Journal of Experimental and Theoretical Physics, Vol. 90, No. 4, 2000, pp. 571–578.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 4, 2000, pp. 659–667.

NUCLEI, PARTICLES, AND THEIR INTERACTION

Structure of Higher Order Corrections to the Lipatov Asymptotic Form

I. M. Suslov
Kapitza Institute of Physical Problems, Russian Academy of Sciences, Moscow, 117334 Russia
e-mail: suslov@kapitza.ras.ru
Received November 1, 1999

Abstract—High orders of perturbation theory can be calculated by the Lipatov method, whereby they are determined by saddle-point configurations, or instantons, of the corresponding functional integrals. For most field theories, the Lipatov asymptotic form has the functional form $ca^N\Gamma(N + b)$ (N is the order of perturbation theory) and the relative corrections to it are series in powers of $1/N$. It is shown that this series diverges factorially and its high-order coefficients can be calculated using a procedure similar to the Lipatov one: the Kth expansion coefficient has the form $\text{const}[\ln(S_1/S_0)]^{-K}\Gamma(K + (r_1 - r_0)/2)$, where S_0 and S_1 are the values of the action for the first and second instantons of this particular field theory, and r_0 and r_1 are the corresponding number of zeroth-order modes; the instantons satisfy the same equation as in the Lipatov method and are assumed to be renumbered in order of their increasing action. This result is universal and is valid in any field theory for which the Lipatov asymptotic form is as specified above. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Lipatov proposed a general method of calculating high orders of perturbation theory whereby these are determined by saddle-point configurations, or instantons, of the corresponding functional integrals [1]. On its appearance, the Lipatov method stimulated major discussion (see the collection of articles [2]) but was subsequently cast into doubt because of the possible existence of additional renormalon contributions [3]. In a recent work [4], the present author put forward a detailed discussion of the existing argumentation in support of renormalons and showed that this is untenable in the broad philosophical sense and in the mathematical sense: this clears any obstacles from applying the Lipatov method to a wide range of problems in theoretical physics.

The Lipatov method can be used to study any quantities [5] but the starting point is that it can be applied to functional integrals having the form

$$I(g) = \int D\varphi\varphi^{(1)}\ldots\varphi^{(M)} \times \exp(-S_0\{\varphi\} - gS_{int}\{\varphi\}), \quad (1)$$

where $\varphi^{(1)}$, …, $\varphi^{(M)}$ is a certain sample φ_{i_1}, …, φ_{i_M} from the set of integration variables φ_i contained within the symbol $D\varphi$. The expansion coefficients I_N of the integral (1) in terms of the coupling constant g are determined by the Cauchy integral

$$I_N = \oint_C \frac{dg}{2\pi i}\frac{I(g)}{g^{N+1}}, \quad (2)$$

in which the saddle-point method can be used for large N. The functional form of the Lipatov asymptote is given by

$$I_N = ca^N\Gamma(N + b), \quad N \longrightarrow \infty, \quad (3)$$

and the relative corrections to it have the form of a regular expansion in terms of $1/N$:

$$I_N = ca^N\Gamma(N + b) \times \left\{1 + \frac{A_1}{N} + \frac{A_2}{N^2} + \ldots + \frac{A_K}{N^K} + \ldots\right\}. \quad (4)$$

Calculation of the corrections to the asymptotic form provides important information on the expansion coefficients and is an alternative to the direct diagrammatic calculations of the lower orders: for instance, instead of calculating the fourth or fifth orders [6, 7], it is more economical to calculate A_1 or A_2. So far, the first corrections to the asymptotic form have only been calculated in φ^4 theory [8] and in a few quantum-mechanical problems [9, 10].

In the present paper, we study the behavior of the coefficients A_K for large K. This topic has not been studied theoretically and the only available data has been obtained by numerical methods: for a perturbation theory series in the problem of an anharmonic oscillator, Bender and Wu [9] determined the first ten coefficients A_K:

$$A_1 = -1.3194444, \quad A_3 = -7.0142876,$$
$$A_2 = -1.9385609, \quad A_4 = -40.118943,$$

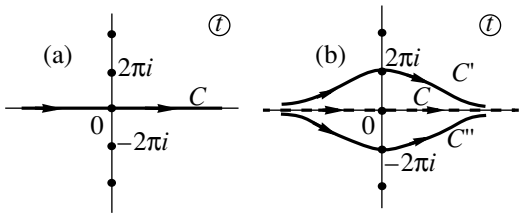

Fig. 1. (a) Saddle points and integration contour in integral (11). (b) In calculations of the asymptotic form of A_K the contour must be deformed since the point $t = 0$ corresponds to a singularity not a saddle point.

$$A_5 = -305.5223, \qquad A_8 = -3.65\times 10^5, \tag{5}$$

$$A_6 = -2808.09, \qquad A_9 = -4.4\times 10^6,$$

$$A_7 = -2.995\times 10^4, \quad A_{10} = -1\times 10^8.$$

The rapid growth of these coefficients indicates that the series in (4) diverges.

Another example which can easily be studied is the zero-dimensional limit of φ^4 theory. In this case, the functional integral in fact reduces to a single one

$$I(g) = \int_0^\infty d\varphi \varphi^M \exp(-\varphi^2 - g\varphi^4), \tag{6}$$

and its expansion coefficients can be calculated in the explicit form:

$$I_N = \frac{2^{M/2}}{2\sqrt{2\pi}} \frac{\Gamma\left(N+\frac{M+3}{4}\right)\Gamma\left(N+\frac{M+1}{4}\right)}{\Gamma(N+1)}(-4)^N. \tag{7}$$

By isolating the asymptotic form for $N \longrightarrow \infty$ and expressing the result in the form (4), we obtain the following expression for the coefficients A_K for $K \longrightarrow \infty$ (see Appendix)

$$A_K = \mathrm{Re}\frac{2(1+e^{\pi i M})}{(2\pi i)^{K+1}}\Gamma(K), \tag{8}$$

whose functional form is similar to the Lipatov asymptotic form (3) but with the complex parameters a and c.

In the present study, we shall show that factorial divergence of the series in (4) also occurs in the general case and a universal result [see formula (47)] valid for any field theory using the Lipatov asymptotic form (3) can be obtained for the asymptotic form A_K.

2. SIMPLE EXAMPLE AND QUALITATIVE PICTURE

The qualitative aspects involved in the calculation of the asymptotic form of A_K can be conveniently demonstrated by calculating the corrections to the Stirling formula:

$$\Gamma(N+1) = \sqrt{2\pi N}e^{-N}N^N \times\left\{1+\frac{A_1}{N}+\frac{A_2}{N^2}+\ldots+\frac{A_K}{N^K}+\ldots\right\}. \tag{9}$$

The result is well known for the logarithmic form of expression (9): in this case it is possible to find a general term of the series known as the Stirling series [11]. By calculating the exponential function of the Stirling series using factorial series algebra [5], we can easily find the asymptotic form of A_K:

$$A_K = -\mathrm{Re}\frac{2\Gamma(K)}{(2\pi i)^{K+1}}, \quad K \longrightarrow \infty. \tag{10}$$

We shall subsequently show how this result is obtained.

Using the definition of a gamma function and making the substitutions $x \longrightarrow Nx$ and $t = \ln x$, we have

$$\Gamma(N+1) = \int_0^\infty dx x^N e^{-x} = Ne^{-N}N^N \times\int_0^\infty dx \exp\{-N[x-1-\ln x]\} \tag{11}$$

$$= Ne^{-N}N^N\int_{-\infty}^{\infty} dt e^t \exp\{-N[e^t-1-t]\}.$$

For large N, the saddle-point condition has the form $e^t - 1 = 0$ so that there is a set of saddle points $t_s = 2\pi i s$, $s = 0, \pm 1, \pm 2, \ldots$, lying on an imaginary axis (Fig. 1a). The integration contour in (11) passes through the saddle point $t = 0$ and satisfies all the conditions for validity of the saddle-point method [12]. Thus, its deformation is not required and the other saddle points can be neglected. Calculating the integral in the saddle-point approximation yields the Stirling's formula.

Formally isolating the asymptotic form, we identically set

$$\Gamma(N+1) = \sqrt{2\pi N}e^{-N}N^N F(1/N) \tag{12}$$

and making the substitution

$$\epsilon = 1/N, \tag{13}$$

we have for the function $F(\epsilon)$ introduced by us

$$F(\epsilon) = \frac{1}{\sqrt{2\pi\epsilon}}\int_{-\infty}^{\infty} dt e^t \exp\left\{-\frac{e^t-1-t}{\epsilon}\right\}. \tag{14}$$

Expanding (14) as a series in terms of ϵ gives the required coefficients A_K which are calculated by analogy with (2):

$$A_{K-1} = \oint_C \frac{d\epsilon}{2\pi i}\frac{1}{\sqrt{2\pi\epsilon}} \times \int_{-\infty}^{\infty} dt e^t \exp\left\{-\frac{e^t-1-t}{\epsilon} - K\ln\epsilon\right\}. \quad (15)$$

This is an exact expression which for large K can be calculated using the saddle-point approximation. The saddle-point conditions yield the set of solutions

$$t_s = 2\pi i s, \quad \epsilon_s = -\frac{t_s}{K}, \quad s = \text{integer}, \quad (16)$$

so that in the complex plane t the saddle points are formally the same as those in the calculations of the leading asymptotic term. However, for the integrand at the sth saddle point we can easily obtain the estimate

$$\sim \exp\{-K - K\ln\epsilon_s\} \sim t_s^{-K} K!, \quad (17)$$

from which it is clear that the solution with $s = 0$ does not in fact correspond to a saddle point but to a singularity.[1] Hence, the integration contour over t cannot pass through the point $t = 0$ but must be deformed and pass through one of the neighboring saddle points, $2\pi i$ or $-2\pi i$ (Fig. 1b) which, because of (17), gives the required asymptotic form of $A_K \sim (2\pi)^{-K}K!$ [see (10)].[2]

A similar situation is encountered in the general case. When the coefficients A_K are calculated in the saddle-point approximation, the instanton equation is the same as that used to calculate the Lipatov asymptotic form. However, using the same solution as in the last case yields a singularity rather than a saddle point (because of the additional integration over ϵ). There is thus a need to consider other solutions of the instanton equation which can be numbered in order of increasing action corresponding to them. If the Lipatov asymptotic form is determined by the first instanton, having the smallest action, the principal contribution to the asymptotic form of A_K is made by the second instanton.

3. GENERAL CASE

Calculations of the Lipatov asymptotic form (3) are fairly cumbersome if the aim is to find all its parameters a, b, and c. However, if the analysis is confined to determining the parameters a and b, it is possible to have simple structural calculations which reduce to a formal expansion near the saddle point and isolate the dependence on N. We shall demonstrate these calculations for the case of φ^4 theory; however, we do not need the explicit form of the action and we shall only use its characteristic properties of homogeneity

$$S_0(\lambda\varphi) = \lambda^2 S_0(\varphi), \quad S_{int}(\lambda\varphi) = \lambda^4 S_{int}(\varphi). \quad (18)$$

Similar properties of homogeneity are encountered in other field theories and, with slight modifications, the scheme put forward subsequently also holds for the general case.

According to (1) and (2), the expansion coefficients are given by

$$I_{N-1} = \oint_C \frac{dg}{2\pi i}\int D\varphi\varphi^{(1)}\ldots\varphi^{(M)} \times \exp(-S_0\{\varphi\} - gS_{int}\{\varphi\} - N\ln g). \quad (19)$$

We introduce the new variable

$$\phi = \varphi\sqrt{g} \quad (20)$$

and set

$$S\{\varphi\} = S_0\{\varphi\} + S_{int}\{\phi\}. \quad (21)$$

In terms of the new variable ϕ, the saddle-point conditions have the form

$$S'\{\phi_c\} = 0, \quad g_c = \frac{S\{\phi_c\}}{N}, \quad (22)$$

and expanding the expression in the exponential function (19) as far as quadratic terms in $\delta\phi = \phi - \phi_c$ and $\delta g = g - g_c$ gives

$$-N - N\ln g_c - \frac{N}{2}\frac{(\delta\phi, S''\{\phi_c\}\delta\phi)}{S\{\phi_c\}} - \frac{N}{2g_c^2}(\delta g)^2. \quad (23)$$

We use a symbolic notation, denoting the first and second functional derivatives by single and double primes and taking these to be a vector and a linear operator, respectively; the variables of integration φ_i contained within $D\varphi$ are taken to be components of the vector φ. Bearing in mind that because of (20)

$$\delta\phi = \sqrt{g_c}\left(\delta\phi + \frac{\delta g}{2g_c}\varphi_c\right), \quad \delta\varphi = \varphi - \varphi_c, \quad (24)$$

and shifting the origin $\delta\varphi$, we have

$$I_{N-1} = e^{-N} g_c^{-N+1-M/2}\int_{-\infty}^{\infty}\frac{dt}{2\pi}\int D\varphi\phi_c^{(1)}\ldots\phi_c^{(M)} \times \exp\left(-\frac{1}{2}(\delta\varphi, S''\{\phi_c\}\delta\varphi) + \frac{N}{2}t^2\right), \quad (25)$$

where we have set $\delta g = ig_c t$.

[1] The solutions (16) are written assuming $\epsilon \neq 0$ which does not hold for $s = 0$. A similar observation must be made with reference to formula (38) below.

[2] The integration contour over ϵ in (15) is conveniently drawn slightly to the right of the imaginary axis, enveloping the left half-plane over an infinitely distant contour; in this case for $\mathrm{Im}\,\epsilon < 0$ the integration contour over t is shifted upward and passes through the point $2\pi i$ whereas for $\mathrm{Im}\,\epsilon > 0$ it is shifted downward and passes through the point $-2\pi i$.

We make the linear substitution $\delta\varphi \longrightarrow \hat{S}\delta\varphi$ with $\det\hat{S} = 1$, which diagonalizes the matrix of the operator $S''\{\phi_c\}$ and we set

$$D\varphi = D'\varphi\prod_{i=1}^{r} d\tilde{\varphi}_i, \tag{26}$$

where we have isolated r variables of integration (denoted by the tilde) which correspond to zero eigenvalues of the operator $S''\{\phi_c\}$ and do not in fact appear in the exponential function (25). In order to ensure correct integration over zero-order modes, the following partition of unit is introduced below in the integrand (25)

$$1 = \prod_{i=1}^{r}\int d\lambda_i\delta(\lambda_i - f_i\{\varphi\}), \tag{27}$$

where λ_i are collective variables. An example of such a variable is the instanton center x_0 defined as

$$\int d^dx|\varphi(x)|^4(x - x_0) = 0, \tag{28}$$

for which integration of the type (27) has the form

$$1 = \int d^dx_0\delta\left(x_0 - \frac{\int d^dx|\varphi(x)|^4x}{\int d^dx|\varphi(x)|^4}\right). \tag{29}$$

By introducing collective variables (which can also be the instanton "orientation," its radius, and so on [5]), we can confine ourselves to homogeneous functions $f_i\{\varphi\}$ [compare with (29)] where the degree of homogeneity can be considered to be zero without limiting the generality: if $f_i\{\mu\varphi\} = \mu^p f_i\{\varphi\}$, the substitution $\lambda_i \longrightarrow \mu^p\lambda_i$ eliminates the factor μ^p from (27). We linearize the arguments of the δ-functions in (27) near the saddle-point configuration

$$\begin{aligned} 1 &= \prod_{i=1}^{r}\int d\lambda_i\delta(\lambda_i - f_i\{\varphi_c\} - (f_i'\{\varphi_c\}, \delta\varphi)) \\ &= \prod_{i=1}^{r}\int d\lambda_i\delta(\lambda_i - f_i\{\varphi_c\} - \sqrt{g_c}(f_i'\{\varphi_c\}, \delta\varphi)), \end{aligned} \tag{30}$$

and select the instanton such that $\lambda_i - f_i\{\phi_c\} = 0$ (in (28) this corresponds to a choice of solution symmetric relative to the point $x = 0$); then ϕ_c becomes a function of λ_i. Substituting (26) and (30) into (25) and eliminating the δ-functions by integrating over $\delta\tilde{\varphi}_i$, we have

$$\begin{aligned} I_{N-1} &= e^{-N}g_c^{-N+1-(M+r)/2}\det[f'\{\phi_c\}]_P \\ &\times \int_{-\infty}^{\infty}\frac{dt}{2\pi}\int D'\varphi\int\prod_{i=1}^{r} d\lambda_i\phi_c^{(1)}(\lambda_i)\ldots\phi_c^{(M)}(\lambda_i) \\ &\times \exp\left(-\frac{1}{2}[(\delta\varphi, S''\{\phi_c\}\delta\varphi) - Nt^2]\right), \end{aligned} \tag{31}$$

where $f'\{\phi_c\}$ is an operator whose matrix consists of the columns $f_i'\{\phi_c\}$ and $[\ldots]_P$ is its projection onto the subspace of the zero-order modes. The integral over $D'\varphi$ and dt is determined by the determinant of the quadratic form in brackets in the exponential function (31) given by $(-N)\det[S''\{\phi_c\}]_{P'}$, although caution must be exercised when reducing this to a sum of squares [13]; the subscript P' indicates a projection on a subspace complementary to the subspace of the zero-order modes. Performing elementary transformations in (31), we obtain a result having the form (3) where

$$\begin{aligned} a &= \frac{1}{S\{\phi_c\}}, \quad b = \frac{M + r}{2}, \\ c &= (S\{\phi_c\})^{-(M+r)/2}\frac{(2\pi)^{(\mathcal{N}-r-2)/2}\det[f'\{\phi_c\}]_P}{\sqrt{-\det[S''\{\phi_c\}]_{P'}}} \\ &\times \int\prod_{i=1}^{r} d\lambda_i\phi_c^{(1)}(\lambda_i)\ldots\phi_c^{(M)}(\lambda_i), \end{aligned} \tag{32}$$

and $\mathcal{N}$ is the number of variables of integration contained within $D\varphi$ [this disappears from the answer on going over to a ratio of integrals of the type (1)].

Similar structural calculations can be made for the asymptotic form of the coefficients A_K. Making the substitution in (19)

$$g \longrightarrow \frac{S\{\phi_c\}}{N}g \tag{33}$$

and isolating the dependence on $\mathcal{N}$ corresponding to the asymptotic form (3), we have

$$\begin{aligned} I_{N-1} &= (S\{\phi_c\})^{-N+1}\exp(-N + N\ln N) \\ &\times N^{(M+r-3)/2}F\left(\frac{1}{N}\right), \end{aligned} \tag{34}$$

where

$$\begin{aligned} F\left(\frac{1}{N}\right) &= N^{-(M+r-1)/2}\oint_C\frac{dg}{2\pi i}\int D\varphi\varphi^{(1)}\ldots\varphi^{(M)} \\ &\times \exp\left(-N\frac{S\{\varphi\}}{S\{\phi_c\}g} + N - N\ln g\right)\Bigg|_{\phi = \varphi\sqrt{S\{\phi_c\}g/N}}. \end{aligned} \tag{35}$$

Setting $\epsilon = 1/N$ and expanding $F(\epsilon)$ as a series

$$F(\epsilon) = \tilde{A}_0 + \tilde{A}_1\epsilon + \tilde{A}_2\epsilon^2 + \ldots + \tilde{A}_K\epsilon^K + \ldots, \tag{36}$$

we have by analogy with (2)

$$\tilde{A}_{K-1} = \oint_C \frac{d\epsilon}{2\pi i}\epsilon^{(M+r-1)/2}\oint_C \frac{dg}{2\pi i}\int D\varphi\varphi^{(1)}\ldots\varphi^{(M)} \times \exp\left(-\frac{1}{\epsilon}\left[\frac{S\{\varphi\}}{S\{\phi_c\}g} - 1 + \ln g\right] - K\ln\epsilon\right)\Bigg|_{\phi=\varphi\sqrt{g\epsilon S\{\phi_c\}}}. \quad (37)$$

The coefficients $\tilde{A}_K$ are simply related to the unknown coefficients A_K but differ from them (see below).

