frac{d^2z}{dx^2} = 0, \tag{5}$$

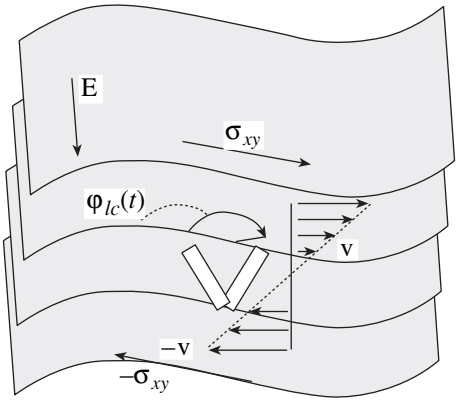

Fig. 2. A mechanism responsible for the development of a viscous stress σ_{xy} in a free-standing ferroelectric film: $\varphi_{lc}(t)$ is the variable azimuthal angle changing under the action of the back-flow; **v** is the liquid crystal velocity field inhomogeneous along the *z* axis.

$$z(a/2) = z(-a/2) = 0, \tag{6}$$

which had a single trivial solution $z \equiv 0$. It is evident that the same result will be obtained for any orientation of **p** with respect to vector **E**.

A more complicated mechanism explaining the appearance of a disturbing moment of forces acting upon the liquid crystal was proposed by Jakli and Eber [19]. According to this, the azimuthal motion of the liquid crystal director is accompanied by the so-called back-flow, which induces a viscous stress acting upon the free surface of the film as depicted in Fig. 2. This mechanism assumes a velocity gradient along the normal to the film surface, which implies the presence of an internal structure in the free-standing film.

Below we will assume for simplicity that a homogeneous density of forces acting upon a unit surface area of a cylindrical film varies according to a harmonic law $f_s = f_{sz}e^{i\omega t}$, where f_{sz} = const. The shape of f_{sz} influences neither the resonance frequency nor the resonance peak width. In the linear case, a mechanical spectrum of the film oscillations depends only on the film parameters and geometry, while f_{sz} influences only the intensity of spectral bands.

Assuming the ρ_s and σ values to be homogeneous and isotropic and taking into account the terms responsible for the internal energy dissipation and aerodynamic drag, a nonlinear one-dimensional equation of motion of a free-standing film can be written as follows:

$$\begin{aligned}(2\sigma - pE(t))\frac{d^2z}{dx^2} + f_{sz}(x,y)e^{i\omega t} \\ = C\rho_{air}\frac{1}{2}\left(\frac{dz}{dt}\right)^2 + \frac{\eta_s}{a^2}\frac{dz}{dt} + \rho_s\frac{d^2z}{dt^2},\end{aligned} \tag{7}$$

where $\rho_{air} = 1.225 \times 10^{-3}$ g/cm^3 is the air density, C is the aerodynamic coefficient ($C \le 1$), $\eta_s \approx \eta_b h$ is the surface viscosity of the film, η_b is the volume viscosity of the liquid crystal, $h = Nl$ is the film thickness, N is the number of smectic layers, l is the thickness of each smectic layer, $p = P_sNl$ is the spontaneous surface polarization, P_s is the spontaneous polarization of the liquid crystal, $\nu = \omega/2\pi$ is the frequency of the alternating electric field E, and a is the distance between electrodes.

In order to simplify the nonlinear equation (7), let us obtain several estimates. For the typical ferroelectric LC film parameters and experimental conditions ($P_s = 10^{-7}$ C/cm^2, $N \sim 100$, $l = 40$ nm, $E < 400$ V/cm, $\sigma \sim 30$ din/cm), the film obeys a strict inequality $\sigma \gg p|E|$. This condition simplifies the coefficient at the second derivative in Eq. (7). The role of the aerodynamic drag can be estimated by comparing this factor to the inertial force. For a film performing harmonic oscillations $z \sim A_0\sin\omega t$, the aerodynamic drag has to be taken into account if the oscillation amplitude is sufficiently large so that $\rho_{air}z_t^2/2 \sim \rho_s z_{tt}$ or $\rho_{air}\omega^2A_0^2/2 \sim \rho_{lc}h\omega^2A_0$. This condition holds when $A_0 \sim 2\rho_{lc}h/\rho_{air}$. Substituting $N = 70$, $l = 4$ nm, and $\rho_{lc}/\rho_{air} \sim 10^3$, we obtain an estimate of $A_0 \sim 500$ μm that is considerably greater than the experimentally measured oscillation amplitudes (1–5 μm). Note that A_0 would additionally increase upon taking into account the virtual (associated) air mass. As will be shown below, an additional reason for ignoring the contribution of air is that the overall losses are determined predominantly by the internal losses in the film, rather than by the friction of entrained air layer, since $\eta_s/a > \eta_{air}$, where $\eta_{air} \approx 2 \times 10^{-4}$ (g cm)/s is the air viscosity.

Thus, neglecting the square term in Eq. (7), we obtain a linear equation of motion with an allowance for the dissipation of mechanical energy:

$$2\sigma\frac{d^2z}{dx^2} + f_{sz}(x,y)e^{i\omega t} = \frac{\eta_s}{a^2}\frac{dz}{dt} + \rho_s\frac{d^2z}{dt^2} \tag{8}$$

which has to be solved with the boundary conditions (6). An exact solution to Eq. (8) is given by the expression

$$z(x,t) = \frac{f_{sz}}{i\omega\eta_s/a^2 - \rho_s\omega^2}\left[1 - \frac{\cosh(q_0x)}{\cosh(q_0a/2)}\right]e^{i\omega t}, \tag{9}$$

where q_0 is the wavevector of the decaying elastic wave:

$$q_0 = \sqrt{\frac{i\omega\eta_s a^{-2} - \rho_s\omega^2}{2\sigma}}. \tag{10}$$

Expression (9) will be used for estimating the surface viscosity and the average two-dimensional density of the films studied.

2. EXPERIMENTAL

The experiments were performed with ferroelectric liquid crystals of two types. The first was a commercial CS-1029 (Chisso) blend featuring the following phase sequence:

$$\text{cryst}(-18^\circ\text{C})\text{sm}C^*(73^\circ\text{C})\text{sm}A(85^\circ\text{C})N^*(95^\circ\text{C})\text{isotr}.$$

Samples of the second type were made of an individual compound, 5-(2-fluoroctyloxy)-2-(4-octylphenyl)pyrimidine (8PPy06, Chisso), changing the sign of spontaneous polarization P_s in the smectic C^* phase:

$$\text{cryst}(46.2^\circ\text{C})[\text{sm}C^*(25^\circ\text{C})\text{sm}C^*(45^\circ\text{C})]\text{sm}A(80^\circ\text{C})\text{isotr}.$$

The free-standing films of both types were prepared using a standard procedure described in [1]. Each film was supported by a rectangular frame composed of two fixed metal electrodes and two mobile poly(ethylene terephthalate) (mylar) barriers (Fig. 1b). The interelectrode distance was fixed at a = 2 or 3 mm, while the spacing of mylar barriers could be varied from 0 to 11 mm. Some experiments were performed with a glass frame of fixed geometry (rectangular slit with an area of 2×10 mm^2 and a thickness of 1 mm). The frame with a film was mounted on a temperature-controlled table, and the entire system was placed into a vacuum chamber (Fig. 1c). The film temperature was monitored with the aid of a calibrated copper-constantan thermocouple. The vacuum in the chamber was provided by a mechanical roughing pump and measured with piezoelectric and mercury pressure gauges. The sample film thickness was calculated using the optical reflection spectrum as described in [1]. The reflection spectra were obtained with the aid of a multichannel spectrum analyzer.

A sinusoidal voltage with variable amplitude (1–120 V) applied to a sample induced periodic oscillations of the film resulting in a pattern of deformation periodic in space and time. The deformation of a given sample region was determined by measuring deviations of a laser beam reflected from this area. The angular deviations of the beam were measured with the aid of a position-sensitive detector comprising a slit diaphragm and silicon photodiode (Fig. 1c). The beam of a He–Ne laser radiation polarized in the direction of the field **E** was incident at an angle of 45° relative to the normal to the film surface. The area of laser-probed region on the film surface did not exceed 0.5×0.5 mm^2. The photodiode response current was analyzed by a Fourier spectrometer and a lock-in amplifier tuned to the first and second harmonics of the sinusoidal electric field applied to the sample. The first-harmonic response corresponds to the amplitude of the vertical displacement (due to a linear deformation in the inverse piezoelectric effect). The second-harmonic optical response component is related primarily to variations in the Fresnel reflection coefficient, which is an even function of the field **E**. In our experiments, the second-harmonic signal amplitude was considerably smaller than that of the first-harmonic response (Fig. 3).

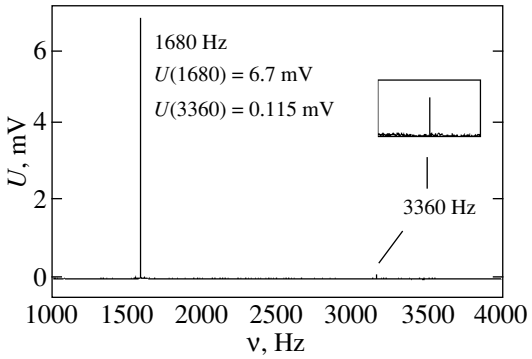

Fig. 3. A typical Fourier spectrum of the laser beam reflected from the surface of an oscillating free-standing film. The inset shows the second-harmonic response signal, the intensity of which did not exceed 2% of the first-harmonic signal amplitude.