For large K in (37) we can use the saddle-point method, for which the conditions have the form

$$S'\{\varphi\} = 0, \quad g_c = \frac{S\{\phi\}}{S\{\phi_c\}}, \quad \epsilon_c = \frac{\ln g_c}{K}, \quad (38)$$

and the function in the integrand for the saddle-point configuration is determined by the factor

$$\exp(-K - K\ln\epsilon_s) \sim (\ln g_c)^{-K}K!. \quad (39)$$

Taking into account the substitution (33), the first two equations (38) coincide with (22) but using the solution

$$\phi = \phi_c, \quad g_c = 1 \quad (40)$$

leads to a singularity not to a saddle point because of (39). Thus, other solutions of the system of the first two equations (38) must be sought for which two possibilities exist.

1. Using other branches of the logarithm. In accordance with (38), ϵ_c is determined by the logarithm of g_c and thus the substitution g_c with integer $g_c \longrightarrow g_c\exp(2\pi is)$ is not an identity transformation: in this case we have $\ln g_c \longrightarrow \ln g_c + 2\pi is$. For $\phi = \phi_c$, $g_c = \exp(2\pi is)$ we have $\epsilon_c = 2\pi is/K$ and the contribution to the asymptotic form of A_K is determined by the saddle points with $s = \pm 1$:

$$A_K \sim (2\pi)^{-K}K!. \quad (41)$$

This is exactly the same mechanism as that used to calculate the corrections to the Stirling formula: the g dependence of the function in the integrand of (19) is similar to the x dependence in (11).

2. Using other instantons. Let us assume that ψ_c is a solution of the equation $S'\{\phi\} = 0$ which differs from ϕ_c; then on account of (38) and (39) we have the contribution to the asymptotic form

$$A_K \sim K!\left[\ln\frac{S\{\psi_c\}}{S\{\phi_c\}}\right]^{-K}, \quad (42)$$

which is larger the smaller $S\{\psi_c\}$. The principal contribution comes from the second instanton (see end of Section 2) and has the lower estimate

$$A_K \gtrsim (\ln 2)^{-K}K!. \quad (43)$$

If $\phi_c(x)$ is a localized solution of the equation $S'\{\phi\} = 0$, we know that there also exists a solution $\psi_c(x)$ corresponding to two infinitely distant instantons $\phi_c(x)$ for which $S\{\psi_c\} = 2S\{\phi_c\}$; in general a solution ψ_c can exist such that $S\{\phi_c\} < S\{\psi_c\} < 2S\{\phi_c\}$ which yields (43). Since the contribution (43) is larger than (41), in any real field theory the second of these mechanisms is the principal one; the first mechanism is only important in various nondegenerate cases such as zero-dimensional theory [see (8)] when the solution of the instanton equation is unique.

Expanding the expression in the exponential function (37) near the second instanton as far as quadratic terms in $\delta\varphi$, δg, and $\delta\epsilon$ and making the substitutions $\delta g = ig_c t$, $\delta\epsilon = i\epsilon_c\tau$, we have

$$\tilde{A}_{K-1} = g_c\epsilon_c^{-K+(r+1)/2}(S\{\psi_c\})^{-M/2}e^{-K} \times \int_{-\infty}^{\infty}\frac{dt}{2\pi}\int_{-\infty}^{\infty}\frac{d\tau}{2\pi}\int D\varphi\psi_c^{(1)}\psi_c^{(M)} \times \exp\left\{-\frac{1}{2}\left[(\delta\varphi, S''\{\psi_c\}\delta\varphi) - \frac{t^2}{\epsilon_c} - K\tau^2\right]\right\}. \quad (44)$$

The number of zero-order modes r' for the second instanton generally differs from r; these are isolated as before by introducing a partition of unit of the type (27), giving a dependence on K having the form

$$\tilde{A}_K \sim (\ln g_c)^{-K}\Gamma\left(K + \frac{r'-r}{2}\right). \quad (45)$$

In order to find the relationship between $\tilde{A}_K$ and A_K, we substitute (36) with $\epsilon = 1/N$ into (34), make the substitution $N \longrightarrow N+1$, and use the series reexpansion rule given in the Appendix. As a result, we obtain

$$A_K = \frac{\tilde{A}_K}{\sqrt{2\pi}cg_c}, \quad (46)$$

where c is a coefficient appearing in the Lipatov asymptotic form (3) and determined by formula (32). Taking into account (46) and performing trivial transformations in (44), we obtain

$$A_K = c_1\left(\ln\frac{S\{\psi_c\}}{S\{\phi_c\}}\right)^{-K}\Gamma\left(K + \frac{r'-r}{2}\right), \quad (47)$$

where

$$c_1 = (S\{\psi_c\})^{-(M+r')/2}\left(\ln\frac{S\{\psi_c\}}{S\{\phi_c\}}\right)^{(r-r')/2} \times \frac{(2\pi)^{(\mathcal{N}-r'-4)/2}\det[f'\{\psi_c\}]_P}{c\sqrt{\det[S''\{\psi_c\}]_{P'}}} \quad (48)$$

$$\times \int \prod_{i=1}^{r'} d\lambda_i \psi_c^{(1)}(\lambda_i) \ldots \psi_c^{(M)}(\lambda_i).$$

In these structural calculations, we used the form of the functional integral (1) and the homogeneity relations (18) typical of φ^4 theory; thus, the parameter M appearing in the result (32) for b was determined by the number of cofactors in the preexponential function (1). In other field theories, several fields of various types generally occur and the homogeneity relations differ from (18); nevertheless, for a wide range of problems the result for b has the previous form (32) but the parameter M has a different meaning. However, the parameter M does not appear in the asymptotic formula (47), indicating that its validity is not confined to φ^4 theory: this is confirmed by the reasoning put forward in the following section.

4. HEURISTIC DERIVATION OF FORMULA (47)

In general, factorial series have an asymptotic form with complex parameters [see (8) and (10)] and the expansion coefficients I_N are determined by the real part of some complex expression. We shall specify this for large N,

$$I_N = \mathrm{Re}\{ca^N\Gamma(N+b)(1+\Delta_N) + \tilde{c}\tilde{a}^N\Gamma(N+\tilde{b}) + \ldots\}, \tag{49}$$

taking into account the Lipatov asymptotic form $ca^N\Gamma(N+b)$, the unknown power corrections to it denoted by Δ_N, and the contribution of the next instanton $\tilde{c}\tilde{a}^N\Gamma(N+\tilde{b})$; the corrections to the latter and the contribution of higher order instantons are shown by the ellipses. Removing the Lipatov asymptotic form from the brackets, we have

$$I_N = \mathrm{Re}\left\{ca^N\Gamma(N+b) \times \left[1+\Delta_N + \frac{\tilde{c}}{c}\left(\frac{\tilde{a}}{a}\right)^N N^{\tilde{b}-b} + \ldots\right]\right\}. \tag{50}$$

Assuming $\epsilon = 1/N$, we can see that the last term has an intrinsic singularity for $\epsilon = 0$ which may be attributed to the imaginary part of some factorial series [5]

$$\epsilon^{-\beta} e^{-\alpha/\epsilon} = \frac{\alpha^{-\beta}}{\pi}\mathrm{Im}\sum_K \Gamma(K+\beta)\left(\frac{\epsilon}{\alpha}+i0\right)^K, \tag{51}$$

whose substitution into (50) leads to an expansion in reciprocal powers of N. It is natural to assume that the expression in brackets (50) is an analytic function whose imaginary and real parts can only appear in a strictly determined combination. The real part of the series (51) is much larger than the imaginary one and should originate from contributions which are higher in the hierarchy than the last term in (50); only Δ_N can fulfill this role. Combining the second and third terms in brackets (50), we obtain[3]

$$I_N = \mathrm{Re}\left[ca^N\Gamma(N+b)\left\{1+\mathrm{const}\cdot\sum_K \Gamma(K+\tilde{b}-b) \times \frac{[\ln(a/\tilde{a})+i0]^K}{N^K} + \ldots\right\}\right]. \tag{52}$$

The singularity on the left-hand side of (51) is associated with high-order terms of the series [see the discussion of formula (4.10) in [5]] and the form of the general term given in (52) is in fact only valid for large K. Bearing in mind that the parameters a and b of the instanton contribution have the form (32) for a wide range of field theories, we return to the result (47) where, however, the coefficient c_1 no longer has the specific form (48).

We shall explain the meaning of these manipulations. As we know, the expansion of the function $f(\epsilon)$ as a power series in ϵ has a radius of convergence equal to the distance between the origin and the nearest singular point $f(\epsilon)$ on the complex plane. For a factorial series the radius of convergence is zero and a singularity should be found for $\epsilon = 0$. Characteristic singularities generating factorial series have the form of branch cuts at which the discontinuity decays exponentially for $|\epsilon| \longrightarrow 0$ [14] [see (51)]. It is deduced from the qualitative pattern established above that (a) Δ_N has the form of a factorial series in $1/N$; (b) the divergence of this series is determined by the second instanton; (c) the contribution of the latter in (50) contains a characteristic singularity generating these series. From this it is logical to conclude that the second and third terms in brackets (50) form a single entity, being related to the real and imaginary parts of the same analytic function.

This reasoning is merely based on the fact that the instanton contribution to the asymptotic form has the functional form (3). Thus, the result (47) is universal: it is not related to the specific field theory nor to the form of the quantity being studied (for example, single-particle or two-particle Green's function).

5. QUANTITATIVE RESULTS

We shall apply these results to the problem of an anharmonic oscillator [9]. This can be reduced to one-dimensional φ^4 theory [15] in which the instantons can easily be investigated (in particular, by using a mechanical analogy [16]); the localized solution of the instanton equation is unique and all other solutions are exhausted by multi-instanton configurations containing

[3] The term Δ_N also contains similar contributions from higher instantons which are small compared with those contained in (52).

several infinitely distant instantons. Thus, as ψ_c in (47) we need to take the two-instanton solution for which $S\{\psi_c\} = 2S\{\phi_c\}$, $r' = r + 1$ (an extraneous zero mode appears corresponding to the motion of two instantons relative to each other). Consequently, for an anharmonic oscillator we have

$$A_K = c_1\left(\frac{1}{\ln 2}\right)^K \Gamma\left(K + \frac{1}{2}\right). \quad (53)$$

The dependence (53) can be compared with the results of Bender and Wu (5) using c_1 as the fitting parameter; results are plotted in Fig. 2 for $c_1 = -1.4$.

Higher order instantons have been little studied in multidimensional φ^4 theory. An exception is the four-dimensional case for which an infinite series of instanton solutions was obtained analytically by Ushveridze [17]. The second instanton in this series, following the Lipatov one ($S\{\phi_c\} = 16\pi^2/3$) has the action $S\{\psi_c\} = 9\pi^2$ which gives the result for the asymptotic form of A_K

$$A_K = c_1\left(\ln\frac{27}{16}\right)^{-K} \Gamma\left(K + \frac{3}{2}\right) \quad (54)$$

(we assumed that $r' = r + 3$ because in view of the absence of spherical symmetry for the second instanton, three zero modes are added corresponding to its rotations in four-dimensional space). Unfortunately, there is no evidence that the Ushveridze series exhausts all the solutions; thus, the result (54) should be understood as a preliminary or lower estimate.

A method of determining the complete series of higher order instantons numerically was proposed in [18]. It would be desirable to use this method to check the result (54) and to find the second instantons in all existing field theories.

6. CONCLUSIONS

Expression (4) can be used to interpolate the coefficient function, by truncating the series at a finite number of terms and selecting the parameters A_K to ensure agreement with the lowest orders of perturbation theory known from direct diagrammatic calculations. This procedure is highly accurate and can reliably estimate the error but nevertheless is unsatisfactory in many respects. This is because when diverging series are summed, the analytic properties of the coefficient function [19] are significant and these are reproduced quite incorrectly in this procedure: the coefficient function is assigned a multiple pole at $N = 0$ but the intrinsic singularity is lost at $N = \infty$ because of the factorial divergence of the series in (4) [see formula (51)]. Qualitative allowance for the functional form of the asymptotic A_K in the form $c_1 a_1^K \Gamma(K + b_1)$ enables us to select basis functions exhibiting correct behavior for $N \longrightarrow \infty$ which should give a positive effect even when the number of fitting parameters is constant. Quantitative calculations of the asymptotic form can be used to determine three parameters a_1, b_1, and c_1 characterizing the coefficient function which is equivalent in efficiency to calculating the next three orders of perturbation theory. The calculations of a_1 and b_1 do not require functional integrals and can be reduced to solving nonlinear differential equations: the calculations of c_1 are of approximately the same complexity as the calculations of the leading Lipatov asymptotic form. This is incomparably easier than calculating the successive terms of a perturbation theory series where progression to the next order takes, on average, ten years

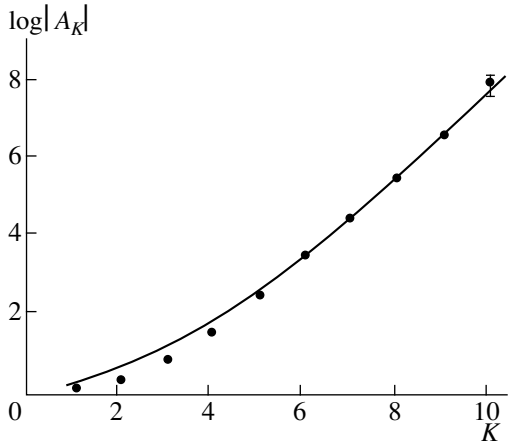

Fig. 2. Comparison between the asymptotic formula (53) for $c_1 = -1.4$ and the coefficients A_K determined numerically in [9]. The value of $|A_{10}|$ is given in [9] with a single significant digit (1×10^8) and the error corresponding to the range $(0.5–1.5) \times 10^8$ is indicated in the figure.

The author thanks L.N. Lipatov for discussions of preliminary results of this work and participants at seminars at the Institute of Physical Problems and the Physics Institute of the Russian Academy of Sciences for their interest in this work and useful discussions.

ACKNOWLEDGMENTS

This work was supported by INTAS (grant no. 96-0580) and the Russian Foundation for Basic Research (project no. 00-02-17129).

APPENDIX

Derivation of Formula (8)

Let us assume that two expansions exist:

$$F_N = 1 + \frac{A_1}{N} + \frac{A_2}{N^2} + \ldots + \frac{A_K}{N^K} + \ldots, \quad (A.1)$$

$$F_N = 1 + \frac{B_1}{N+p} + \frac{B_2}{(N+p)^2} + \ldots + \frac{B_K}{(N+p)^K} + \ldots. \quad (A.2)$$

If the second series is factorial,

$$B_K = \mathrm{Re}\{ca^K\Gamma(K+b)\}, \quad K \longrightarrow \infty, \tag{A.3}$$

it is easy to show by direct reexpansion that

$$A_K = \mathrm{Re}e^{-p/a}ca^K\Gamma(K+b), \quad K \longrightarrow \infty. \tag{A.4}$$

Making the substitution $N \longrightarrow N + \beta - 1$ in (9) and using (A.3) and (A.4), we obtain the result

$$\Gamma(N+\beta) = N^{\beta-1}\sqrt{2\pi N}e^{-N}N^N \times\left\{1 + \ldots - \frac{2\Gamma(K)}{N^K}\mathrm{Re}\frac{e^{-2\pi i\beta}}{(2\pi i)^{K+1}} + \ldots\right\} \tag{A.5}$$

and, applying it to relation (7) and using factorial series algebra [5], derive formula (8).

REFERENCES

1. L. N. Lipatov, Zh. Éksp. Teor. Fiz. **72**, 411 (1977) [Sov. Phys. JETP **45**, 216 (1977)].
2. *Large Order Behavior of Perturbation Theory*, Ed. by J. C. Le Guillou and J. Zinn-Justin (North-Holland, Amsterdam, 1990).
3. G.'t Hooft, in *The Whys of Subnuclear Physics*, Ed. by A. Zichichi (Plenum, New York, 1979).
4. I. M. Suslov, Zh. Éksp. Teor. Fiz. **116**, 369 (1999) [JETP **89**, 197 (1999)].
5. I. M. Suslov, Usp. Fiz. Nauk **168**, 503 (1998) [Phys. Usp. **41**, 441 (1998)].
6. S. G. Gorishny *et al.*, Phys. Lett. B **256**, 81 (1991); T. van Ritbergen, J. A. M. Vermaseren, and S. A. Larin, Phys. Lett. B **400**, 379 (1997).
7. H. Kleinert *et al.*, Phys. Lett. B **272**, 39 (1991).
8. Yu. A. Kubyshin, Teor. Mat. Fiz. **57**, 363 (1983).
9. C. M. Bender and T. T. Wu, Phys. Rev. D **7**, 1620 (1973).
10. S. V. Faleev and P. G. Silvestrov, Phys. Lett. A **197**, 372 (1995).
11. H. B. Dwight, *Tables of Integrals and Other Mathematical Data* (Macmillan, London, 1961, 4th ed.; Nauka, Moscow, 1977).
12. Yu. V. Sidorov, M. V. Fedoryuk, and M. I. Shabunin, *Lectures in the Theory of Complex-Variable Functions* (Nauka, Moscow, 1976).
13. I. M. Suslov, Zh. Éksp. Teor. Fiz. **105**, 560 (1994) [JETP **79**, 307 (1994)].
14. E. B. Bogomolny, Phys. Lett. B **67**, 193 (1977).
15. E. Brezin and G. Parisi, J. Stat. Phys. **19**, 269 (1978).
16. R. Finkelstein, R. Le Lovier, and M. Ruderman, Phys. Rev. **83**, 326 (1951).
17. A. G. Ushveridze, Yad. Fiz. **30**, 845 (1979) [Sov. J. Nucl. Phys. **30**, 436 (1979)]; Yad. Fiz. **32**, 1446 (1980) [Sov. J. Nucl. Phys. **32**, 747 (1980)].
18. G. G. Leptukh and A. G. Ushveridze, J. Phys. A **14**, 3085 (1981).
19. Yu. A. Kubyshin, Teor. Mat. Fiz. **58**, 137 (1984).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 4, 2000, pp. 579–583.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 4, 2000, pp. 668–672.

ATOMS, SPECTRA, RADIATION

Theory of Relativistic-Electron Transition Radiation in a Thin Metal Target

N. F. Shul'ga* and S. N. Dobrovol'skiĭ

Kharkov Institute of Physics and Technology, Kharkov, 310108 Ukraine
**e-mail: shulga@kipt.kharkov.ua*

Received February 9, 1999

Abstract—An analysis is made of the transition radiation of a relativistic electron in a thin metal target in the infrared wavelength range. It is shown that in this case, the spatial transverse dimensions responsible for forming the radiation may have macroscopic dimensions comparable with the size of the target and allowance for this effect may lead to significant distortion of the transition-radiation spectrum compared with a target having infinitely large transverse dimensions. *© 2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

When ultrashort bunches of relativistic electrons propagate through a thin layer of material, a coherent effect may occur in the transition radiation where all the bunch particles emit as a single particle having an effective total charge [1–3]. This effect is similar to coherent synchrotron radiation [4, 5] and occurs if the longitudinal dimension of the bunch is smaller than the length of the emitted wave. Experimental investigations of this effect were recently carried out using electron beams having energies of the order $\varepsilon \sim 100$ MeV in the far infrared ($\lambda \sim 0.1$ cm) [1–3]. The experimental results were analyzed using formulas from transition-radiation theory for targets having infinite transverse dimensions [6–8].