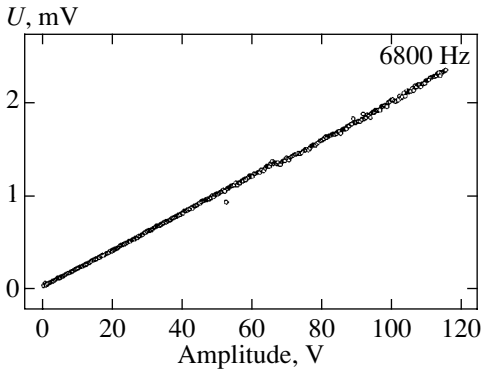

Fig. 4. The optical response U measured for the laser beam reflected from the surface of an oscillating free-standing LC film (CS-1029) as a function of the amplitude of a sinusoidal voltage applied to the film. The applied electric field frequency was fixed at ν_0 = 6800 Hz.

3. LINEAR ELECTROMECHANICAL EFFECT IN FREE-STANDING FILMS OF CS-1029 LIQUID CRYSTAL: RESULTS AND DISCUSSION

For the experimental verification of linearity of the equation of motion, we have studied the film deformation as a function of the amplitude of applied sinusoidal voltage. The results presented in Figs. 4 and 5 give unambiguous evidence that the linear approximation (9) is valid.

However, the results obtained by some researchers [3, 20] led to a conclusion that the air inertia may significantly influence the free-standing film dynamics via the so-called associated air mass effect [21]. For example, the data in [3, 20] indicate that the oscillating free-standing film moves together with an associated air layer having a thickness of several millimeters. It is interesting to note that restricting the air space around the oscillating film by the glass walls parallel to the film

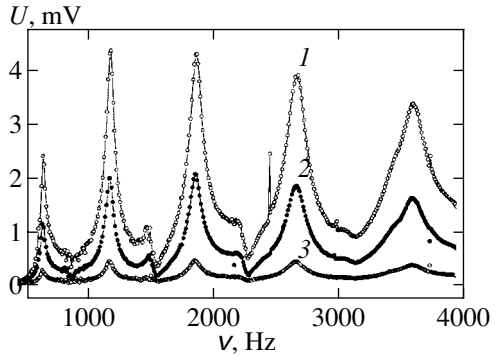

Fig. 5. The spectra of mechanical oscillations of a free-standing LC film (CS-1029, $N = 73$, $S = 3 \times 11$ mm^2) measured for three values of the applied voltage amplitude A = 40 (*1*), 20 (*2*), and 4 V (*3*).

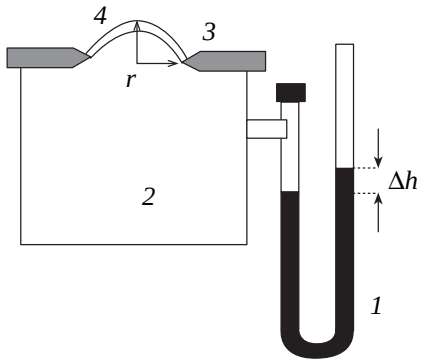

Fig. 6. A schematic diagram of the experimental setup for measuring the surface tension of a smectic liquid crystal: (*1*) U-shaped water pressure gauge; (*2*) glass cylinder; (*3*) round lid with a hole of radius r; (*4*) free-standing film sample.

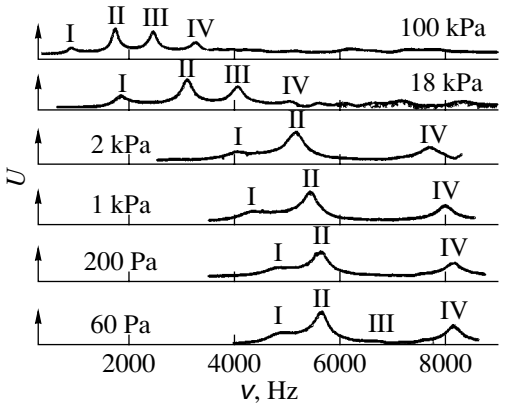

Fig. 7. Evolution of the spectra of mechanical oscillations of a free-standing LC film (CS-1029, $N = 57$, $S = 2 \times 10$ mm^2, $A = 80$ V) with a change in the atmospheric pressure.

plane, with the spacing d variable from 10 μm to 10 mm (Fig. 1c), did not significantly affect the spectrum of mechanical oscillations. At the same time, replacing air in the chamber by helium, possessing a considerably smaller density ($\rho = 0.17 \times 10^{-3}$ g/cm^3) than air, led to a significant (twofold) shift of the spectral bands toward higher frequencies, in accordance with the decreasing gas medium density.

Subsequent calculations involve the surface tension of a CS-1029 film. The surface tension was measured using a setup schematically depicted in Fig. 6. The excess air pressure Δp in a closed volume of glass cylinder *2* was measured by an U-shaped water pressure gauge *1*. As the air pressure was gradually increased, the LC film in a circular frame of radius r deformed from plane to hemisphere *3*. The surface tension was calculated by the Laplace formula $\Delta p = 4\sigma/r$. The room-temperature surface tension for a CS-1029 liquid crystal was $\sigma = 35.3$ din/cm.

Figure 7 shows the evolution of the spectrum of mechanical oscillations of a rectangular free-standing film (CS-1029, $N = 57$, $S = 2 \times 10$ mm^2, $A = 80$ V) measured on decreasing the air pressure from 100 kPa to 60 Pa. Measured at atmospheric pressure, the spectrum contained four rather narrow spectral bands I-IV with $Q_{\rm I} = 8$, $Q_{\rm II} = 17$, $Q_{\rm III} = 17.3$, and $Q_{\rm IV} = 18$ peaked at 860, 1690, 2420, and 3240 Hz, respectively. This pattern can be considered with a good accuracy as a superposition of two symmetric modes, ν_{11} and ν_{33}, and two asymmetric modes, ν_{22} and ν_{44}, obeying the following relationships:

$$\nu_{11} \approx \nu_{22}/2 \approx \nu_{33}/3 \approx \nu_{44}/4. \qquad (11)$$

The simultaneous appearance of even and odd modes in the spectrum indicates that a real distribution of external forces acting upon the film surface is more complicated than that in a homogeneous case [5, 6].

As the air pressure decreases, the spectrum shifts toward high frequencies and transforms in shape until reaching the following characteristics at a minimum pressure of 60 Pa (Fig. 8): (I) 860 ⟶ 4970 Hz, $Q_{\rm I} = 9.4$; (II) 1690 ⟶ 5590 Hz, $Q_{\rm II} = 18.6$; (III) 2420 ⟶ 6750 Hz, $Q_{\rm III} = 33.7$; (IV) 3240 ⟶ 8180 Hz, $Q_{\rm IV} = 35$. It is interesting to note that, on reaching rough vacuum conditions, the Q values of peaks I a II did not change, whereas the Q values of peaks III and IV almost doubled evidencing a considerable decrease in the level of dissipation.

Using the values of eigenmodes of a rectangular membrane [5, 6], it is possible to estimate the average two-dimensional film density $\langle \rho_s \rangle = N_{\rm eff} h \rho_{lc}$ and, assuming the film to be homogeneous, calculate an effective number of smectic layers $N_{\rm eff}$. The expressions for eigenmodes is as follows:

$$\nu_{nm} = \frac{1}{2}\sqrt{\frac{2\sigma}{\rho_s}\left(\frac{n^2}{a^2} + \frac{m^2}{b^2}\right)}, \qquad (12)$$

where n and m are integers. For $\sigma = 35.3$ din/cm, $\rho_{lc} = 0.95$ g/cm^3, $a = 2$ mm, $b = 10$ mm, and $n = m = 1$, we obtain at a normal air pressure (for $\nu_{11} = 860$ Hz) the average density $\langle \rho_s(100\ \text{kPa}) \rangle = 6.2 \times 10^{-4}$ g/cm^2 and

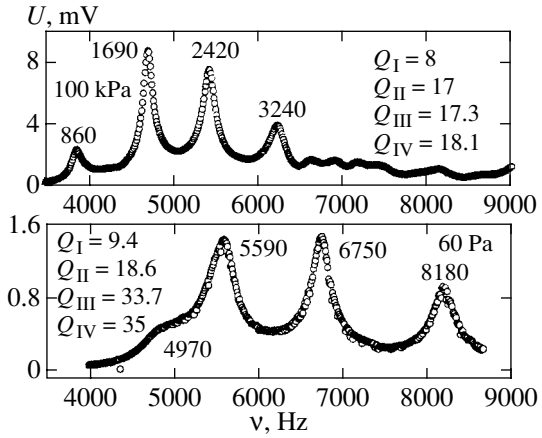

Fig. 8. The spectra of mechanical oscillations of a free-standing LC film (CS-1029, $N = 57$, $S = 2 \times 10$ mm^2, $A =$ 80 V) measured at an atmospheric pressure of 100 kPa and 60 Pa. The latter spectrum was measured under the same conditions as that in Fig. 7, except that a different area was probed. Figures at the peaks indicate resonance frequencies (Hz); also indicated are the corresponding Q values (numbered as in Fig. 7). The resonance bandwidth was measured at an 0.7 peak height level. The Q values of overlapped peaks (I and II) in the bottom spectrum were determined upon mathematical separation into components.