Studies [9, 10] of the diffraction radiation accompanying the passage of charges through an aperture and near screens have shown that for relativistic particles the order-of-magnitude transverse dimensions responsible for the formation of diffraction radiation at wavelength λ are determined by

$$\rho_{\text{eff}} \sim \lambda_\gamma \tag{1.1}$$

and may have macroscopic dimensions (γ is the electron Lorentz factor). This is because the transverse component of the Fourier component of the particle field is spatially bounded and nonzero in a circle of radius $\rho_{\text{eff}} \sim \lambda\gamma$. In this case, a particle will only effectively emit electromagnetic waves at wavelength λ when $\lambda\gamma > a$, where a is the distance between the particle and the edge of the screen (see § 31 in [9]).

The situation where large transverse distances are important in the interaction processes of high-energy particles is encountered in problems involving bremsstrahlung and the formation of electron–positron pairs in colliding beams at high energies. In [11–17] attention was drawn to the fact that for high-energy colliding e^+e^-, ep, and γp beams, impact parameters much larger than the transverse dimensions of the beams may play an important role in bremsstrahlung processes and the formation of e^+e^- pairs. This may lead to a decrease or an increase in the number of observable events compared with standard calculations in which this effect is not taken into account.

In [18] the present authors noted that a similar situation may arise for the transition radiation of an ultrarelativistic electron in targets of finite transverse dimensions. Specifically, for ultrarelativistic particles the order-of-magnitude transverse dimensions responsible for the transition radiation are determined by $\rho_{\text{eff}} \sim \lambda\gamma$ and may exceed the transverse dimension of the target itself. This situation in particular has been observed experimentally [1, 3]. It was shown in [18] that allowance for the finite macroscopic dimensions of the target appreciably distorts the spectrum of the transition radiation compared with the case when the transverse dimension of the target is infinite. The effect occurs not only for coherent transition radiation but also for the transition radiation of an isolated electron in a thin target. The analysis made in [18] referred to the simplest case when an electron passes through a thin layer of transparent material. The radiation was considered in the wave zone, i.e., at large distances from the target.

In the present study, we consider the transition radiation accompanying the propagation of a relativistic electron through a thin layer of material which completely reflects the particle field. We show that allowance for the transverse dimensions of the target appreciably distorts the radiation spectrum of the fast electrons in the wavelength range of the emitted photons $\lambda \geq a/\gamma$, where a is the characteristic transverse dimension of the target, compared with the case where the transverse dimension is infinite ($a \longrightarrow \infty$). This effect occurs for both the "forward" and "backward" radiation. The radiation is analyzed by placing a detector at various distances from the target (both smaller and larger than the

formation length of the radiation $l_c \sim 2\gamma^2\lambda$ [19]). We show that in this particular case the radiation consists of transition and diffraction radiation and also various forms of interference between these types of radiation. We consider the simplest case of the perpendicular propagation of a relativistic electron through the center of a thin disk of radius a.

2. FIELD FORMED AFTER PROPAGATION OF A RELATIVISTIC ELECTRON THROUGH A THIN DISK

We shall analyze the radiation formed when a relativistic electron propagates through the center of a thin disk of radius a consisting of a material which completely reflects the particle field (ideal conductor). We shall assume that the disk is perpendicular to the vector of the electron velocity. The disk is assumed to be fairly thin, i.e., $a_z \ll \lambda$, where a_z is the thickness of the disk and λ is the length of the emitted wave. It can be assumed that the field not incident on the target is diffracted at an infinitely thin disk of radius a positioned in the plane $z = 0$ [20]. After the electron has left the target, the electric field surrounding it may be written in the form

$$\mathbf{E}(\mathbf{r}, t) = \mathbf{E}^{(e)}(\mathbf{r}, t) + \mathbf{E}'(\mathbf{r}, t), \tag{2.1}$$

where $\mathbf{E}'(\mathbf{r}, t)$ is the radiation field, and $\mathbf{E}^{(e)}(\mathbf{r}, t)$ is the self-induced (Coulomb) field of an electron propagating uniformly at velocity $\mathbf{v}$ in vacuum. The Fourier component of the field $\mathbf{E}^{(e)}(\mathbf{r}, t)$ with respect to time is determined by

$$\mathbf{E}^{(e)}_\omega(\mathbf{r}) = \left(\frac{\boldsymbol{\rho}}{\rho}\frac{e\omega}{v^2\gamma}2K_1\left(\frac{\omega\rho}{v\gamma}\right) - i\frac{\mathbf{v}}{v}\frac{e\omega}{v\gamma^2}2K_0\left(\frac{\omega\rho}{v\gamma}\right)\right)\exp\left(\frac{i\omega z}{v}\right), \tag{2.2}$$

where e is the electron charge, $\boldsymbol{\rho}$ is the transverse coordinate, and K_0 and K_1 are the zeroth- and first-order Macdonald functions [21].

The self-induced field for a relativistic electron can be considered to be transverse relative to the velocity to within a factor of $1/\gamma$:

$$\mathbf{E}^{(e)}_\omega(\mathbf{r}) \approx \frac{\boldsymbol{\rho}}{\rho}\frac{e\omega}{v^2\gamma}2K_1\left(\frac{\omega\rho}{v\gamma}\right)\exp\left(\frac{i\omega z}{v}\right). \tag{2.3}$$

Since the electron field (2.3) incident on the disk was completely reflected by the target, immediately after the electron has left the target the field $\mathbf{E}'(\mathbf{r}, \omega)$ should completely screen its self-induced (Coulomb) field, i.e.,

$$\mathbf{E}^{(e)}_\omega(\mathbf{r}) + \mathbf{E}'_\omega(\mathbf{r}) = \theta(\rho - a)\mathbf{E}^{(e)}_\omega(\mathbf{r}), \quad z = 0, \tag{2.4}$$

where $\theta(x)$ is the Heaviside step function:

$$\theta(x) = 1, \quad \text{if} \quad x \geq 0 \quad \text{and} \quad \theta(x) = 0, \quad \text{if} \quad x < 0.$$

The field $\mathbf{E}'(\mathbf{r}, t)$ can be represented as a superposition of plane electromagnetic waves of frequency ω and wave vector $\mathbf{k}$:

$$\boldsymbol{E}'(\mathbf{r}, t) = \frac{1}{(2\pi)^4}\int d^3k\, d\omega \exp(i(\mathbf{k}\cdot\mathbf{r} - \omega t)) \times \mathbf{E}_{\mathbf{k},\omega}\delta\left(\frac{\omega^2}{c^2} - \mathrm{k}^2\right), \tag{2.5}$$

where $\delta(\omega^2/c^2 - \mathbf{k}^2)$ is the Dirac delta function, c is the velocity of light, and $\mathbf{E}_{\mathbf{k},\omega}$ are the expansion coefficients which are determined from the condition for screening of the field $\mathbf{E}^{(e)}(\mathbf{r}, t)$ by the field $\mathbf{E}'(\mathbf{r}, t)$ at $z = 0$:

$$\mathbf{E}'(\mathbf{r}, \omega) = -\theta(a - \rho)\mathbf{E}^{(a)}(\mathbf{r}, \omega), \quad z = 0. \tag{2.6}$$

Taking into account this condition, after integrating over k_z in (2.5), we obtain the following expression for the radiation field $\mathbf{E}'_\pm(\mathbf{r}, t)$ (the "+" sign corresponds to the forward radiation and the "–" sign to the backward radiation):

$$\mathbf{E}'_\pm(\mathbf{r}, t) = -\frac{ie}{v}\boldsymbol{\chi}\frac{\mathrm{Re}}{2\pi^2}\int d^2\chi\, d\omega \exp(i(\mathbf{k}\cdot\boldsymbol{\rho} - \omega t)) \times \frac{F}{\chi^2 + (\omega/v\gamma)^2}\exp\left(\pm iz\sqrt{\left(\frac{\omega}{c}\right)^2 - \chi^2}\right), \tag{2.7}$$

where

$$F = \left(\chi^2 + \left(\frac{\omega}{v\gamma}\right)^2\right)\frac{\omega}{v\gamma\chi}\int_0^a \rho\, d\rho\, J_1(\chi\rho)K_1\left(\frac{\omega\rho}{v\gamma}\right), \tag{2.8}$$

$\boldsymbol{\chi}$ is the transverse component of the wave vector $\mathbf{k}$ ($\mathbf{k} = \boldsymbol{\chi} + \mathbf{e}_z k_z$), and J_1 is a first-order Bessel function [21]. The relationship $\mathbf{E}_{\mathbf{k},\omega} = \mathbf{E}^*_{-\mathbf{k},-\omega}$ was used to derive (2.7) [6, 9].

3. INTENSITY OF ELECTRON TRANSITION RADIATION AT A DISK OF FINITE DIMENSIONS

We shall now consider an energy flux passing across a certain plane (detector) perpendicular to the particle trajectory and separated from the target (disk) by the distance z. We shall calculate this energy flux over the entire observation time. To this end we determine the projection of the Poynting vector $(\mathbf{E} \times \mathbf{H})_z$ on the z-axis and calculate the integral of this quantity over the surface xy and time t [6, 9]:

$$S_z = \frac{c}{4\pi}\int (\mathbf{E}\times\mathbf{H})_z\, dx\, du\, dt. \tag{3.1}$$

Here, we assume that the *xy* plane is infinite. Expanding the fields in (3.1) into Fourier integrals over *t* and *r* and expressing $\mathbf{H}_{\omega,\chi}$ in terms of $\mathbf{E}_{\omega,\chi}$, we obtain

$$S_z = \frac{c}{(2\pi)^4}\int\limits_{\omega>0} d\omega\int d^2\chi|\mathbf{E}_{\chi,\omega}(z)|^2. \tag{3.2}$$

Now using the relationship

$$\chi = \frac{\omega}{c}\sin\vartheta$$

to convert from the variable χ to the angle ϑ which determines the direction of emission, and noting that

$$d^2\chi = \frac{\omega^2}{c^2}d\Omega,$$

where $d\Omega = \sin\vartheta d\vartheta d\varphi$ is the solid-angle element in the direction of emission, we obtain the following expression for the spectral–angular density of the electromagnetic energy flux across the *xy* plane:

$$\frac{d^3S_z}{d\omega d\Omega} = \frac{1}{(2\pi)^4 c}\omega^2|\mathbf{E}_{\chi,\omega}(z)|^2, \quad \omega>0. \tag{3.3}$$

Thus, in order to find the spectral–angular density of the radiation (3.3), we need to know the Fourier component of the electric field $\mathbf{E}_{\chi,\omega}(z)$ in terms of the variables χ and ω for a fixed value of the coordinate *z*. Using formulas (2.7) and (2.1), we find that, after an electron has propagated through the disk,

$$\mathbf{E}_{\pm\chi,\omega}(z) = -\frac{4\pi ie}{v}\frac{\chi}{\chi^2+\left(\frac{\omega}{v\gamma}\right)^2} \times\left\{\exp\left(i\frac{\omega}{v}z\right) - F\exp\left(\pm i\frac{\omega}{c}z\cos\vartheta\right)\right\}. \tag{3.4}$$

Substituting this relationship into (3.3), we obtain

$$\frac{d^3S_z^+}{d\omega d\Omega} = \frac{e^2}{\pi^2 v}\frac{\sin^2\vartheta\cos^2\vartheta}{(\sin^2\vartheta+\gamma^{-2})^2} \times\left\{(1-F)^2 + 4F\sin^2\left[\frac{\omega z}{4v}(\gamma^{-2}+\sin^2\vartheta)\right]\right\}, \tag{3.5a}$$

$$\frac{d^3S_z^-}{d\omega d\Omega} = \frac{e^2}{\pi^2 v}\frac{\sin^2\vartheta\cos^2\vartheta}{(\sin^2\vartheta+\gamma^{-2})^2}F^2. \tag{3.5b}$$

The formulas (3.5) determine the spectral–angular distribution of the density of the electromagnetic energy flux across the plane (detector) positioned at a distance *z* from the target after propagation of a relativistic electron through a thin disk having finite transverse dimensions. The distance *z* can be either smaller or larger than the coherence length of the emission process $l_c \sim 2\gamma^2\lambda$. These formulas determine the intensity of the forward and backward radiation relative to the direction of the particle velocity vector. For the forward radiation, interference between the self-induced electron field and the radiation field is an important factor. The backward radiation is determined by the self-induced field of the impinging electron reflected from the surface of the target.

Note that a similar result for the spectral–angular density of the forward radiation was obtained in [22]. However, the function *F* obtained in [22] must be replaced by the function (2.8) in the present study. This is because in [22] the spectral–angular density of the radiation was calculated in terms of the vector potential of the electromagnetic field whereas for a target with sharp edges we need to use constraints at the target boundary for the electric field and induction vectors.

The function *F* appearing in (3.5) has simple asymptotic forms at low and high frequencies:

$$F(y,x)\approx\begin{cases}1, & x\geq 1\\ \dfrac{(xy)^2}{8}, & x\ll 1,\end{cases} \tag{3.6}$$

where

$$x = \frac{\omega}{\omega_\perp}, \quad \omega_\perp = \frac{\gamma v}{a}, \quad y = \gamma\sin\vartheta$$

(it is assumed that $y\approx 1$).

In the frequency range $\omega \gg c\gamma/a$ according to (3.6) we find $F \longrightarrow 1$ and the spectral–angular density of the radiation at the disk is the same as the result for an infinite plate ($a \longrightarrow \infty$). In this case, the influence of the transverse boundaries of the target on the transition radiation is insignificant and

$$\frac{d^3S_z^+}{d\omega d\Omega}\approx\frac{d^3S_z^{(\infty)}}{d\omega d\Omega}4\sin^2\left[\frac{\omega z}{4v}(\gamma^{-2}+\sin^2\vartheta)\right]. \tag{3.7}$$

Here, $d^3S_z^{(\infty)}/d\omega d\Omega$ is the spectral–angular density of the transition radiation of a fast electron propagating through a thin metal plate having infinite transverse dimensions when the detector is positioned at a large distance from the target [6, 9]:

$$\frac{d^3S_z^{(\infty)}}{d\omega d\Omega} = \frac{e^2}{\pi^2 v}\frac{\sin^2\vartheta\cos^2\vartheta}{(\sin^2\vartheta+\gamma^{-2})^2}. \tag{3.8}$$

In accordance with (3.7), the characteristic angles of emission $\vartheta_{\text{eff}} \sim 2\sqrt{c/\omega z}$ for $z \leq c\gamma^2/\omega$ are much larger than the characteristic angles of emission $\vartheta \approx 1/\gamma$ for $z \gg c\gamma^2/\omega$ [3]. In the range $z \ll c\gamma^2/\omega$, the argument of the sine in (3.7) is small compared with unity and formula (3.7) acquires the form [3, 23]

$$\frac{d^3S_z^+}{d\omega d\Omega}\approx\frac{d^3S_e}{d\omega d\Omega}\left[\frac{\omega z}{2v}(\gamma^{-2}+\sin^2\vartheta)\right]^2. \tag{3.9}$$

In this case, the density of the energy flux across the detector is much lower than the energy flux density associated with the self-induced electron field. This effect is attributed to interference between the self-induced electron field and the radiation field following the emission of an electron from a completely absorbing material.

For $z \longrightarrow \infty$, i.e., when the detector is positioned outside the radiation formation zone, the square of the sine in (3.5a) may be replaced by its average value of 1/2 and formula (3.5a) becomes

$$\frac{d^3S_z^+}{d\omega d\Omega} \approx \frac{d^3S_z^{(\infty)}}{d\omega d\Omega}\{1 + F^2\}. \tag{3.10}$$

The first term in this expression determines the spectral–angular density of the energy flux across the detector surface (infinite plane) produced by the self-induced (Coulomb) electron field. The second term determines the spectral–angular density of the radiation produced when an electron has propagated through a disk of radius a. If $\omega \ll c\gamma/a$, then $F \approx 1$ and the value of (3.10) is twice the spectral–angular density of the energy flux across the detector for the self-induced electron field. In this case, the boundaries of the disk are not important for the radiation and this is the same as that for a target with $a \longrightarrow \infty$. The characteristic angles of emission in this frequency range are $\vartheta_{\text{eff}} \sim \gamma^{-1}$. In the frequency range $\omega \ll c\gamma/a$, in accordance with (3.6), the intensity of the transition radiation at the disk [second term in (3.10)] is appreciably suppressed compared with the case $\omega > c\gamma/a$. For the backward radiation it can be seen from (3.5b) and (3.6) that the intensity of the transition radiation is also suppressed in the frequency range $\omega \ll c\gamma/a$.

For the case of forward radiation, formula (3.4) can be rewritten in the form

$$\mathbf{E}_{\pm\boldsymbol{\chi},\omega}(z) = -\frac{4\pi ie}{v}\frac{\boldsymbol{\chi}}{\chi^2 + \left(\frac{\omega}{v\gamma}\right)^2}\left\{\exp\left(i\frac{\omega}{v}z\right)\right.$$
$$\left. - \exp\left(i\frac{\omega}{c}z\cos\vartheta\right) + (1 - F)\exp\left(i\frac{\omega}{c}z\cos\vartheta\right)\right\}. \tag{3.11}$$

The first term in this expression is the self-induced electron Coulomb field, the second term is the transition-radiation field, and the third term is the field formed as a result of the diffraction of the impinging electron Coulomb field at a disk of finite dimensions. The last term in this expression can also be obtained from the Huygens principle if we consider the diffraction of an impinging electron field at an infinitely thin disk/screen having finite transverse dimensions.

It is therefore important to allow for the finite transverse dimensions of the target when analyzing the emission process of relativistic electrons in the frequency range $\omega \leq c\gamma/a$ since this leads to an appreciable difference between the spectral–angular density of the transition radiation compared with the result for a target having infinite transverse dimensions. The effect occurs for both the forward and the backward radiation. In experiments [1–3] where $\gamma \sim 200$ and $a \sim 5$ cm this constraint yields the following characteristic wavelengths of the emitted photons: $\lambda_0 \approx 1/\omega \geq 2.5 \times 10^{-2}$ cm for which the transverse dimensions of the target have a significant influence on the radiation.

We also note that we have considered the case where the entire energy flux produced by the electron emission in the target passes through the detector. In other words we assumed that the transverse dimensions of the detector are infinite. If this detector has finite transverse dimensions, this factor must also be taken into account when analyzing the emission process. This is because the transverse distances responsible for the formation of radiation at wavelength λ are of the order of magnitude of $\gamma\lambda$. As we have seen, these distances may have a macroscopic scale comparable to the size of the target. If $\rho_{\text{eff}} \sim \gamma\lambda$ is comparable to the transverse dimension of the detector, only part of the electromagnetic energy flux will enter the detector. For example, if the transverse dimension of the detector is small compared with the size of the target and the detector is positioned on the electron beam axis at the distance $z \ll c\gamma^2/\omega$, some of the radiation produced by diffraction of the particle field at the disk will not enter the detector. In this case, the detector will only record the energy flux produced by the formation of the particle field after the electron has left the target and the influence of the target boundaries on the radiation will be suppressed.