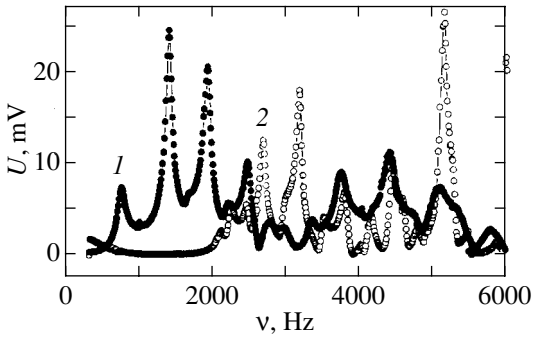

Fig. 9. The spectra of mechanical oscillations of a thick free- standing LC film (CS-1029, $N = 122$, $S = 2 \times 10$ mm^2, $A = 80$ V) measured at (*1*) normal atmospheric pressure of 100 kPa and (*2*) in a rough vacuum of 60 Pa.

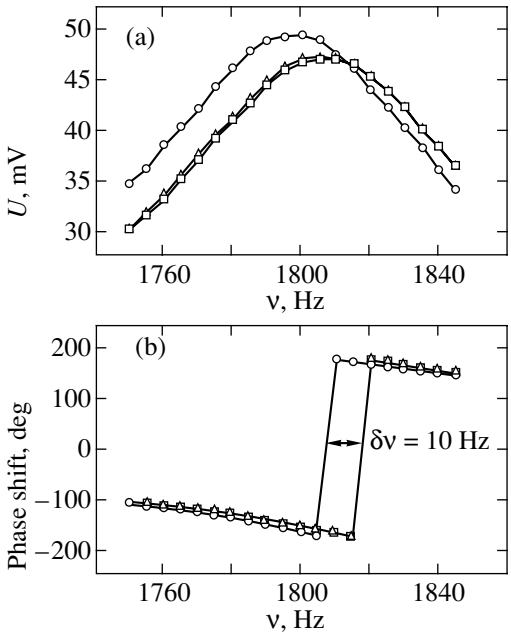

Fig. 10. (a) A portion of the spectrum of mechanical oscillations of a thick free-standing film in the vicinity of one resonance peak and (b) the corresponding plot of phase shift versus frequency. The data were obtained (▵, ▫) in air at normal pressure and (○) in the presence of ethanol vapors (CS-1029, $N = 72$, $S = 3 \times 11$ mm^2, $A = 100$ V, $T = 24$°C).

the effective number of layers $N_{\text{eff}}(100\text{ kPa}) = 1715$. At a low pressure (for $\nu_{11} = 4970$ Hz), the corresponding values are $\langle\rho_s(60\text{ Pa})\rangle = 1.9 \times 10^{-5}$ g/cm^2 and $N_{\text{eff}}(60\text{ Pa}) = 52$. Therefore, the effective film thickness in a rough vacuum ($h_{\text{eff}} = 52l$ almost coincides with the value ($h_{\text{eff}} = 57l$) determined independently using the experimentally measured reflection spectra.

The spectra presented in Fig. 8 were analyzed by a two-parametric procedure of fitting to formula (9). The parameters were (i) the experimentally measured resonance frequency and (ii) the resonance bandwidth determined at an 0.7 peak height level. This treatment yielded the average two-dimensional film densities $\langle\rho_s(100\text{ kPa})\rangle = 6.0 \times 10^{-4}$ g/cm^2 and $\langle\rho_s(60\text{ kPa})\rangle = 1.88 \times 10^{-5}$ g/cm^2, which almost exactly coincide with the analogous values estimated using the eigenmodes (12) of the two-dimensional membrane. The corresponding surface viscosities at normal and reduced pressure are $\eta_s(100\text{ kPa}) = 8.8 \times 10^{-3}$ g/s and $\eta_s(60\text{ kPa}) = 1.5 \times 10^{-3}$ g/s, respectively. Note that the pressure dependence of the surface viscosity may be related to a change in the film oscillation modes.

Figure 9 shows the spectrum of mechanical oscillations of a thicker film measured at normal pressure and in a rough vacuum (60 Pa). As expected, the sensitivity of the film with respect to the pressure decreases with increasing thickness. As the film thickness grows from $N = 57$ to 122, the absolute frequency shift decreased (see Figs. 8 and 9) by a factor of more than three, from $\Delta\nu_{11}(N = 57) = 4110$ Hz to $\Delta\nu_{11}(N = 122) = 1380$ Hz.

Figures 10–12 demonstrate the effect of the vapors of volatile organic solvents on the intrinsic frequencies of a free-standing film. Depending on the solvent type, the spectrum reversibly shifted toward lower (kerosene, ethanol) or higher (toluene) frequencies. Obviously, this behavior cannot be explained only by the associated mass effect or by the adsorption of solvent vapor (with the corresponding decrease in the surface tension).

It should be noted that the associated mass effect was also observed for solid membranes [21]. However, the frequency shift achieved at a low air pressure in

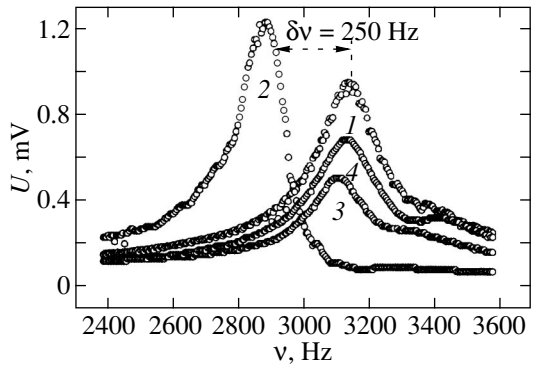

Fig. 11. Evolution of the spectrum of mechanical oscillations of a free-standing film measured under various conditions (a time interval between sequential measurements was about 12 min): (*1*) in air; (*2*) in the presence of kerosene vapors (a Petri dish with kerosene placed at a distance of 2 cm from the film); (*3, 4*) two sequential spectra recorded after removal of the dish with kerosene from the chamber.

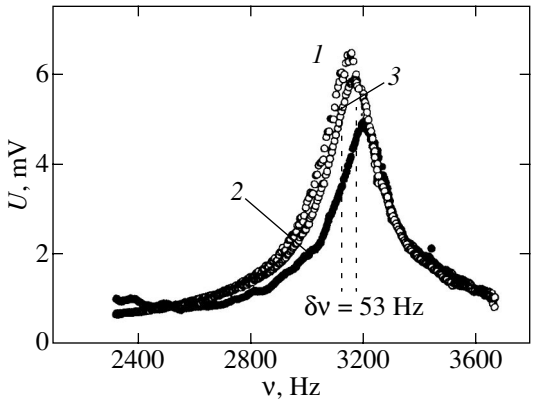

Fig. 12. Evolution of the spectrum of mechanical oscillations of a free-standing film measured under various conditions: (*1*) in air; (*2*) in the presence of toluene vapors; (*3*) after evacuating toluene vapors.

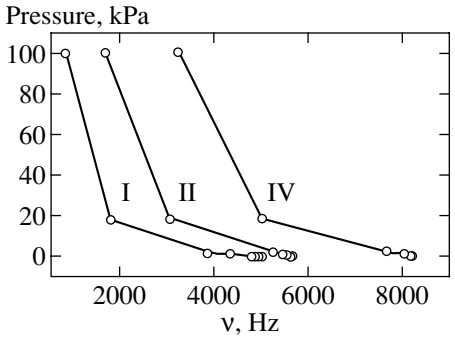

Fig. 13. The plots of air pressure versus resonance band positions.

such membranes was relatively small because it was difficult to obtain sufficiently thin (i.e., low-weight) samples. For example, evacuating air in experiments with a 50-μm-thick silicon membrane led to a resonance band shift as low as 2% [21].

As is seen in Fig. 7, a decrease in the pressure is accompanied, besides the frequency shift, by a change in the relationship (11). A comparison of the experimental spectra presented in Fig. 7 to the theoretical formula (12) derived for a Rayleigh membrane shows that the spectra measured at low pressures disagree with the model of an elastic membrane with homogeneous ρ_s and σ. Apparently, the conditions of low pressure (and a small associated air mass) increase the role of film inhomogeneities in the region of meniscus. The spectrum of an inhomogeneous linearly elastic film may, in principle, significantly differ from that determined by formula (12).

Possessing sufficiently high Q, a small-size thin ($h \sim$ 200 nm) free-standing ferroelectric film of CS-1029 represents essentially a nanomechanical pressure detector free of any moving parts. The range of measured pressures (760–0.5 Torr) and sensitivity threshold (0.5 Torr) of this detector are almost the same as those of a classical U-shaped liquid pressure gauge. Figure 13 shows a typical plot of the air pressure versus resonance band position.