ACKNOWLEDGMENTS

This work was supported in part by the Ukrainian Foundation for Basic Research (project LPM effect) and by the MNOP (project QSU 082231).

REFERENCES

1. Y. Shibata, K. Ishi, T. Takahashi, *et al.*, Phys. Rev. A **44**, R3449 (1991).
2. U. Happek, A. J. Sievers, and E. B. Blum, Phys. Rev. Lett. **67**, 2962 (1991).
3. Y. Shibata *et al.*, Phys. Rev. E **49**, 785 (1994).
4. J. S. Nodvick and D. S. Saxon, Phys. Rev. **96**, 180 (1954).
5. F. C. Michel, Phys. Rev. Lett. **48**, 580 (1982).
6. G. M. Garibyan and Yan Shi, *Transient X-ray Radiation* (Izd. Akad. Nauk Arm. SSR, Yerevan, 1983).
7. V. L. Ginzburg and V. N. Tsytovich, *Transient Radiation and Transient Scattering* (Nauka, Moscow, 1984).
8. F. G. Bass and V. M. Yakovenko, Usp. Fiz. Nauk **86**, 189 (1965) [Sov. Phys. Usp. **8**, 420 (1965)].
9. M. L. Ter-Mikaelyan, *Medium Effects in High-Energy Electromagnetic Processes* (Izd. Akad. Nauk Arm. SSR, Yerevan, 1969) [Sov. Phys. Usp. **9**, 73 (1966)].

10. B. M. Bolotovskiĭ and G. V. Voskresenskiĭ, Usp. Fiz. Nauk **88**, 209 (1966).
11. V. N. Baĭer, V. M. Katkov, and V. M. Strakhovenko, Yad. Fiz. **36**, 163 (1982) [Sov. J. Nucl. Phys. **36**, 95 (1982)].
12. A. V. Burov and Ya. S. Derbenev, Preprint No. 82-07, Institute of Nuclear Physics, Siberian Division of USSR Academy of Sciences (Novosibirsk, 1982).
13. G. L. Kotkin, S. I. Polityko, and V. G. Serbo, Yad. Fiz. **42**, 692 (1985) [Sov. J. Nucl. Phys. **42**, 440 (1985)].
14. G. L. Kotkin, S. I. Polityko, and V. G. Serbo, Yad. Fiz. **42**, 925 (1985) [Sov. J. Nucl. Phys. **42**, 587 (1985)].
15. I. F. Ginzburg, G. L. Kotkin, S. I. Polityko, and V. G. Serbo, Pis'ma Zh. Éksp. Teor. Fiz. **55**, 614 (1992) [JETP Lett. **55**, 637 (1992)].
16. I. F. Ginzburg, G. L. Kotkin, S. I. Polityko, and V. G. Serbo, Yad. Fiz. **55**, 3310 (1992) [Sov. J. Nucl. Phys. **55**, 1847 (1992)].
17. I. F. Ginzburg, G. L. Kotkin, S. I. Polityko, and V. G. Serbo, Yad. Fiz. **55**, 3324 (1992) [Sov. J. Nucl. Phys. **55**, 1855 (1992)].
18. N. F. Shul'ga and S. N. Dobrovol'skiĭ, Pis'ma Zh. Éksp. Teor. Fiz. **65**, 581 (1997) [JETP Lett. **65**, 611 (1997)].
19. A. I. Akhiezer and N. F. Shul'ga, *High-Energy Electrodynamics in Matter* (Nauka, Moscow, 1993).
20. L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields* (Nauka, Moscow, 1973, 6th ed.; Pergamon Press, Oxford, 1975, 4th ed.).
21. H. B. Dwight, *Tables of Integrals and Other Mathematical Data* (Macmillan, London, 1961, 4th ed.; Nauka, Moscow, 1973).
22. N. F. Shul'ga and S. N. Dobrovol'sky, Nucl. Instrum. Methods Phys. Res., Sect. B **145**, 180 (1998).
23. Luke C. L. Yuan, C. L. Wang, and H. Uto, Phys. Rev. Lett. **25**, 1513 (1970).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 4, 2000, pp. 584–591.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 4, 2000, pp. 673–681.

ATOMS, SPECTRA, RADIATION

Stochastic Correlation Tomography

M. G. Karimov
Dagestan State University, Makhachkala, 367025 Dagestan, Russia
e-mail: karmaggas@mail.ru
Received June 7, 1999

Abstract—An analysis is made of the possibility of using stochastic generation of correlated quanta (random in time) to obtain rapid volume information on the state of a medium and to develop the physical principles of a real-time tomograph. Flux theory and mathematical modeling with a computer experiment are successfully used for these investigations. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The application of recent developments in laser physics and, in particular, the application of nonlinear optical diagnostic methods [1] giving the highest possible spatial and temporal resolution, determined by the diffraction limit of the focusing and the ultrashort (<10 fs) pulse duration [2, 3], to problems in tomography [4] quite clearly leads to a qualitatively new level of development of laser chronotomography. Increased interest in laser diagnostic methods is also being shown in the solution of problems in quantum tomography [5], i.e., inverse quantum-mechanics problems such as optical tomography and simplectic tomography [6–8]. Despite the various different schemes, problems of quantum tomography can be reduced to determining the density matrix (or Wigner function) for a marginal probability density distribution. In particular, the reconstruction of even and odd coherent states for an ion in a trap, achieved experimentally in [9], is of major importance for organizing the quantum calculations in the quantum computers currently being postulated. It is important to note that in quantum tomography the quantum state is determined using reversible integral transformations of the density matrix of the measurable probability distribution function (marginal distribution) by analogy with the inverse Radon problem. The methodology of quantum tomography will evidently be used to solve reconstructive problems in quantum electronics and laser physics. However, actual nonlinear processes in quantum electronics are only used to solve some problems in classical tomography [10–12]. As will be shown below, the phenomenon of correlated quantum generation, which occurs in many-quantum processes such as two-photon luminescence, parametric processes, and so on, can be used for reconstructive systems. Ultimately, the problem of determining the quantum flux parameter of a spatial point from measurements of the distribution in integral quantum fluxes needs to be solved. This semiclassical problem is solved using flux theory [13]. The scheme being discussed is called stochastic correlation tomography, since it is based on the property that the correlated-quantum generation events are random in time and with respect to the coordinate, i.e., the randomness of the events. Theoretical analyses and computer modeling show that this scheme is effective in positron tomography [14] if the generation of gamma quanta accompanying electron–positron annihilation is considered as a stochastic process of correlated-quantum generation. The correspondence between stochastic correlation tomography and the inverse Radon problem, and, in particular, chronotomography, is discussed.

2. FLUX THEORY AS THE BASIS FOR DESCRIBING EMITTING MEDIA

It is assumed that any point in an emitting medium having the coordinate r is a source of correlated quanta generated as a result of many-photon processes which are random in time and propagate in all possible directions. Assuming that the process of quantum generation in time at point r is a sequence $\Phi(r, t)$ of random events of the same type and the generation number at this point is denoted by $N(r, t)$, this process is determined by the analytic expressions

$$\begin{aligned} \Phi(r, t) &= \sum_{l}^{\infty} \delta(t - t_l(r)), \\ N(r, t_N) &= \int_{-\infty}^{t_N} \Phi(r, t)dt + 1, \end{aligned} \tag{1}$$

where $t_l(r)$ are the times of quantum generation at point r. The time interval between successive events, $T(r) = t_{l+1}(r) - t_l(r)$ is a random continuous flux variable. In particular, if the flux of random quantum generations is steady-state, with no aftereffect (Markov) and ordinary, this flux is the simplest flux [13, 15] having the exponential distribution

$$W(r, t) = \lambda(r)\exp[-t\lambda(r)], \tag{2}$$

where t is the instantaneous variable of the random continuous interval $T(r)$ and the intensity of the generation flux at point r is determined using the average interval between quantum generations, i.e., $\lambda(r) = 1/\bar{T}(r)$. In particular, we note that the flux formed as a result of the superposition of the simplest fluxes generated by various isolated generation sources and propagating in the same direction, will also be the simplest flux and the intensity of this quantum flux is determined by the intensities of the fluxes created by all isolated sources distributed on the straight line L, i.e.,

$$\lambda = \int_{-L/2}^{L/2} \lambda(r)dr.$$

The general system used for the measurements, as shown in Fig. 1, assumes that an emitting medium bounded in space with the linear dimension L is the object being studied, where the intensity $\lambda(r)$ of the generation flux is proportional to the density $\rho(x, y)$ of the actively emitting medium. In the present study, we merely note that correlated quanta are generated and do not consider the reason for this generation. The important thing is that fluxes of correlated quanta are generated and these can be located in a moving coordinate system (s, r) for various angles of rotation φ.

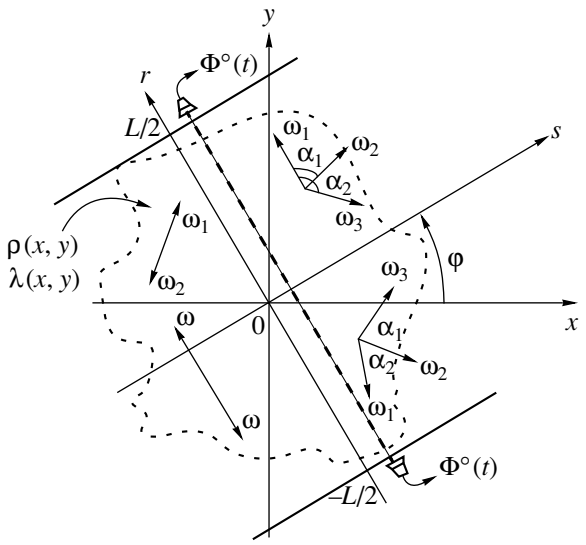

Fig. 1. Schematic diagram of experiment to study the correlated quantum-coincidence method for the two-dimensional case: α_i are the angles between the directions of emission of quanta at frequencies ω_i, L are the linear dimensions of the operative field of measurement, φ is the angle of rotation of the coordinates (s, r) relative to the fixed coordinates (x, y), and $\Phi^\circ(t)$ is the sum quantum flux generated by all points in the emitting medium having the density $\rho(x, y)$ positioned on a line in the direction of measurement.

3. INVERSE PROBLEM, METHOD OF CORRELATED QUANTUM COINCIDENCE

Of practical interest is the problem of determining the distribution laws of a quantum flux created by isolated sources under conditions where the initial quantum source is a sum flux and its distribution laws are known. This is the inverse problem of tomography using flux theory, which should ultimately determine the distribution density of the emission activity $\rho(x, y)$. The solution of this problem can be simplified appreciably, although this is not of fundamental importance, if it is assumed that the correlated quanta propagate in opposite directions at the velocity c. In order to separate the flux of an isolated point having the coordinate r, we use the principle of correlated quantum coincidence where one of the two fluxes is rarefied assuming that its distribution density $W(t)$ is known. For this case, the expressions for the sum fluxes as given by (1) have the following form:

$$\Phi^{\ominus}(t) = \sum_{i=1}^{\infty} \delta(t - t_i^{\ominus}),$$
$$\Phi^{\oplus}(t + 2\tau) = \sum_{j=1}^{\infty} \delta(t - (t_j^{\oplus} - 2\tau)), \quad (3)$$

where $\tau \in [-\tau_0/2, +\tau_0/2]$ is the delay time variable, $\tau_0 = L/c$, and $t_i^{\ominus}$ and $t_j^{\oplus}$ are the quantum fixing times to the left and right of the operative space, respectively. This is equivalent to the formation of a new flux $\Phi_X(t)$ consisting of quanta extracted from the fluxes $\Phi^{\ominus}(t)$ and $\Phi^{\oplus}(t)$ for which the condition $t_i^{\ominus} - t_j^{\oplus} = 2\tau$ is satisfied, as shown in Fig. 2. Using probability theory [13], we can establish that the average period of this flux is equal to the average period of the flux created by the point having the coordinate r, i.e., $\bar{T}_X = \bar{T}(r) = 1/\lambda(r)$. However, establishing the distribution laws for a synthesized flux for the general case may be represented as the problem of determining the total probability of all possible coincidences of the moments of the two principal initial fluxes $\Phi^{\ominus}(t)$ and $\Phi^{\oplus}(t)$. The flux thus synthesized is the set of event coincidences obtained when comparing an infinite number of Erlang fluxes [13, 16] obtained from an initial flux and the distribution density of the new flux is given by

$$W_X(t) = \sum_{s=1}^{\infty} Pq^{s-1} W_s(t), \quad (4)$$

where $P = P(r) = \lambda(r)/\lambda$, $q = 1 - P$, and the coefficient Pq^{s-1} defines the contribution of the sth-order Erlang flux having the distribution density $W_s(t)$ to the distribution density of the synthesized flux $\Phi_X(t)$. Hence, the problem of determining the distribution density of the synthesized flux $W_X(t)$ can be reduced to determining

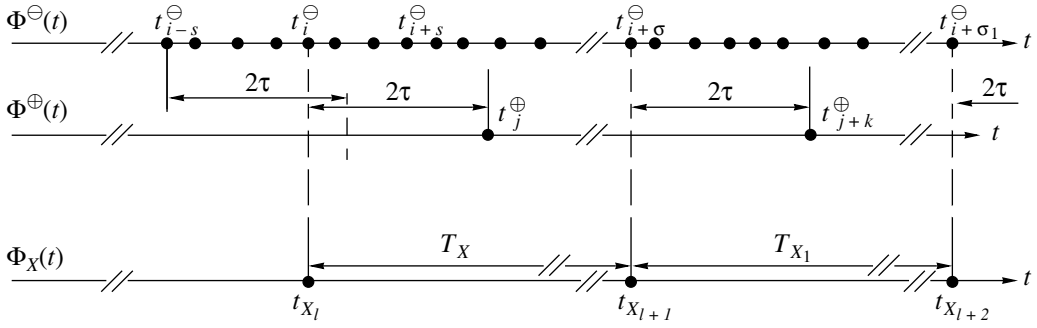

Fig. 2. Principle of the formation of the flux $\Phi_X(t)$ from the two fixed fluxes $\Phi^{\ominus}(t)$ and $\Phi^{\oplus}(t)$ by the correlated quantum-coincidence method. The duration of the random time interval $T_{X_l} = t_{X_{l+1}} - t_{X_l}$ is defined as the difference between the successive times t_X and $t_{X_{l+1}}$. The times t_{X_l} are selected from points in the sequence $\{t_i^{\ominus}\}$ for which points exist in the sequence $\{t_j^{\oplus}\}$ separated from the first by exactly 2τ, where $\tau = r/c$, and the points $t_i^{\ominus}$ and $t_j^{\oplus}$ are the quantum fixing times at the ends of the observation line.

the distribution densities of the Erlang fluxes $W_s(t)$. The distribution density of the Erlang fluxes is related to the distribution density of the initial fluxes by the recurrence formula [16]

$$W_s(t) = \frac{d}{dt}\int_0^t f_{s-1}(t-\tau)W(\tau)d\tau, \tag{5}$$

where $f_0(t) = 1$ and $f_1(t) = f(t)$ is the distribution function of the initial flux. Determining the distribution density can be simplified by using Hartley transformations [17] (symmetric and real-valued), i.e.,

$$\begin{aligned}\Theta(\nu) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} W(t)\mathrm{cas}(\nu t)dt,\\ W(t) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\Theta(\nu)\mathrm{cas}(\nu t)d\nu,\end{aligned} \tag{6}$$

where the function $\mathrm{cas}(\nu t) = \cos(\nu t) + \sin(\nu t)$. In accordance with (5) and (6), we can obtain another recurrence formula corresponding to expression (5):

$$\Theta_s(\nu) = \Theta_{s-1}(\nu)E(\nu) + \Theta_{s-1}(-\nu)O(\nu), \tag{7}$$

where

$$E(\nu) = \frac{1}{2}[\Theta(\nu) + \Theta(-\nu)],$$

$$O(\nu) = \frac{1}{2}[\Theta(\nu) - \Theta(-\nu)],$$

and $\Theta_s(\nu)$ is the Hartley transform of the distribution density of the sth-order Erlang flux. Taking into account expressions (7) and (4), after simple mathematical transformations we can easily obtain an expression for the Hartley transform of the synthesized flux:

$$\begin{aligned}\Theta_X(\nu) &= \sum_{s=1}^{\infty} Pq^{s-1}\\ &\times[\Theta_{s-1}(\nu)E(\nu) + \Theta_{s-1}(-\nu)O(\nu)].\end{aligned} \tag{8}$$

When calculating $\Theta_s(\nu)$, it is convenient to use various properties of the function $\mathrm{cas}(\nu t)$ such as

$$\begin{aligned}\int_{-\infty}^{\infty}\frac{dW(t)}{dt}\mathrm{cas}(\nu t)dt &= -\nu\Theta(-\nu),\\ \int_{-\infty}^{\infty}\left(\int W(t)dt\right)\mathrm{cas}(\nu t)dt &= \frac{1}{\nu}\Theta(-\nu),\end{aligned} \tag{9}$$

which allows us to calculate the sum (8) and obtain the simpler expression:

$$\begin{aligned}\Theta_X(\nu) &= P\\ &\times\frac{\Theta(\nu) - \dfrac{q}{2}[\Theta^2(\nu) + \Theta^2(-\nu)]}{1 - q[\Theta(\nu) + \Theta(-\nu)] + \dfrac{q^2}{2}[\Theta^2(\nu) + \Theta^2(-\nu)]},\end{aligned} \tag{10}$$

which establishes the relationship between the initial and synthesized fluxes. Using this expression (10), we can directly calculate the Hartley transform $\Theta_X(\nu)$ of the flux obtained by the correlated quantum-coincidence method if the Hartley transform $\Theta(\nu)$ of the initial flux is known. For the density of a Poisson distribution, in accordance with expression (2) and the definition (6) of

the Hartley transformation, the corresponding Hartley transform is given by

$$\Theta(\nu) = \frac{\lambda(\lambda + \nu)}{\lambda^2 + \nu^2}. \quad (11)$$

Then, in accordance with (10) and (11), the expression for the Hartley transform of the synthesized flux will have the following form:

$$\Theta_X(\nu) = \frac{P\lambda(P\lambda + \nu)}{(P\lambda)^2 + \nu^2}, \quad (12)$$

which corresponds to the distribution density

$$W_X(t) = P\lambda\exp(-P\lambda t) \quad (13)$$

based on the property of single-valued correspondence between the direct and inverse Hartley transformations. Assuming that the parameter is $P = P(r) = \lambda(r)/\lambda$, expression (13) may be given as follows:

$$W_X(t) = \lambda(r)\exp[-\lambda(r)t] = W(r, t), \quad (14)$$

which shows that the density of the flux synthesized by the correlated quantum-coincidence method corresponds to the distribution density of the flux created by a source of quantum generation positioned at a point having the coordinate r and that this flux is the simplest. Consequently, by sweeping the delay time τ of the initial fluxes $\Phi^\circ(t)$relative to one another in the range $-\tau_0/2$ to $\tau_0/2$, we can organize flux filtering for all points in the operative space L. The proposed algorithm can form a completely different and new basis for solving the reconstructive problem and specifically for the stochastic reconstructive problem of tomography. Thus, in accordance with the result (14), the analytic formula for the flux synthesized by the correlated quantum-coincidence method regardless of the type of distribution of the stochastic fluxes $\Phi^\circ(t)$ may be represented as the logical product of the sum fluxes (3):

$$\Phi_X(t) = \Phi(r, t) = \Phi^{\ominus}(t)\&\Phi^{\oplus}(t + 2\tau),$$
$$P_X = P(r).$$

However, this formula holds theoretically, i.e., only for stochastic (continuous) fluxes and only for the case of fixing with infinitely high accuracy.

4. CORRELATED QUANTUM-COINCIDENCE METHOD FOR REAL (DISCRETE) FLUXES

For real measurements using the principle of correlated quantum coincidence we must bear in mind that the fixing of quanta in fluxes does not take place instantaneously but with finite accuracy and over a specific observation time. Assuming that the observation time ($\Delta\tau$) is equal to the resolution of the time measurements, in accordance with the definition of the general probability P^0 [16, 18] of n independent incompatible generation sources each having the probability $P(r_k)$, the probability of fixing quanta (events) in a sum flux at an arbitrary discrete time is given by

$$P^0 = 1 - \prod_{k=1}^{n}[1 - P(r_k)],$$

where $n = L/c\Delta\tau$ is the number of discrete points or sources on the line L. In this case, the intensity of the steady-state flux created by a discrete point having the coordinate r_k is given by

$$\lambda^T(r_k) = \int_{r_k - \Delta\tau c/2}^{r_k + \Delta\tau c/2} \lambda(r)dr.$$

Consequently, the parameter of the flux created by a discrete point will be

$$a^T(r_k) = \lambda^T(r_k)\Delta\tau.$$

For real measurements using the correlated quantum-coincidence method the process of reconstruction can be represented as a probabilistic process consisting of two independent random effects having the probabilities $(P^0)^2$ and $P(r_k)$ which are compatible. In this case, in accordance with [18], the probability of the simultaneous manifestation of an event at the ends of the line of observation is given by

$$P^{st}(r_k) = (P^0)^2 + [1 - (P^0)^2]P(r_k), \quad (15)$$

where $P(r_k)$ is the probability of the generation of correlated quanta at the point r_k, propagating in opposite directions, and P^0 is the total probability of quantum generation at all points located on the line of observation. Formula (15) shows that a constant pedestal $(P^0)^2$ is present for all the reconstructed points, it is determined by all the sources located on the line, and the following inequality holds:

$$P^{st}(r_k) - P(r_k) = [1 - P(r_k)](P^0)^2 \geq 0, \quad (16)$$

which indicates that the reconstruction process is multivalued since the condition $P^{st}(r_k) - P(r_k) = 0$ is only satisfied when $P^0 \equiv 0$ and/or $P(r_k) \equiv 1$. However, this implies that a total flux is absent and/or the flux created by the point having the coordinate r_k is a regular flux. However, both these regimes are "unworkable" since the fluxes do not possess stochastic properties, i.e., the fluxes are regular or completely absent. Assuming that the generated fluxes are the simplest, the generation probability is given by

$$P(r_k) = 1 - \exp(-a^T(r_k)), \quad (17)$$

i.e., the probability $P = P(r_k) = P(a^T(r_k))$ is only a function of $a^T(r_k)$ and $P^0 = P^0(a) = 1 - \exp(-a)$ is only a function of the total flux parameter a. Thus, on the basis of (15) and (17), the reconstruction probability

$$P^{st}(r_k) = P^{st}(a^T(r_k), a)$$

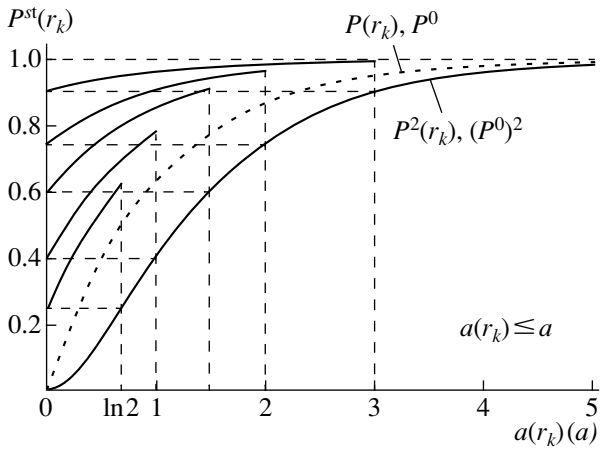

Fig. 3. Dependence of the probability $P^{st}(r_k)$ of the coincidence of quanta fixed by the correlated quantum-coincidence method on the parameters of the simplest fluxes $a(r_k)$ for various values of the summarized parameter a. For clarity graphs are also plotted for the generation probability $P(r_k)$ for the point r_k on the line of investigation and for the probability of the total quantum flux P^0 along this line.

is a function of two variables. Figure 3 shows how the probability depends on the flux parameters ($a^T(r_k)$, a). It should be noted that each curve of the reconstruction probability $P^{st}(r_k)$ has a bounded range of determination [0, a] and the state reconstructed "in front of" the point r_k depends on the "surroundings" of this point, i.e., on a, and the constant component is also entirely and completely determined by the summarized parameter a. It can be seen from (15) and (17) that for the points $r_k = r^0$ for which $a^T(r_k) \equiv 0$ and thus $P(r_k = r^0) \equiv 0$ the equality $(P^0)^2 = P^{st}(r^0)$ is valid. Then, using this fact to modify formula (15) and also expression (17) which establishes a single-valued correlation between the generation probability and the flux parameter, we obtain an expression for the true flux parameter of the point having the coordinate r_k:

$$a^T(r_k) = \ln\frac{1 - P^{st}(r^0)}{1 - P^{st}(r_k)}, \tag{18}$$

which is the basic formula for stochastic tomography, the formula for image reconstruction using the correlated quantum-coincidence method. Note that points specially selected as "instrumental" and known to be not positioned on the object may be taken as points with zero flux parameters or so-called reference points. Thus, the reconstructed image of the object is entirely and completely determined only by the stochastic tomographic image. Bearing in mind that the quantity which can directly be measured experimentally by the correlated quantum-coincidence method is the number of coincidences $N^{st}(r_k, l)$ (coincident photon counting regime) during the experiment, these data are used to calculate a new parameter

$$a^{st}(r_k, l) = \ln\frac{1 - N^{st}(r^0, l)/l}{1 - N^{st}(r_k, l)/l}, \tag{19}$$

which is the parameter determined by the correlated quantum-coincidence method for real quantum fluxes and will depend on the duration of the experiment (l). It is readily observed that the parameters calculated using formulas (18) and (19) are interrelated by $a^T(r_k) = \lim_{l \to \infty} a^{st}(r_k, l)$ provided that the limit exists:

$$P^{st}(r_k) = \lim_{l \to \infty}[N^{st}(r_k, l)/l]. \tag{20}$$

It should be noted that the parameter $a^{st}(r_k, l)$ is calculated using the number of quantum coincidences $N^{st}(r_k, l)$ using formula (19) and corresponds to its time delay τ.

5. COMPUTER MODELING

We can make a purely theoretical analysis of the correlated quantum-coincidence method and establish the system parameters for which the measurable flux parameter $a^{st}(r_k, l)$ will correspond to the flux parameter $a^T(r_k)$ or the true parameter $a(r_k)$. However, this problem is of independent interest. In particular, this problem was investigated by computer modeling at all stages, beginning with the generation of correlated quanta as far visualizing the measurable parameter $a^{st}(r_k, l)$. For the computational experiment, we developed a computer multichannel analyzer system with facilities for input and output of on-line two-dimensional information. The ideology behind the construction of multichannel systems is to visualize images obtained at all stages of the experiment with relevant reference information on the operating conditions of the program at a particular stage. Information on the experimental data is imaged using a specially developed operating protocol. Figure 4 gives the results of mathematical modeling and a computer experiment to reconstruct volume information at various stages using the correlated quantum-coincidence method. A microscopic section of the lungs (cross section) with various pathologies is taken as an example. The results are given as two-dimensional normalized "images" of the actively emitting medium and it is shown how these change as we systematically go through all (four) stages of modeling. Results of three experiments are given for various parameters of the total flux. At the first stage of the computer experiment Fig. 4 shows an anatomical atlas of the cross section (true image) while the second stage shows an image of the statistical activity of the cross section which is determined only by the number of generations $N^g = N^g(r_k, l)$ at each point over the duration of the experiment l. In fact, this image corresponds to the statistical manifestation of the cross

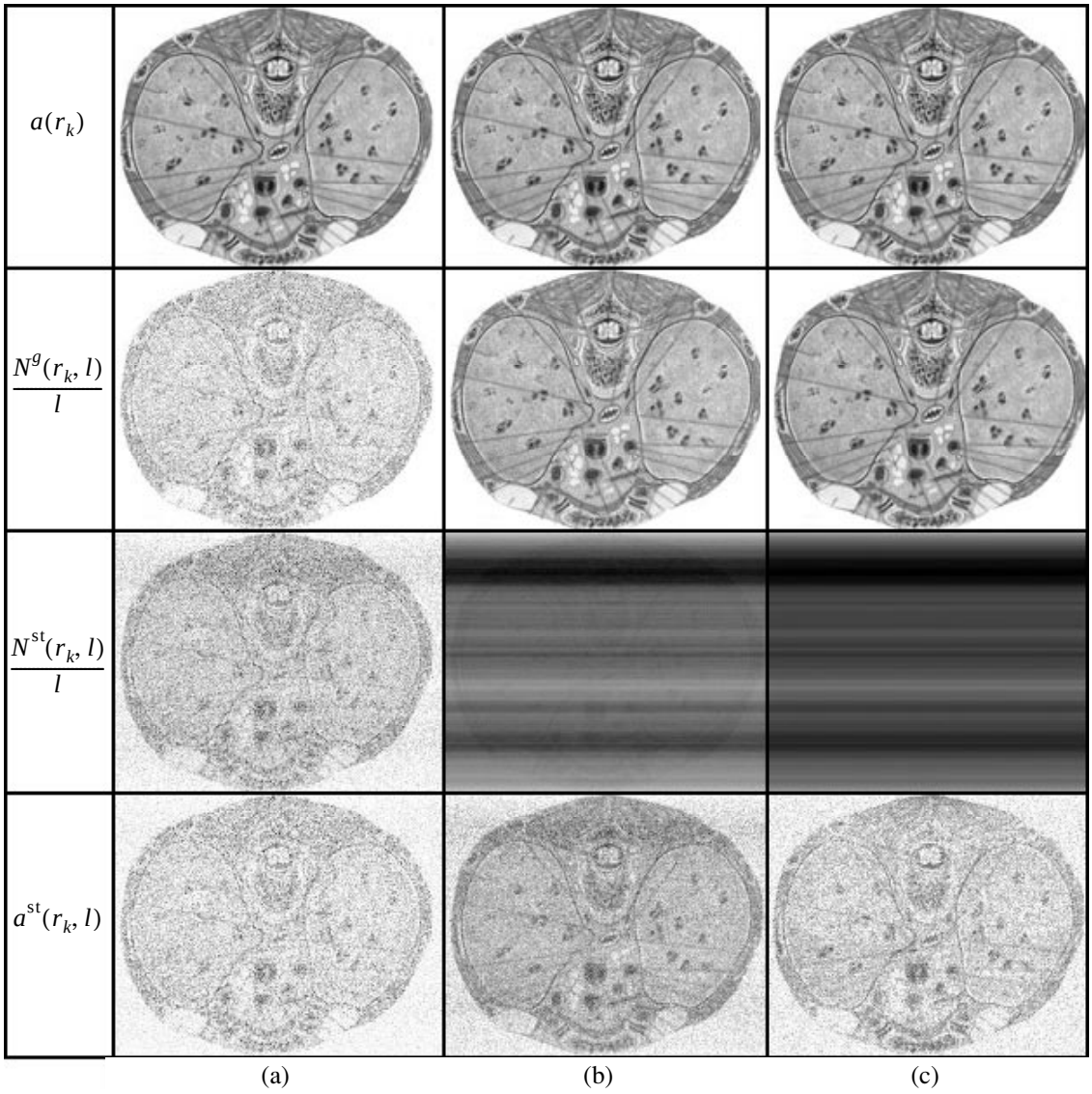

Fig. 4. Results of mathematical modeling and computer experiment using the coincidence method to reconstruct a two-dimensional "image" of an actively emitting medium for various parameters of the total flux. $N^g(r_k)$ is the number of generations, $N^{st}(r_k)$ is the number of coincidences for measurements by the correlated quantum-coincidence method, Π is the similarity parameter, $a^{st}(r_k, l)$ is the flux parameter estimated using formula (19): (a) $a = 1/512$, $l = 10^6$, $\Pi = 0.3440$; (b) $a = 1/10$, $l = 10^6$, $\Pi = 0.3560$; (c) $a = 1$, $l = 1.5 \times 10^6$, $\Pi = 0.2000$.

section. The image formed at the third stage of modeling is obtained directly by the correlated quantum-coincidence method and is determined by the number of quantum coincidences $N^{st}(r_k, l)$. In particular, for values of the total flux parameter of 0.1 and 1, the images obtained at the third stage are saturated, which confirms formula (15) and Fig. 3. The experiment shows that for values of the total flux parameter less than $1/n$, where $n = 512$ is the number of divisions, the image obtained at this stage corresponds to the true one and for values of 0.1 and 1 these are far from similar. Finally, the images $a^{st} = a^{st}(r_k, l)$ obtained at the fourth stage are formed entirely and completely from the image data obtained at the second stage using expression (19). The results of the computer experiment show that the agreement between the images obtained at this stage and the initial image is maintained for all three values of the total flux parameter. Then the similarity parameter $\Pi = \Pi(N^g/l, a^{st})$ is estimated for all cases where $a^{st} = a^{st}(r_k, l)$ which is one of the main measurable characteristics in classical tomography.

6. RECONSTRUCTIVE FORMULA FOR CHRONOTOMOGRAPHY

The results of this investigation of the correlated quantum-coincidence method, including the computer

modeling, suggest that the proposed method is in turn a method of chronometry which can be used for measurements along the longitudinal (volume) coordinate, i.e., according to the time-of-flight principle. Thus, it can be used as a chronometric method in tomography and in particular, the results of the modeling can be used to form the technical basis for developing the main components of a chronotomograph. In this case, the theory of chronotomography is constructed on the basis of the fact that the most common physical characteristic for chronometric systems, in particular for nonlinear optical diagnostic systems, is the instrumental function which is determined by a single and fundamental physical parameter and by the parameter of the measurement accuracy. Assuming that this system has a Gaussian instrumental function

$$h(\tau) = \frac{1}{\sigma\sqrt{\pi}}\exp\left(-\frac{\tau^2}{\sigma^2}\right),$$

where

$$\tau = \frac{r}{c}, \quad \sigma^2 = \frac{\Delta\tau^2}{\ln 2},$$

and $\Delta\tau$ is the accuracy of the time measurements, this (by definition) establishes the interrelation between the true and measured parameters using the contraction integral:

$$a^T(s, \tau) = \int_{-\infty}^{\infty} a(x', y')h(\tau - \tau')d\tau'. \qquad (21)$$

In fact, the solution of this integral equation is the solution of the chronotomography problem which was obtained using the Hartley transformation. After simple mathematical operations we can obtain an expression for the true value of the flux parameter $a(x, y)$ in terms of the measurable quantities $a^T(s, \tau, \varphi)$:

$$a(x, y) = \int_0^{\pi} d\varphi \int_{-\infty}^{\infty} ds' a^T(s', \tau, \varphi)Q(s - s'), \qquad (22)$$

which is the reconstructive formula for chronotomography where

$$Q(s) = \int_{-\infty}^{\infty} d\omega \frac{1}{H(\omega)}\mathrm{cas}(2\pi\omega s), \quad \omega = \frac{\nu}{2\pi},$$

$$H(\omega) = \int_{-\infty}^{\infty} d\tau h(\tau)\mathrm{cas}(2\pi\omega\tau).$$

This expression shows good agreement with the result of solving the reconstructive problem allowing for the finite rate of propagation of the volume information signal which was obtained directly in [19]. Formula (21) may be transformed to give a more convenient form without using the Hartley transform of the instrumental function $H(\omega)$:

$$a(x, y) = \pi\sigma^2 \times \int_0^{\pi} d\varphi \int_{-\infty}^{\infty} ds' a^T(s', \tau, \varphi)h(s - s')Q_0(s - s'), \qquad (23)$$

where

$$Q_0(s) = \int_{-\infty}^{\infty} d\omega|\omega|\mathrm{cas}(2\pi\omega s).$$

This formula is common to all laser chronotomographs for which the characteristic accuracy of the time measurements is $\Delta\tau$, including tomographs using nonlinear optical diagnostic methods. Formulas (21) and (22) are the reconstructive formulas for chronotomography which can be successfully used for computer chronotomographs since the method of calculating the contraction integrals in these formulas for computer tomography has been fairly well developed and studied [14].

7. CONCLUSIONS

By using probabilistic (stochastic) theory or flux theory to describe actively emitting media, a fundamentally new possibility for obtaining volume information by a coincidence method can be substantiated theoretically for continuous correlated fluxes and for real discrete fluxes. Mathematical modeling and a computer experiment have shown that this method can be used to develop a time-of-flight chronotomograph. In particular, expressions (19) and (22) which establish a relationship between the measured and true parameters, can be used to describe the volume nonlinear optical properties of the medium, i.e., for tomography.

ACKNOWLEDGMENTS

The author thanks Professor A. S. Chirkin of Moscow State University for valuable and useful advice which helped to make substantial improvements to the content of this article and its qualitative formulation.

REFERENCES

1. S. A. Akhmanov and N. I. Koroteev, *Nonlinear-Optics Methods in the Spectroscopy of Light Scattering* (Nauka, Moscow, 1981).
2. S. A. Akhmanov, V. A. Vysloukh, and A. S. Chirkin, *The Optics of Femtosecond Laser Pulses* (Nauka, Moscow, 1988).
3. *Ultrashort Light Pulses*, Ed. by S. Shapiro (Springer, Heidelberg, 1977; Mir, Moscow, 1980).
4. J. Radon, Ber. Verh. Saechs. Akad. Wiss. Leipzig, Math.–Phys. Kl. **69**, 262 (1917).
5. V. I. Man'ko and O. V. Man'ko, Zh. Éksp. Teor. Fiz. **112**, 796 (1997) [JETP **85**, 430 (1997)].

6. S. Walentowitz and W. Vogel, Phys. Rev. A **53**, 4528 (1996).
7. K. Banaszek and K. Wodkiewicz, Phys. Rev. Lett. **76**, 4344 (1996).
8. R. L. de Matos Filho and W. Vogel, Phys. Rev. A **54**, 4560 (1996).
9. O. M. Meekhof, G. Monroe, B. E. King, *et al.*, Phys. Rev. Lett. **76**, 1796 (1996).
10. M. R. Hee, J. A. Izatt, E. A. Swanson, and J. G. Fujimoto, Opt. Lett. **18**, 1107 (1993).
11. L. H. Wang, P. P. Ho, C. Lui, *et al.*, Science **259**, 769 (1993).
12. M. A. O'Leary, D. A. Boas, B. Chance, and A. G. Yodh, Proc. SPIE **1888**, 334 (1993).
13. Yu. P. Pyt'ev and I. A. Shishmarev, *Probability Theory and Mathematical Statistics for Physicists* (Mosk. Gos. Univ., Moscow, 1983).
14. F. Natterer, *The Mathematics of Computerised Tomography* (Wiley, Chichester, 1986; Mir, Moscow, 1990),
15. S. A. Akhmanov, Yu. E. D'yakov, and A. S. Chirkin, *Introduction to Statistical Radio Physics and Optics* (Nauka, Moscow, 1981).
16. B. V. Gnedenko and I. N. Kovalenko, in *Introduction to the Queueing Theory* (Nauka, Moscow, 1987), p. 336.
17. R. Bracewell, *The Hartley Transform* (Oxford Univ. Press, Oxford, 1986; Mir, Moscow, 1990).
18. G. A. Korn and T. M. Korn, in *Mathematical Handbook for Scientists and Engineers* (McGraw-Hill, New York, 1961; Nauka, Moscow, 1974), p. 832.
19. M. G. Karimov and A. A. Aliverdiev, Vestn. Mosk. Univ., Ser. 3: Fiz., Astron. **15**(4), 35 (1996).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 4, 2000, pp. 592–599.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 4, 2000, pp. 682–690.