4. STUDY OF THE SPONTANEOUS POLARIZATION SIGN INVERSION IN FERROELECTRIC LIQUID CRYSTALS OF THE 8PPyO6 TYPE

As was demonstrated above, the free-standing film shape must not change as a result of interaction of the electric field with an arbitrarily directed surface polarization. In a real experiment, the film surface is deformed as a result of the electric field interaction with a spontaneous volume polarization of the smectic *C** liquid crystal inducing the back-flow. The resulting crystal deformation depends on the applied field polarity and on the spontaneous polarization sign. This combination of properties makes the measurement of free-standing film oscillations a useful tool for studying the phenomenon of spontaneous polarization sign inversion in ferroelectric chiral liquid crystals [22, 23].

The spontaneous polarization is usually measured by two routine techniques based on detecting a change in the current passing through a capacitor, when a substance studied is placed between the capacitor plates [22]. The current is induced either by applying a voltage of some special shape or by heating. In this case, there is always an open question concerning the ratio of contributions of the surface and volume polarizations to the current measured. In most cases, the volume polarization effect dominates. However, in the case of ferroelectric liquid crystals capable of changing the spontaneous polarization sign in the vicinity of the inversion

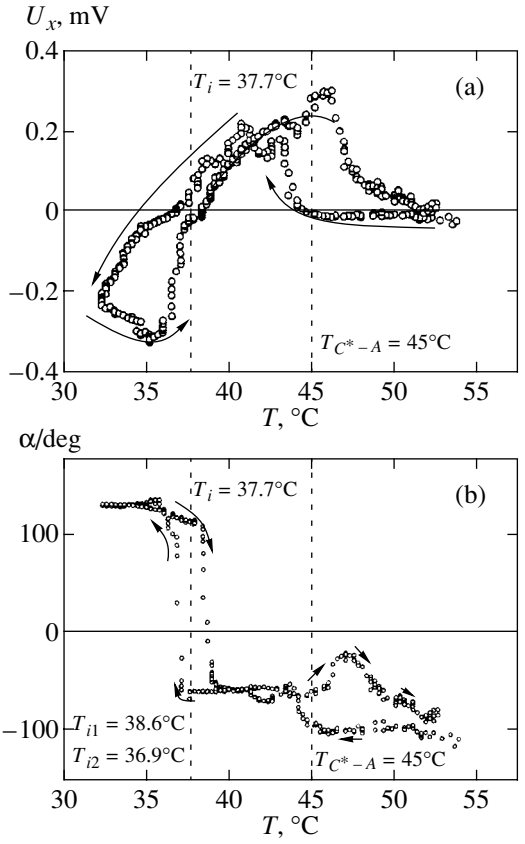

Fig. 14. Plots of (a) $U_x(\nu)$ and (b) phase angle α versus temperature for an 8PPyO6 free-standing film ($S = 3 \times 8$ mm^2, $\nu_0 = 2500$ Hz, $A = 44$ V). Arrows indicate the direction of temperature variation.

point or upon transition through this point, the surface polarization contribution (e.g., due to a thermal or electric modulation of the electric double layer) may become comparable to or even greater than the volume spontaneous polarization contribution.

We have studied the linear electromechanical effect depending on the temperature in a liquid crystal of the 8PPyO6 type exhibiting inversion of the spontaneous polarization sign at $T_i = 37.7$°C, as determined by the repolarization current and pyroelectric measurements [22].

The free-standing films were prepared using an sm*A* phase of 8PPyO6 at $T = 54$°C, and then the polarization was measured on slowly ($dT/dt = 1$ K/min) cooling the sample down to $T = 32$°C. Then the procedure was repeated in the heating mode. The response signal amplitude and phase were detected on the first harmonic $U_x(\nu) = U(\nu)\cos\alpha$ of a sinusoidal voltage of constant frequency and amplitude applied to the film. The results of these experiments are presented in Fig. 14. As change in the electrooptical response sign in the vicinity of $T_i = 37.7$°C (Fig. 14a) was accompanied by an almost 180° phase jump (Fig. 14b). The temperature variation exhibits a hysteresis, with the points of inversion (albeit differing by $T_{i1} - T_{i2} = 1.7$°C) being close to the value determined in the independent experiment [22]. A decrease in the rate of the sample heating and cooling virtually did not affect the T_{i1} and T_{i2} values. The hysteresis curve did not change depending on whether the sample was (or not) heated above the phase transition temperature $T_{C^*-A} = 45$°C. It should be noted that a hysteresis behavior is typical of the contact angle between meniscus and substrate [24].