ATOMS, SPECTRA, RADIATION

Quantum–Classical Correspondence and Nonclassical State Generation in Dissipative Quantum Optical Systems¶

K. N. Alekseev[1, *], N. V. Alekseeva[1], and J. Peřina[2, **]

[1] *Kirensky Institute of Physics of Russian Academy of Sciences, Krasnoyarsk, 660036 Russia*

[2] *Department of Optics and Joint Laboratory of Optics, Palacky University, 77207 Olomouc, Czech Republic*

** e-mail: kna@tnp.krascience.rssi.ru*

*** e-mail: perina@optw.upol.cz*

Received September 3, 1999

Abstract—We develop a semiclassical method to determine the nonlinear dynamics of dissipative quantum optical systems in the limit of large number of photons N; it is based on the $1/N$-expansion and the quantum–classical correspondence. The method is used to tackle two problems: the study of the dynamics of nonclassical state generation in higher order anharmonic dissipative oscillators and the establishment of the difference between the quantum and classical dynamics of the second-harmonic generation in a self-pulsing regime. In addressing the first problem, we obtain an explicit time dependence of the squeezing and the Fano factor for an arbitrary degree of anharmonism in the short-time approximation. For the second problem, we analytically find a characteristic time scale at which the quantum dynamics differs insignificantly from the classical one. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The situation when nonlinear interactions involve a large number of photons, N, is quite typical of many problems in quantum and nonlinear optics [1–3]. Heidmann *et al.* suggested [4] the use of the $1/N$-expansion method [5] to describe the nonlinear dynamics of the mean values and second-order cumulants of a quantum system in the $N \gg 1$ limit. Following the general scheme of that method [5], an exact or approximate solution can be found in terms of the coherent state representation in the classical limit as $N \longrightarrow \infty$ and can then be adjusted by adding the quantum corrections. The method proves to be particularly convenient when the dynamics of nonclassical state generation must be determined [4]. We have recently developed the method further to study the enhanced squeezing at the transition to quantum chaos [6–8].

Papers [4, 6, 7] are concerned with the problems of nondissipative quantum systems only. In this paper, we extend the method to dissipative quantum systems. For quantum systems without dissipation, the lowest order of the $1/N$-expansion is equivalent to the linearization in terms of the classical solution [6, 7], whereas in dissipative systems, as is demonstrated in what follows, the solution of the equations of motion for variations near the classical trajectory cannot provide complete information on the dynamics of quantum fluctuations even in the lowest order of $1/N$. We show that the influence of the reservoir on the dynamics of expectation values and dispersions, which is different from the energy dissipation, always exists: It has the quantum nature and cannot be neglected even in the semiclassical limit. However, specific manifestations of the effect depend on the type of the attractor in the underlying classical dynamic system. For systems with a simple attractor in the classical limit, the "quantum diffusion" associated with the quantum fluctuations of the reservoir do not lead to any new physical effects in the dynamics of the main system, at least in the short-time limit. For a stable limit cycle, on the other hand, such a diffusion appears to be the main mechanism responsible for the difference between the classical and quantum dynamics for $N \gg 1$.

Along with the presentation of a general formalism, we consider two typical examples of quantum optical systems with a simple attractor and a stable limit cycle in the classical limit as $N \longrightarrow \infty$: the dissipative higher order anharmonic oscillator and the self-pulsing regime of intracavity second-harmonic generation (SHG). We show how the $1/N$-expansion method can be used to investigate the dynamics of the nonclassical state generation and to determine the time scale for a correct classical description of the dissipative quantum dynamics.

The quantum anharmonic oscillator with a Kerr-type nonlinearity is one of the simplest and most popular models used in the description of quantum statistical properties of light interacting with a nonlinear medium [1, 9]. The Kerr oscillator model with a third-order nonlinearity yields an exact solution in both the nondissipative [10] and dissipative limits [9]. However, because of the complexity of the solution in the dissipative case, numerical methods or special approximate analytic

¶ This article was submitted by authors in English.

methods must be used to determine statistical properties of the radiation in the most relevant experimental case involving a large number of photons. Moreover, there are no exact solutions available for the model of the anharmonic oscillator with a higher order nonlinearity.

In this paper, we analytically obtain a simple and explicit time dependence of the degree of squeezing and the Fano factor in the anharmonic oscillator model of an arbitrary order for the most interesting experimental situation featuring higher intensities ($N \gg 1$) and short-time interactions. As another example of application of the $1/N$-expansion, we consider the self-pulsing in SHG [11]. Such an oscillatory regime corresponding to the limit cycle was observed experimentally in [12]. There are several papers dealing with the development of approximate analytic and numerical methods with the purpose of describing different dynamical regimes in SHG in terms of quantum mechanics [13–17]. In particular, Savage [14] calculated the Gaussian approximation of the Q distribution function about the classical limit cycle. He demonstrated numerically that in the classical limit, the initial rapid collapse of the Q distribution in the neighborhood of the limit cycle is followed by the diffusion around the limit cycle. However, the author did not offer any analytical solution of the problem or an explanation of the physics of the effect observed.

In this paper, we show that the diffusion around the classical limit cycle can be obtained as a solution of the equations of motion for low-order cumulants by using the $1/N$-expansion technique. This enables us to find the time scale $t \ll t^*$ with $t^* \simeq 2N\gamma^{-1}$ (where γ is a damping constant) for a correct classical description of self-oscillations in SHG. The resultant estimate is consistent with that obtained for t^* numerically in [14]. Finally, we interpret the quantum diffusion around the limit cycle as a diffusion caused by the effect of the reservoir vacuum on the SHG dynamics.

This paper is organized as follows. In Section 2, we describe a general formalism of the $1/N$-expansion applicable to an arbitrary single-mode quantum dissipative system and present the solution of the equations of motion for mean values and second-order cumulants obtained in the first order of $1/N$. In Sections 3 and 4, we deal with the nonclassical state generation dynamics in higher order anharmonic oscillators and the quantum–classical correspondence for the self-pulsing regime in SHG, respectively. The final section contains a summary and concluding remarks.

2. $1/N$-EXPANSION AND QUANTUM–CLASSICAL CORRESPONDENCE

We begin with generalizing the approach of [7] systems with dissipation. As an illustrative example, we consider a quantum anharmonic oscillator with the Hamiltonian in the interaction picture

$$H = \Delta b^\dagger b + \frac{\lambda_l}{l+1}(b^\dagger b)^{l+1}, \quad [b, b^\dagger] = 1, \tag{1}$$

where the operators b and $b^\dagger$ describe a single quantum field mode and the constant λ_l is proportional to a $(2l+1)$-order nonlinear susceptibility of a nonlinear medium (l is an integer), Δ is the light frequency detuning from the characteristic quantum transition frequency, and $\hbar \equiv 1$. Everywhere in this paper, we use the normal ordering of operators. The oscillator interacts with an infinite linear reservoir at a finite temperature. The Hamiltonians of the reservoir and of the oscillator-reservoir interaction are defined as

$$\begin{aligned} H_r &= \sum_j \psi_j (d_j^\dagger d_j + 1/2), \\ H_{int} &= \sum_j (\kappa_j d_j b^\dagger + \text{H.c.}), \end{aligned} \tag{2}$$

where the Bose operator d_j ($[d_j, d_k^\dagger] = \delta_{jk}$) describes an infinite reservoir with the characteristic frequencies ψ_j and κ_j are the coupling constants between reservoir modes and the oscillator. We introduce new scaled operators $a = b/N^{1/2}$ and $c_j = d_j/N^{1/2}$ and their Hermitian conjugates satisfying the commutation relations

$$[a, a^\dagger] = 1/N, \quad [c_j, c_k^\dagger] = \delta_{jk}/N. \tag{3}$$

In the classical limit as $N \longrightarrow \infty$, we obtain commuting classical c-numbers instead of operators. The full Hamiltonian

$$H = H_0 + H_r + H_{\text{int}}$$

can be rewritten as

$$H = N\mathcal{H},$$

where $\mathcal{H}$ is as in (1) and (2) but with the replacements

$$\begin{aligned} &b \longrightarrow a, \quad b^\dagger \longrightarrow a^\dagger, \quad d_j \longrightarrow c_j, \\ &d_j^\dagger \longrightarrow c_j^\dagger, \quad \lambda_l \longrightarrow g_l(N) \equiv \lambda N^l. \end{aligned} \tag{4}$$

It can be shown that the photon-number dependent constant $g_l(N)$ provides a correct time scale of oscillations for nonlinear oscillator (1) in the classical limit (for the Kerr nonlinearity with $l = 1$, see, e.g., [18]). We note that $\mathcal{H}$ can have an explicit time dependence in the general case [7]. Within a standard Heisenberg–Langevin approach, the equation of motion has the form ([1], Chap. 7)

$$\dot{a} = -i\left(\Delta - i\frac{\gamma}{2}\right)a + V + L(t), \tag{5}$$

where $V = \partial\mathcal{H}_0/\partial a^\dagger$, $\gamma = 2\pi|\kappa(\omega)|^2\rho(\omega)$ is the damping constant, with $\rho(\omega)$ being the density function of reservoir oscillators whose spectrum is considered to be flat. The Langevin force operator $L(t)$ is in a standard relation to the operators $\{c_j\}$ of the reservoir [1]. In our notation (4), the properties of $L(t)$ [1] can be rewritten as

$$\begin{aligned} \langle L(t)\rangle_R &= \langle L^\dagger(t)\rangle_R = 0, \\ \langle L^\dagger a\rangle_R + \langle a^\dagger L\rangle_R &= \gamma\frac{\langle n_d\rangle}{N}, \\ \langle La\rangle_R + \langle aL\rangle_R &= 0. \end{aligned} \tag{6}$$

Here, the averaging is performed over the reservoir variables and $\langle n_d\rangle$ is a single-mode mean number of the reservoir quanta (phonons) related to temperature T as

$$\langle n_d\rangle = \left[\exp\left(\frac{\omega}{kT}\right) - 1\right]^{-1},$$

where k is the Boltzmann constant and ω is the characteristic phonon frequency. From the Heisenberg–Langevin equations for a, a^2 and the Hermitian conjugated equations, using (5) and (6), we obtain

$$\begin{aligned} i\frac{d}{dt}\langle\alpha\rangle &= \langle V\rangle - i\frac{\gamma}{2}\langle\alpha\rangle, \\ i\frac{d}{dt}\langle(\delta\alpha)^2\rangle &= 2\langle V\delta\alpha\rangle + \langle W\rangle - i\gamma\langle(\delta\alpha)^2\rangle, \\ i\frac{d}{dt}\langle\delta\alpha^*\delta\alpha\rangle &= -\langle V^*\delta\alpha\rangle + \langle\delta\alpha^* V\rangle \\ &\quad - i\gamma\langle\delta\alpha^*\delta\alpha\rangle + i\gamma\frac{\langle n_d\rangle}{N}, \end{aligned} \tag{7}$$

where

$$\begin{aligned} W &= (1/N)\partial V/\partial a^\dagger, \quad z \equiv \langle a\rangle, \\ \langle(\delta\alpha)^2\rangle &= \langle a^2\rangle - z^2, \quad \langle\delta\alpha^*\delta\alpha\rangle = \langle a^\dagger a\rangle - |z|^2, \end{aligned}$$

and the averaging is performed over both the reservoir variables and the coherent state

$$|\alpha\rangle = \exp(N\alpha a^\dagger - N\alpha^* a)|0\rangle$$

corresponding to the mean photon number $\simeq N$. In deriving (7), we neglect the insignificant additional detuning introduced to Δ by the interaction with the reservoir [1]. In the absence of damping, $\gamma = 0$, our equations for the mean values and the second-order cumulants (7) are reduced to the corresponding equations in [4, 7].

The set of equations (7) is not closed and is basically equivalent to the infinite dynamical hierarchy system for the cumulants of a different order. To truncate the system to the second-order cumulants, we make the substitution $a \longrightarrow z + \delta\alpha$, where, at least initially, the mean value is $z \simeq 1$ and the quantum correction are

$$|\delta\alpha(t = 0)| \simeq N^{-1/2} \ll 1.$$

Using the Taylor expansion of the functions V and W and after some algebra analogous to that used in [7], we obtain from (7) in the first order of $1/N$ the following self-consistent system of equations for the mean value and the second-order cumulants (for details see [19]):

$$i\dot{z} = -i\frac{\gamma}{2}z + \langle V\rangle_z + \frac{1}{N}Q(z, z^*, C, C^*, B), \tag{8}$$

$$i\dot{C} = 2\left(\frac{\partial V}{\partial\alpha}\right)_z C + 2\left(\frac{\partial V}{\partial\alpha^*}\right)_z B - i\gamma C, \tag{9}$$

$$i\dot{B} = -\left(\frac{\partial V^*}{\partial\alpha}\right)_z C + \left(\frac{\partial V}{\partial\alpha^*}\right)_z C^* - i\gamma(B - B^{(0)}). \tag{10}$$

The corresponding equation for $C^*(t)$ can be obtained from (9) by complex conjugation. The quantum correction to the classical motion Q in (8) has the form

$$\begin{aligned} Q &= \frac{1}{2}\left(\frac{\partial^2 V}{\partial\alpha^2}\right)_z C + \frac{1}{2}\left(\frac{\partial^2 V}{\partial\alpha^{*2}}\right)_z C^* \\ &\quad + \left(\frac{\partial^2 V}{\partial\alpha^*\partial\alpha}\right)_z\left(B - \frac{1}{2}\right). \end{aligned} \tag{11}$$

In (8)–(11), the subscript z means that the values of V and its derivatives are calculated for the mean value z; we have introduced

$$B = N\langle\delta\alpha^*\delta\alpha\rangle + 1/2, \quad C = N\langle(\delta\alpha)^2\rangle. \tag{12}$$

The initial conditions for system (8)–(10) are of the form

$$B(0) = 1/2, \quad C(0) = 0, \tag{13}$$

and an arbitrary $z(0) \equiv z_0$ which is of the order of unity. The equilibrium value of the cumulant B in (10) is determined by the mean number of the reservoir quanta and its zero-point energy as

$$B^{(0)} = \langle n_d\rangle + 1/2. \tag{14}$$

We note that the zero-point energy of the reservoir appears in the equations of motion for the cumulants, though it is not present in the Heisenberg equations of motion and can even be dropped from the Hamiltonian by redefining a zero of energy. Such a "reappearance" of a zero-point field energy is quite common in other quantum theory problems where the vacuum is responsible for physical effects [20].

The equations of motion for the second-order cumulants B and C [(9), (10)] are linear inhomogeneous equations. Their solution consists of two parts: a general

solution of the homogeneous set of equations, (i.e., without the term $+i\gamma B^{(0)}$ in (10)) that we denote as $(\bar{B}(t), \bar{C}(t))$ and the particular solution of the inhomogeneous equations

$$(B(t), C(t)) = (\bar{B}(t), \bar{C}(t)) + (\gamma B^{(0)}t, 0). \quad (15)$$

To find $(\bar{B}(t), \bar{C}(t))$, we use the perturbation theory for $N \gg 1$ and as a first step, neglect the quantum correction Q/N in (8). It is easy to see that the homogeneous equations of motion for cumulants (9) and (10) can be obtained from the classical equation (i.e., from (8) with $Q/N \longrightarrow 0$) by linearization around z (which goes by substituting $z \longrightarrow z + \delta z$, $|\delta z| \ll |z|$), if one writes the dynamical equations for the variables $(\delta z)^2$ and $|\delta z|^2$. The only difference between the linearization of the classical equations of motion and equations for quantum cumulants (9) and (10) lies in the impossibility of obtaining the initial conditions (13) for C and B from only the initial conditions for the linearized classical equations of motion (see also the discussion of this problem in [7]). Hence, we first need to know the classical solution $z_{\rm cl}(t)$, find the differentials $dz_{\rm cl}$ and $dz_{\rm cl}^*$, and then use the substitution

$$(\bar{B}(t), \bar{C}(t)) \longrightarrow (|dz|^2, (dz)^2).$$

Thus, it has become apparent that assuming the actual field deviations from the coherent state to be small and treating the small deviation as a first-order correction is not equivalent to the direct linearization around the classical trajectory. Even in the limit as $N \longrightarrow \infty$, we always deal with the effect of reservoir on the dynamics of the quantum system via the second-order cumulant B, which has the form of the quantum diffusion

$$B(t) = \bar{B}(t) + (\langle n_d \rangle + 1/2)\gamma t, \quad (16)$$

where $\bar{B}$ is obtained by linearizing around a large mean field. In particular, as follows from (16), the quantum diffusion also exists for a quiet reservoir $\langle n_d \rangle = 0$.

We now discuss the validity range of the $1/N$-expansion and the role of the quantum diffusion in different classical dynamical regimes. The validity criterion of the $1/N$-expansion can be represented in two forms. First, the $1/N$-expansion works well, provided the difference between the classical and quantum solutions is small,

$$\left|\frac{z(t) - z_{cl}(t)}{z_{cl}(t)}\right| \simeq \frac{1}{N}\frac{\left|\int^t Q(t')dt'\right|}{|z(t)|} \ll 1, \quad (17)$$

where $z_{\rm cl}(t)$ is the solution of (8) for $N \longrightarrow \infty$. To write the second form of the validity criterion of the $1/N$-expansion, we follow [6, 7] in introducing the "convergence radius"

$$R = \{[\mathrm{Re}(\delta\alpha)^2] + [\mathrm{Im}(\delta\alpha)]^2\}^{1/2}.$$

The expansion is then correct over a time interval when

$$\frac{R(t)}{|z(t)|} \simeq \frac{B^{1/2}(t)}{N^{1/2}|z(t)|} \ll 1. \quad (18)$$

As a rule, both conditions (17) and (18) determine the same time interval for the validity of the $1/N$-expansion [6, 7]. (For a physically interesting exception, the problem of SHG, see Section 4.)

For dissipative systems with a simple attractor, the classical field intensity $|z_{\rm cl}(t)|^2$ as well as the cumulants $\bar{B}(t)$ and $C(t)$ and the quantum correction $Q(t)$ are proportional to the factor $\exp(-\gamma t)$; therefore, as follows from (17) and (18), with (16) taken into account, the $1/N$-expansion is well defined only in the time interval of the order of several relaxation times:

$$t^* \simeq \gamma^{-1}$$

(see [19]). Moreover, during this time interval, the effect of quantum diffusion on the system dynamics is small.

A quite different behavior is characteristic of the stable limit cycle. Here, a variation near the classical trajectory collapses to zero ($\delta\alpha \longrightarrow 0$), hence,

$$\bar{B}(t) \simeq |\delta\alpha|^2 \longrightarrow 0, \quad C(t) \simeq (\delta\alpha)^2 \longrightarrow 0.$$

However, $|z_{\rm cl}(t)| \simeq 1$ for the limit cycle and, as a result, the time interval of the validity of the $1/N$-expansion is rather large,

$$t^* \simeq N\gamma^{-1}.$$

It is important that the diffusion is a major physical mechanism responsible for the difference between the classical and quantum dynamics for a stable limit cycle. In the following two sections, we consider two typical examples of dissipative optical systems with a simple attractor and a limit cycle.

3. NONCLASSICAL STATE GENERATION IN HIGHER ORDER ANHARMONIC OSCILLATORS

We start by defining the squeezing and the Fano factor. We define the general field quadrature as

$$X_\theta = a\exp(-i\theta) + a^\dagger\exp(i\theta),$$

where θ is the local oscillator phase. A state is called squeezed if there exists a value of θ for which the variance of X_θ is smaller than the variance for the coherent state or the vacuum [1, 9]. Minimizing the variance of

X_θ over θ, we obtain the condition of the so-called principal squeezing [1, 9, 10] in the form

$$S \equiv 1 + 2N(\langle|\delta\alpha|^2\rangle - |\langle(\delta\alpha)^2\rangle|) = 2(B - |C|) < 1. \quad (19)$$

The determination of the principal squeezing S is very useful because it gives the maximum squeezing measurable by the homodyne detection [1, 9].

Another important characteristic of nonclassical properties of light is the Fano factor

$$F = (\langle n^2\rangle - \langle n\rangle^2)/\langle n\rangle$$

that determines the deviation of the probability distribution from the Poisson distribution [1, 9]. After the substitution $a \longrightarrow z + \delta\alpha$ in the expressions

$$\langle n\rangle = N\langle a^\dagger a\rangle$$

and

$$\langle n^2\rangle = N^2\langle a^\dagger a a^\dagger a\rangle = N^2\langle a^{\dagger 2}a^2\rangle + \langle n\rangle,$$

and after the Taylor expansions to the first order of $1/N$, we obtain

$$F = 2B + \left(\frac{z^*}{z}C + \text{c.c.}\right). \quad (20)$$

We see that in order to determine the time dependence of the principal squeezing S in (19) and the Fano factor (20) for nonlinear oscillators, we must find the time dependence of z, C, and B in (8)–(10) for Hamiltonian (1). Following the general procedure described in previous section, we first neglect the quantum correction Q/N in (8). In this case, equation (8) has the exact solution

$$z(t) = z_0\exp[(-i\Delta - \gamma/2)t]\exp[-ig_l|z_0|^{2l}\mu_l(t)],$$
$$\mu_l(t) \equiv [1 - \exp(-\gamma l t)]/\gamma l. \quad (21)$$

We find the differentials dz and dz^* of classical solution (21) and using the substitutions $|dz|^2 + \tilde{B} \longrightarrow B$ and $(dz)^2 \longrightarrow C$, we obtain

$$C(t) = -lz_0^2|z_0|^{2(l-1)}g_l\mu_l(t)(l|z_0|^{2l}g_l\mu_l(t) + i)$$
$$\times \exp[(-\gamma - i2\Delta)t - i2|z_0|^{2l}g_l\mu_l(t)],$$
$$B(t) = \exp(-\gamma t)[1/2 + l^2|z_0|^{4l}g_l^2\mu_l^2(t)]$$
$$+ (\langle n_d\rangle + 1/2)\gamma t, \quad (22)$$

where we took the initial conditions for B and C, (13), into account. Inserting (22) in (19), we obtain in the limits $\tau \equiv g_l(N)t \ll 1$ and $\gamma t \ll 1$ a very simple time dependence of S,

$$S(t) = 1 - [lx_0^{2l} - (\gamma/g_l)\langle n_d\rangle]2\tau < 1, \quad (23)$$

where, for the sake of simplicity, we assume that the initial value z_0 is real, $x_0 = \text{Re}\,z_0$, and only the terms that are linear in τ and γt are taken into account. The short-time approximation $\tau \ll 1$ and the limit of a large photon number $N \gg 1$ are quite realistic for a nonlinear medium modeled by the anharmonic oscillators (for numerical estimates, see [1, Chap. 10] and [10]). It should be noted that our formula (23) coincides with the corresponding formula for $S(t)$ in [10] for the Kerr nonlinearity ($l = 1$) with zero loss ($\gamma = 0$). In the case where $\gamma = 0$, our formula (23) shows that the rate of squeezing is determined by the factor

$$2lx_0^{2l}\lambda_l N^t \equiv 2l\mathcal{P}^{(2l+1)}.$$

Since λ_l is proportional to the $(2l + 1)$-order nonlinear susceptibility, the factor $\mathcal{P}^{(2l+1)}$ has a physical meaning of nonlinear polarization. Therefore, the stronger the nonlinear polarization induced by light in the medium, the greater the possibility of effective squeezing of light. For a finite dissipation $\gamma \neq 0$, the squeezing is determined by an interplay between the polarization of nonlinear medium modeled by the anharmonic oscillator and the thermal fluctuations of the reservoir. As follows from (23), there exists a critical number of phonons

$$\langle n_d\rangle^{(\text{cr})} = (l/\gamma)\mathcal{P}^{(2l+1)}$$

such that the squeezing is no longer possible for $\langle n_d\rangle \geq \langle n_d\rangle^{(\text{cr})}$.

In the same approximation, we obtain from (20) the following time dependence of the Fano factor

$$F(t) = 1 + 2\langle n_d\rangle\gamma t. \quad (24)$$

Thus, the statistics is super-Poissonian for any $\gamma \neq 0$ and is independent of the degree of nonlinearity l. This is in good agreement with the earliest result of [9] for a dissipative Kerr oscillator ($l = 1$), where the impossibility of sub-Poissonian statistics and antibunching were found from the exact solution.

We now discuss the validity ranges of our approach. It is easy to see that in terms of this approach, the time dependence of the number of quanta for $l = 1$ is

$$\langle n\rangle(t) + 1/2 = N|z|^2 + B \approx N|z_0|^2(1 - \gamma t) + \langle n_d\rangle\gamma t,$$
$$\gamma t \ll 1, \quad g_l t \ll 1, \quad (25)$$

where we have used (22) for cumulants B and C. It is instructive to compare (25) with the exact solution for $\langle n\rangle(t)$ for the Kerr nonlinearity [9],

$$\langle n\rangle(t) = \langle n_0\rangle\exp(-\gamma t) + [1 - \exp(-\gamma t)]\langle n_d\rangle. \quad (26)$$

Equations (25) and (26) both describe the evolution of an initially coherent state to a final chaotic state that is characteristic of the reservoir. It is evident that (26) and (25) coincide when $\gamma\tau \ll 1$ and $\langle n_0\rangle \simeq N \gg 1$. A more accurate analysis of the validity condition of the $1/N$-expansion should include a comparison of the solution of quantum motion equation (8), which takes into account the quantum correction Q/N given by (11), with the solution of classical motion equation (21). It

may be shown after some algebra, that if $\gamma t \ll 1$ and $\tau \ll 1$, the effect of the quantum correction Q/N on the dynamics of the mean value z is of the order of $1/N$ and, therefore, our cumulant expansion is well defined for $N \gg 1$. The same conclusion could be obtained from another criterion of validity (18).

4. QUANTUM–CLASSICAL CORRESPONDENCE IN SELF-PULSING REGIME OF SECOND-HARMONIC GENERATION

We now consider another example of a quantum optical system, namely intracavity SHG. The Hamiltonian describing two interacting quantum modes in the interaction picture has the form [11, 14]

$$H = \sum_{j=1}^{2} \Delta_j b_j^\dagger b_j + iEN^{1/2}(b_1^\dagger - b_1) + \frac{i\chi}{2}(b_1^{\dagger 2} b_2 - b_1^2 b_2^\dagger), \quad (27)$$

where the boson operators b_j ($j = 1, 2$) describe the fundamental and second-harmonic modes, respectively, Δ_j is the cavity detuning of mode j, $EN^{1/2}$ is the classical field driving first mode (E is of the order of unity), and χ is a second-order nonlinear susceptibility. The linear reservoir and its interaction with a second-order nonlinear medium are described by Hamiltonians (2). Now, we can rewrite the full Hamiltonian of the problem as $H = N\mathcal{H}$, where $\mathcal{H}$ has the same form as (27) and (2) with replacements analogous to (4) taking into account and with the new coupling constant defined by

$$g = \chi\sqrt{N}, \quad (28)$$

which is of the order of unity. Formally, the $1/N$-expansion procedure developed in Section 2 cannot be applied to the problem of SHG; however, its straightforward generalization to two interacting modes gives in the first order of $1/N$ the following self-consistent set of equations

$$\dot{z}_1 = -\frac{\gamma_1}{2} z_1 + E + g z_1^* z_2 + \frac{1}{N} g B_{12}, \quad (29)$$

$$\dot{z}_2 = -\frac{\gamma_2}{2} z_2 - \frac{g}{2} z_1^2 - \frac{1}{N}\frac{g}{2} C_1, \quad (30)$$

$$\dot{B}_1 = -\gamma_1(B_1 - B^{(0)}) + g B_{12}^* z_1 + g B_{12} z_1^* + C_1^* z_2 + C_1 z_2^*, \quad (31)$$

$$\dot{B}_2 = -\gamma_2(B_2 - B^{(0)}) - g B_{12}^* z_1 - g B_{12} z_1^*, \quad (32)$$

$$\dot{C}_1 = -\gamma_1 C_1 + 2g(C_{12} z_1^* + B_1 z_2), \quad (33)$$

$$\dot{C}_2 = -\gamma_2 C_2 - 2g C_{12} z_1, \quad (34)$$

$$\dot{C}_{12} = -0.5(\gamma_1 + \gamma_2) C_{12} + g B_{12} z_2 - C_1 z_1 + C_2 z_1^*, \quad (35)$$

$$\dot{B}_{12} = -0.5(\gamma_1 + \gamma_2) B_{12} + g C_{12} z_2^* + g z_1 (B_2 - B_1), \quad (36)$$

where

$$z_j \equiv \langle a_j \rangle = N^{1/2}\langle b_j \rangle, \quad B_j = N\langle \delta\alpha_j^* \delta\alpha_j \rangle + 0.5,$$

$$C_j = N\langle (\delta\alpha_j)^2 \rangle \ (j = 1, 2),$$

$$B_{12} = N\langle \delta\alpha_1^* \delta\alpha_2 \rangle, \quad C_{12} = N\langle \delta\alpha_1 \delta\alpha_2 \rangle,$$

and $B^{(0)}$ is defined in (14). The initial conditions for system (29)–(36) are

$$B_j(0) = 1/2, \quad C_j(0) = C_{12}(0) = B_{12}(0) = 0,$$

$$z_2(0) = 0, \quad z_1(0) = z_0,$$

where z_0 is of the order of unity. In this work, we limit ourselves by the values of the field strength z_0 corresponding to self-oscillations [11] and $\Delta_1 = \Delta_2 = 0$.

It is easy to see that in the limit as $N \longrightarrow \infty$ and for $g = \text{const} \simeq 1$, we obtain from (29) and (30) the correct classical equations of motion for the scaled field amplitudes. The solution of equations of motion (31)–(36) for the second-order cumulants has the form

$$\mathbf{X}(t) = \bar{\mathbf{X}}(t) + (\gamma B^{(0)} t, \gamma B^{(0)} t, 0, 0, 0, 0),$$
$$\mathbf{X}(t) \equiv [B_1(t), B_2(t), C_1(t), C_2(t), B_{12}(t), C_{12}(t)], \quad (37)$$

where the vector $\bar{\mathbf{X}}$ describes the part of $\mathbf{X}$ that can be obtained by linearization around the classical trajectory. Variations near a stable limit cycle rapidly approach zero and, therefore, $\bar{\mathbf{X}}(t) \longrightarrow 0$. As a result, we have only a diffusive growth of cumulants B_j ($j = 1, 2$) as

$$B_j(t) = 0.5\gamma_j t, \quad (38)$$

where we considered the case of a quiet reservoir $\langle n_d \rangle$. This result indicates that the effect of reservoir zero-point energy on the dynamics of the nonlinear system is the principal physical mechanism responsible for the difference between the classical and quantum dynamics in the semiclassical limit. A time scale t^* for a correct description of the quantized SHG dynamics in terms of classical electrodynamics can be found using criterion (18). Taking into account that $|z(t)| \simeq 1$, we obtain $t^* \simeq 2N\gamma^{-1}$.

We note that the quantum corrections to the classical equations of motion (29) and (30) do not include the cumulants $B_{1,2}$. Therefore, in the first order of $1/N$, there is no difference between the evolution of quantum mean values and the classical limit cycle dynamics. In other words, the quantum correction $Q \longrightarrow 0$, and therefore, criterion (17) of the $1/N$-expansion validity does not work. In this respect, the quantized SHG is a somewhat singular problem. In other quantum optical

systems, for instance, for a nonlinear oscillator with $l \geq 1$, both validity criteria (18) and (17) typically give the same result.

Over a decade ago, Savage addressed the same quantum–classical correspondence problem for self-oscillations in SHG numerically [14]. He calculated the Q distribution function in the Gaussian approximation centered at a deterministic trajectory corresponding to a limit cycle. He worked in a large field and small nonlinearity limits, $\chi/\gamma_{1,2} \longrightarrow 0$, which correspond to the classical limit [14]. It is easy to see that the condition $\chi/\gamma_{1,2} \longrightarrow 0$ is consistent with our condition $N \gg 1$, if one additionally considers the natural condition of a not very strong dissipation in (29)–(36), $\gamma_{1,2}/g < 1$ together with $g \simeq 1$ (28). In other words, Savage's small parameter χ/γ corresponds to our large parameter N as $\chi/\gamma \longrightarrow N^{-1/2}$. To establish the difference between the classical and quantum dynamics, the equations of motion for low-order cumulants were obtained in [14] and solved numerically for particular values of the parameters [21]. Based on the results of numerical simulations, Savage concluded that it is a quantum diffusion that is mostly responsible for the difference between the classical and quantum dynamics in the semiclassical limit. Moreover, his numerical estimate for a characteristic time for the classical description scales as $(\gamma/\chi)^2$, which is in a good agreement with our analytic result $t^* = 2\gamma^{-1}N$. In summary, our analytic results for the quantum–classical correspondence at self-pulsing in SHG are consistent with the previous numerical investigation of same problem in [14].

5. CONCLUSION

We developed the $1/N$-expansion method to consider the nonlinear dynamics and nonclassical properties of light in dissipative optical systems in the limit of a large number of photons. The method was applied to the investigation of squeezing in higher order dissipative nonlinear oscillators. We would like to note that our method can also be directly applied to an important case of the generation of nonclassical states in a medium involving competing nonlinearities [22].

We found a time scale of validity of the $1/N$-expansion for a classical description of the dynamics of nonlinear optical systems with a simple attractor and with a limit cycle. For systems with a simple attractor, this time scale is of the order of unity, and for the limit cycle, is proportional to large N. Qualitatively, this result can be understood as follows. For time of the order of unity, the trajectory spirals around a stable stationary point with a small amplitude, and therefore, by virtue of the uncertainty principle, the contribution of quantum corrections to the classical equations of motion becomes very important. Unlike the previous case, the oscillations corresponding to the limit cycle are often close to harmonic and, thus, their quantum and classical descriptions can coincide for a sufficiently long period of time. The basic difference between the classical and quantum dynamics in the latter case originates from the influence of reservoir zero-point fluctuations, which, in our notation, are of the order of $1/N$. This result is in a good agreement with the result of earlier numerical simulations of self-oscillations in the quantized second-harmonic generation [14]. Finally, it should be noted that our findings are of a rather general nature and can be applied to the investigations of self-oscillations in other optical systems, for example, in those involving optical bistability [23–25].

ACKNOWLEDGMENTS

We would like to thank A. Heidmann, E. Bulgakov, and Z. Hradil for their useful discussions. The work was partially supported by Czech Grant Agency (grant no. 202/96/0421) and Czech Ministry of Education (grant no. VS96028).

REFERENCES

1. J. Peřina, *Quantum Statistics of Linear and Nonlinear Optical Phenomena* (Kluwer, Dordrecht, 1991).
2. C. Fabre, Phys. Rep. **219**, 215 (1992).
3. S. Reynaud, A. Heidmann, E. Giacobino, *et al.*, in *Progress in Optics XXX*, Ed. by E. Wolf (Elsevier, Amsterdam, 1992), p. 1.
4. A. Heidmann, J. M. Raimond, S. Reynaud, and N. Zagury, Opt. Commun. **54**, 189 (1985).
5. L. G. Yaffe, Rev. Mod. Phys. **54**, 407 (1982).
6. K. N. Alekseev, Opt. Commun. **116**, 468 (1995).
7. K. N. Alekseev and J. Peřina, Phys. Lett. A **231**, 373 (1997); Phys. Rev. E **57**, 4023 (1998).
8. K. N. Alekseev and D. S. Priimak, Zh. Éksp. Teor. Fiz. **113**, 111 (1998) [JETP **86**, 61 (1998)].
9. V. Peřinová and A. Lukš, in *Progress in Optics XXXIII*, Ed. by E. Wolf (Elsevier, Amsterdam, 1992), p. 130.
10. R. Tanaš, A. Miranowicz, and S. Kielich, Phys. Rev. A **43**, 4014 (1991).
11. P. D. Drummond, K. J. McNeil, and D. F. Walls, Opt. Acta **27**, 321 (1980).
12. H. J. Kimble and J. L. Hall, in *Quantum Optics IV*, Ed. by J. D. Harvey and D. F. Walls (Springer-Verlag, Berlin, 1986).
13. J. Peřina, J. Křepelka, R. Horák, *et al.*, Czech. J. Phys., Sect. B **37**, 1161 (1987).
14. C. M. Savage, Phys. Rev. A **37**, 158 (1988).
15. R. Schack and A. Schenzle, Phys. Rev. A **41**, 3847 (1990).
16. P. Szlachetka, K. Grygiel, J. Bajer, and J. Peřina, Phys. Rev. A **46**, 7311 (1992); K. Grygiel and P. Szlachetka, Phys. Rev. E **51**, 36 (1995).
17. C. M. Savage, Comput. Phys. **6**, 513 (1992); X. Zheng and C. M. Savage, Phys. Rev. A **51**, 792 (1995).
18. K. N. Alekseev, G. P. Berman, A. V. Butenko, *et al.*, J. Mod. Opt. **37**, 41 (1990); Kvantovaya Élektron. (Mos-

cow) **17**, 425 (1990) [Sov. J. Quant. Electron. **20**, 359 (1990)].

19. K. N. Alekseev and J. Peřina, Phys. Scr. **61**, 7 (2000); E-print archives, quant-ph/9812019.
20. P. W. Milonni, *The Quantum Vacuum: An Introduction to Quantum Electrodynamics* (Academic, Boston, 1994), Section 2.13.
21. The analogous motion equations for cumulants in the problem of SHG were also considered in the papers [13, 15, 16]. However, the limit of large photon numbers was not analysed in these works.
22. P. Tombesi, Phys. Rev. A **39**, 4288 (1989); M. A. M. Marte, J. Opt. Soc. Am. B **12**, 2296 (1995); J. Peřina and J. Bajer, J. Mod. Opt. **42**, 2071 (1995); C. Cabrillo, J. L. Roldán, and P. García-Ferández, Phys. Rev. A **56**, 5131 (1997).
23. K. Ikeda and O. Akimoto, Phys. Rev. Lett. **48**, 617 (1982).
24. L. A. Lugiato, L. M. Narducci, D. K. Bandy, *et al.*, Opt. Commun. **4**, 281 (1982); L. A. Lugiato, R. J. Horowicz, G. Strini, *et al.*, Phys. Rev. A **30**, 1366 (1984).
25. L. A. Orozco, A. T. Rosenberger, and H. J. Kimble, Phys. Rev. Lett. **53**, 2547 (1984).