Thus, the results of investigation of the linear electromechanical effect depending on the temperature provide unambiguous evidence of the inversion of the spontaneous polarization sign in 8PPyO6.

5. CONCLUSION

The results of our experimental study of the linear electromechanical effect in a free-standing ferroelectric LC film showed that, as the external pressure decreases, the spectrum of mechanical oscillations varies and eventually fails to obey the Rayleigh model of a homogeneous elastic membrane This deviation from the classical behavior can be explained by inhomogeneity of the film, in particular, by dependence of the dynamic characteristics of the membrane on the properties of meniscus participating in the system oscillations.

It was demonstrated that a free-standing ferroelectric LC film oscillating due to the inverted piezoelectric effect can be used as an air pressure transducer or a sensor of volatile chemical compounds.

Measurements of the electromechanical effect in the liquid crystals studied allowed the surface viscosity, an average two-dimensional density, and the temperature of inversion of the spontaneous polarization sign of a free-standing ferroelectric film to be determined.

ACKNOWLEDGMENTS

One of the authors (S.V.Ya) is grateful to L.M. Blinov, A.A. Saranin, and A.N. Zotov for fruitful discussions.

This study was supported by the Ministry of Education, Science, Sport, and Culture of Japan and by the Russian Foundation for Basic Research (project no. 99-02-16484).

REFERENCES

1. P. Pieranski, L. Beliard, J.-Ph. Tournellec, *et al.*, Physica A (Amsterdam) **194**, 364 (1993).
2. G. Fridel, Ann. Phys. (Paris) **18**, 273 (1922).
3. E. I. Kats and V. V. Lebedev, *Fluctuational Effects in the Dynamics of Liquid Crystals and Films* (Springer-Verlag, New York, 1993).
4. I. Kraus, Ch. Bahr, I. V. Chikina, and P. Pieranski, Phys. Rev. E **58**, 610 (1998).
5. J. W. Rayleigh, *The Theory of Sound* (McMillan, London, 1929; Gostekhizdat, Moscow, 1955).

6. P. M. Morse, *Vibration and Sound* (McGraw Hill, New York, 1936; GITTL, Moscow, 1949), p. 142.
7. Ch. Bahr and D. Fliegner, Phys. Rev. Lett. **70**, 1842 (1993).
8. K. Miyano, Phys. Rev. A **26**, 1820 (1982).
9. I. Kraus, Ch. Bahr, I. V. Chikina, and P. Pieranski, Phys. Rev. E **58**, 610 (1998).
10. P. E. Cladis, P. L. Finn, and H. R. Brand, Phys. Rev. Lett. **75**, 1518 (1995).
11. M. Brazovskaya, H. Dumoulin, and P. Pieranski, Phys. Rev. Lett. **76**, 1655 (1996).
12. A. Boudaoud, Y. Couder, and M. Ben Amar, Phys. Rev. Lett. **82**, 3847 (1999).
13. S. V. Yablonskii, Toshiyasu Oue, Hidetaka Nambu, *et al.*, Appl. Phys. Lett. **75**, 64 (1999).
14. E. I. Demikhov, S. A. Pikin, and E. S. Pikina, Phys. Rev. E **52**, 6250 (1995).
15. Sadahito Uto, Eisuke Tazoh, Masanori Ozaki, and Katsumi Yoshino, J. Appl. Phys. **82**, 2791 (1997).
16. Sadahito Uto, Masanori Ozaki, and Katsumi Yoshino, Appl. Phys. Lett. **74**, 117 (1999).
17. A. Jakli, L. Bata, A. Buka, and N. Eber, Ferroelectrics **69**, 153 (1986).
18. A. Jakli and A. Saupe, Liq. Cryst. **9**, 519 (1991).
19. A. Jakli and N. Eber, in *Modern Topics in Liquid Crystals* (World Scientific, Singapore, 1993), p. 235.
20. I. Kraus, Ch. Bahr, and P. Pieranski, Mol. Cryst. Liq. Cryst. **262**, 1289 (1995).
21. B. S. Berry and W. C. Pritchet, IBM J. Res. Dev. **19**, 334 (1978).
22. L. M. Blinov, L. A. Beresnev, D. Demus, *et al.*, Mol. Cryst. Liq. Cryst. **292**, 277 (1997).
23. L. M. Blinov, L. A. Beresnev, D. Demus, S. V. Iablonski, and S. V. Pikin, in *Dynamics and Defects in Liquid Crystals* (Gordon and Breach, Canada, 1998), p. 305.
24. P. G. de Genes, Rev. Mod. Phys. **57**, 827 (1985).

Translated by P. Pozdeev