Journal of Experimental and Theoretical Physics, Vol. 90, No. 4, 2000, pp. 600–608.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 4, 2000, pp. 691–701.

ATOMS, SPECTRA, RADIATION

Two-Dimensional Photonic Crystals with a Lattice Defect: Spectrum of Defect Modes, Localization of Light, and Formation of Evanescent Waves

A. M. Zheltikov*, S. A. Magnitskiĭ, and A. V. Tarasishin

International Laser Center, Faculty of Physics, Moscow State University, Moscow, 119899 Russia
**e-mail: zheltikov@nlsl.ilc.msu.su*

Received September 14, 1999

Abstract—Properties of an electromagnetic field localized in the defect modes of two-dimensional photonic crystals are studied. The defect-mode spectrum of these structures is calculated, electromagnetic field localization and channeling effects are analyzed, and the properties of the field inside and beyond a photonic crystal with a lattice defect are also studied. The calculations show that the electromagnetic field is localized in the defect mode of a photonic crystal in a region smaller than the wavelength. The dependence of the defect-mode spectrum on the parameters of the photonic crystal is investigated and possibilities for controlling the spectrum of defect modes are indicated. It is shown that the optical field leaving a photonic crystal possesses the properties of a evanescent wave, which means that spatial resolution substantially greater than the wavelength of the radiation can be achieved in the near field and opens up possibilities for using photonic crystals with a lattice defect in near-field optical microscopy. The possibility of externally controlling an optical field localized in the defect modes of a photonic crystals is demonstrated. *© 2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Photonic crystals [1–4] are a new type of artificial, structurally organized media whose dielectric properties vary periodically in one, two, or three dimensions with a characteristic spatial periodicity scale of the order of the optical wavelength. An important characteristic of these structures, associated with the periodicity of the refractive index distribution, is the presence of photonic band gaps (PBGs) which are spectral regions in which light waves cannot propagate in the photonic crystal, being exponentially attenuated and reflected from the structure. As a result of the existence of PBGs and their unusual dispersion properties, photonic crystals can sustain various light wave, pulse, and beam propagation regimes which are of physical interest and important for numerous applications. Over the last few years, these aspects have been intensively studied and discussed in the extensive literature on this topic (see, for example, [2–7]). In particular, photonic-crystal structures can be used to solve various fundamental problems associated with controlling spontaneous emission [1], reducing the group velocity of light [8], localization and channeling of light [9, 10], and also problems associated with increasing the efficiency of nonlinear optical interactions and possible methods of controlling phase locking in these interactions [11–14]. The relevance of the practical applications of PBG structures is indicated by the rapid progress of Bragg reflectors and chirped mirrors [15], the development of photonic-crystal waveguides and devices for rotating laser radiation [9, 16], and also the recently widely discussed ideas of developing low-threshold optical switches and limiters [17–19], compact optical delay lines [8] and light pulse compressors [20], channel drop filters [21], and also nonlinear optical diodes [22].

The incorporation of a defect in the crystal lattice of a photonic-crystal structure leads to the appearance of defect modes in the photonic band gap, which allows us to observe new physical phenomena. An important property of these structures is the localization of the electromagnetic field in the defect modes [4]. These effects in photonic crystals have been widely studied by analyzing the dispersion properties of PBG structures [23] which can reveal the main properties of the transmission spectrum of the PBG structure. Various modifications of the slowly varying amplitude approximation [24] have also been used to describe the main laws governing the localization of a plane-wave field with a slowly varying envelope in one-dimensional photonic crystals. However, a detailed physical understanding of the effects of electromagnetic field localization in defect modes of photonic crystals and also of the properties of the field formed at the exit from these structures cannot be obtained by analyzing the dispersion properties of PBG structures or using the plane-wave approximation, because these approaches do not allow us to determine the field distribution in the photonic crystal. Quite clearly, the only possible method of studying the field distribution in these structures involves a numerical integration of the Maxwell equations since the problem of obtaining an analytic description of the field of light beams inside and at the exit from two-dimensional and

three-dimensional PBG structures is almost impossible to solve.

In the present paper, we investigate the defect-mode spectrum of a two-dimensional photonic crystal, we analyze the effects of electromagnetic-field localization and channeling in a two-dimensional PBG structure with a lattice defect, and we also study the properties of the electromagnetic field inside and at the exit from such a structure. The calculations show that the electromagnetic field can be localized in the defect mode of a PBG structure in a region smaller than the wavelength. We investigate the dependence of the defect-mode spectrum on the parameters of a photonic crystal and suggest possibilities for controlling the defect-mode spectrum. An optical field leaving a photonic crystal possesses the properties of a evanescent wave which means that the near-field spatial resolution can be substantially greater than the radiation wavelength and opens up possibilities for using photonic crystals with a lattice defect in near-field optical microscopy. It is shown that placing a dielectric sphere close to the exit face of a PBG structure with a lattice defect can change the optical-field distribution in the defect mode of a photonic crystal. This effect can be used to control the localization of the optical field in photonic crystals from the outside.

2. CALCULATION METHOD

2.1. Difference Scheme

The method predominantly used for theoretical analyses of PBG structures at present is the plane-wave approximation which involves a numerical solution of the Hill equation using the Floquet theorem (this approach is discussed in detail in the reviews [7, 25]). This technique can ensure fairly high accuracy in calculations of the dispersion properties of PBG structures but gives rise to serious difficulties in calculations of the field distribution in and at the exit from a PBG structure.

A promising approach to study the field distribution in PBG structures is based on a numerical solution of the Maxwell equations using a finite-difference time-domain (FDTD) method [26]. This method is now being increasingly widely used to study photonic crystals. In particular, this approach has been used to analyze the phenomenon of optical switching [18, 27] and the formation of ultrashort optical pulses in one-dimensional PBG structures [20, 28], and also to study the influence of the material dispersion on the reflection and transmission spectra of multilayer media. The last two years have seen the publication of studies in which the FDTD method was used to investigate the propagation of light in two-dimensional photonic crystals [21, 29].

Unlike methods based on the plane-wave approximation, the FDTD method does not require the inversion of large matrices and consequently does not require a large computer memory. As a result, the FDTD method can be used to calculate the propagation of spatially bounded laser pulses in photonic-crystal structures and to analyze the possibility of controlling light pulses and beams. In connection with the problems addressed in the present study, it is important to note that since the FDTD procedure involves calculations in real space not in Fourier space, the addition of defects to the photonic-crystal structure does not significantly complicate the calculations made by the FDTD scheme. Finally, by using the FDTD technique we can go over from analyzing linear to nonlinear problems and analyze a wide range of phenomena associated with the development of photonic-crystal compressors, switches, and modulators using ultrashort optical pulses [18, 20, 27, 28].

In the FDTD method, the region of space being studied is divided into cells of length $\delta x = \delta y = \delta z = \delta$, and the time interval being studied is divided into sections of length δt. The real electric and magnetic fields are replaced by discrete functions of the time variable n and the spatial coordinates i, j, k:

$$\begin{aligned}\mathbf{E}(t, x, y, z) &\longrightarrow \mathbf{E}^n(i, j, k),\\ \mathbf{H}(t, x, y, z) &\longrightarrow \mathbf{H}^n(i, j, k).\end{aligned} \tag{1}$$

The Maxwell equations for the two-dimensional case are written in the following form:

$$\begin{aligned}H_x^{n+1/2}\left(i, j+\frac{1}{2}\right) &= H_x^{n-1/2}\left(i, j+\frac{1}{2}\right)\\ &+ \frac{\delta t}{\delta}[E_z^n(i, j) - E_z^n(i, j+1)],\\ H_y^{n+1/2}\left(i+\frac{1}{2}, j\right) &= H_y^{n-1/2}\left(i+\frac{1}{2}, j\right)\\ &+ \frac{\delta t}{\delta}[E_z^n(i+1, j) - E_z^n(i, j)],\\ E_z^{n+1}(i, j) &= E_z^n(i, j) + \frac{\delta t}{\delta\varepsilon(i, j)}\\ &\times\left[H_x^{n+1/2}\left(i, j-\frac{1}{2}\right) - H_x^{n+1/2}\left(i, j+\frac{1}{2}\right)\right.\\ &\left. + H_y^{n+1/2}\left(i+\frac{1}{2}, j\right) - H_y^{n+1/2}\left(i-\frac{1}{2}, j\right)\right].\end{aligned} \tag{2}$$

Defining the initial distribution of the fields **H** and **E**, we can calculate the field distribution for any spatial cell for each time interval using the difference scheme described above. The relationship between the displacement and electric field vectors is defined phenomenologically:

$$D_z^n(i, j) = \varepsilon(i, j)E_z^n(i, j), \tag{3}$$

where $\varepsilon(i, j)$ is the permittivity of the medium.

In this formulation of the problem, we neglect the spatial and frequency dispersions of the material form-

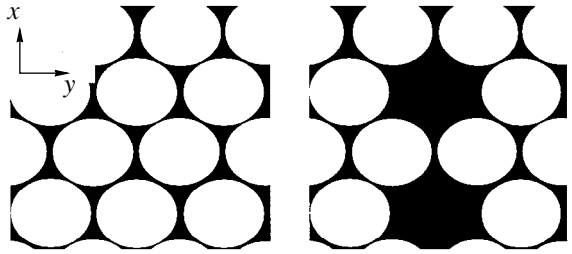

Fig. 1. Fragment of two-dimensional PBG structure without (a) and with (b) a lattice defect.

ing the PBG structure. However, in principle the FDTD scheme can also allow for these dispersion effects (see [30]).

2.2. Boundary Conditions

The solution of initial and initial–boundary-value wave problems for regions of finite spatial dimensions (photonic crystals of finite dimensions) encounters difficulties because an electromagnetic wave reaching the boundary of the spatial region being studied is reflected specularly from this boundary, reentering the calculation region and thereby distorting the real distribution of the electromagnetic field. One method of solving this problem is to enlarge the calculation region to such an extent that the light wave (pulse) does not have time to reach the boundaries during its time of propagation through the region of interest to us. In this case, our solution will not contain waves produced by artificial reflection from the boundaries of the calculation region. A similar procedure gives good results for short (in time and space) optical pulses but excessively increases the calculation time for relatively long pulses. At the same time, short optical pulses can undergo appreciable spreading in a photonic crystal as a result of the substantial dispersion of the group velocity, which leads to an increase in the duration of the pulses transmitted by and reflected from the structure.

With this reasoning in mind, the boundary conditions should be formulated so that a light wave reaching the boundary is completely absorbed by the boundary without being reflected. Unfortunately, no ideally absorbing, fairly simple boundary conditions exist for two-dimensional and three-dimensional problems. For our calculations, we used second-order absorbing boundary conditions [31]

$$\left(\frac{\partial^2}{\partial x \partial t} - \frac{\partial^2}{\partial t^2} + \frac{1}{2}\frac{\partial^2}{\partial y^2}\right)E_z^n\Big|_{x = 0, L} = 0, \qquad (4)$$

where $x = 0, L$ correspond to the boundaries of the calculation region. The conditions (4) can significantly reduce the influence of reflection from the boundaries on the accuracy of the numerical calculations.

In order to check the efficacy of the boundary conditions (4), we modeled the reflection of strongly focused short pulses numerically. These calculations showed that for a Gaussian beam whose diameter is of the order of the wavelength, the reflection coefficient for an angle of incidence of 45° is less than 3% in terms of amplitude, which agrees with the results of calculations made by the authors [31] for a plane monochromatic wave.

By analogy with higher order boundary conditions, the second-order boundary conditions give reflection coefficients of unity for glancing incidence. However, the corresponding field components propagate along the boundaries and generally have no significant influence on the field distribution in the region of interest.

2.3. Photonic Crystals

For our investigations we selected a structure consisting of a variable number of periods (between five and ten) of cylindrical air pores forming a triangular lattice in a silicon matrix (Fig. 1). This type of photonic crystal was selected because silicon technologies are extremely promising for fabricating one-dimensional [14, 32], two-dimensional [33], and three-dimensional [34] PBG structures.

In order to introduce a defect in the two-dimensional PBG structure described above, we removed one row of air pores (Fig. 1b). Since, in the region of the complete band gap in our selected triangular lattice of air cylinders in silicon, the depth of penetration of the optical field in the photonic crystal (skin layer) does not exceed a single period of the structure, in order to determine the transmission spectrum of a lattice with a defect we considered structures in which the defect was introduced periodically (with the period Λ along the y axis). In other words, we solved the problem of light propagation in a superlattice. Control calculations made for superlattices differing only in terms of the period Λ showed that the electromagnetic-field distribution in the defect mode and the defect-mode spectrum do not depend on Λ. As well as being convenient, this calculation scheme can also be used to determine the noise level associated with the influence of light localized in neighboring defects (cross talk) which is of interest for applications of these PBG structures in optical memory systems and optical information processing.

2.4. Calculations of Transmission Spectra

The transmission spectrum for a two-dimensional photonic crystal with no lattice defects was calculated by modeling the propagation of a plane monochromatic wave

$$E_z^{n+1}(2, j) = E_z^n(2, j) + \sin\left(\frac{2\pi}{\lambda}\delta t n\right), \qquad (5)$$

incident on a photonic crystal located in the xy plane, parallel to the x axis. The absorbing boundary condi-

tions (4) were set in the planes $x = 0$ and $x = L$. The calculation region along the y axis corresponded to a single period of the crystal superlattice. By virtue of the symmetry of the problem, the field components in the planes $y = 0$ and $y = \Lambda$ are equal:

$$\begin{aligned} H_x^n(i, 0) &= H_x^n(i, N_y), \\ H_y^n(i, 0) &= H_y^n(i, N_y), \\ H_z^n(i, 0) &= H_z^n(i, N_y). \end{aligned} \quad (6)$$

Here, $N_y = \Lambda/\delta$ is the number of mesh divisions per period of the photonic crystal in the direction of the y axis.

Thus, the periodic Bloch conditions in real space simply took the form of equal field components (6) on these two planes. The calculations were made for the squares of the electric and magnetic fields averaged over the period of the oscillations. The calculation procedure continued until these values remained the same to within the required accuracy after doubling the computation time.

This scheme was used to calculate the transmission spectrum of a square lattice consisting of sixteen layers of circular air cylinders of radius $R = 0.504$ μm in a PbO matrix (permittivity $\varepsilon = 2.72$) having the period $a = 1.17$ μm. The results of the calculations completely agreed with the calculations made for the same structure using a Fourier expansion of the fields and dielectric constants in [15], which indicates that this calculation procedure is reliable.

The transmission spectrum of two-dimensional PBG structures with a lattice defect was calculated using a test pulse scheme. In this approach, the shortest possible (one or two optical periods) pulse is defined at the entrance to a photonic crystal at zero time:

$$E_z^0(i, j) = Ae^{-0.5((i-i_0)\delta/l)^2}\cos\left(\frac{2\pi}{\lambda_0}(i-i_0)\delta\right), \quad (7)$$

where l is the pulse length in space (in our calculations $l = 1 - 2\lambda_0$), λ_0 is the wavelength corresponding to the pulse carrier frequency, and $x = i_0\delta$ is the coordinate of the pulse center.

In order to determine the transmission spectrum of the photonic crystal, the spectrum of the pulse transmitted by the PBG structure was normalized to the spectrum of the initial pulse. By performing a similar procedure for several test pulses whose spectra cover the entire spectral range of interest to us, we can find the transmission spectrum of the crystal, including the spectrum of the defect modes. This approach can be used to draw important conclusions from the physical point of view on the nature of light channeling in the defect modes of PBG structures.

3. RESULTS AND DISCUSSION

3.1. Spectrum of Defect Modes

In order to study the transmission spectrum of two-dimensional PBG structures, the field incident at the boundary of the PBG structure was defined as a plane wave (5). For the investigation we selected a structure consisting of a variable number of periods (between five and ten) of cylindrical air pores forming a triangular lattice in a silicon matrix (see Section 2.3 and Fig. 1).

Numerical calculations made for these PBG structures with no defects (Fig. 1a) showed that a band gap exists for the direction along the x axis, corresponding to the minimum width of the band gap [36], for the ratio a/λ (where a is the period of the PBG structure, λ is the wavelength of the optical beam) which varies between 0.35 and 0.52 for H-waves and between 0.44 and 0.57 for E-waves. Thus, the results of our calculations indicate that a closed band gap exists and show good agreement with the results of calculations made by the plane-wave method in [36] and also with the calculated and experimental results reported in [33].

The numerical analysis shows that whereas the field intensity in a PBG structure with no defects decreases on a spatial scale of the order of the wavelength λ and the PBG structure has a transmission coefficient of the order of 10^{-3}, in a PBG structure having a defect (Fig. 1b) the optical beam can propagate along the narrow channel formed by this defect. In this structure, the transmission coefficient for E-waves in the range of a/λ between 0.44 and 0.47 increases from 10^{-3} to 0.5 which indicates the appearance of a defect level in the band gap. Figure 2 shows the spectrum of defect modes for propagation along the x axis for E-waves. This spectrum was calculated using the wide-band test pulse method described in Section 2.4.

An analysis of the transmission spectrum of a PBG structure with defects reveals that the defect-mode spectrum can be controlled by varying the parameters of the photonic crystal. It can be seen from Fig. 2 that a decrease in the ratio of the refractive index of the matrix material to that of the cylinder material (in our case air) shifts the defect modes of the PBG structure toward higher frequencies. A qualitatively clear interpretation of this effect can be obtained by representing the defect mode of the PBG structure as a standing wave formed as a result of reflection from the walls of the channel formed by the defect, whose length corresponds to that of the lowest order natural mode:

$$\lambda_s = 2dn,$$

where d is the channel width and n is the refractive index of the channel material. Then, the wave number of this standing wave decreases with increasing refractive index n, as is observed in the numerical modeling.

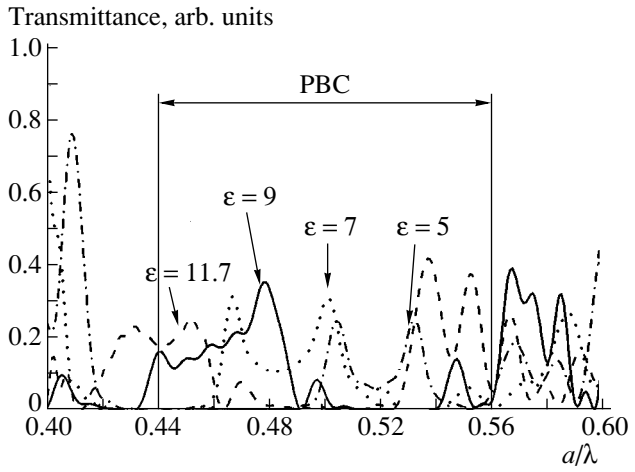

Fig. 2. Defect-mode spectrum of the two-dimensional PBG structure whose fragment is shown in Fig. 1b for various permittivities of the matrix material: $\varepsilon = 11.7$ (dashed curve), 9 (solid curve), 7 (dotted curve), and 5 (dash–dotted curve).

3.2. Localization of the Field Inside a Two-Dimensional Photonic Band Gap Structure with a Lattice Defect

We used the FDTD procedure (see Section 2.1) with the absorbing boundary conditions (4) and initial condition (5) to determine the field distribution in a photonic crystal. The calculations were continued until the period-averaged values of the electric and magnetic fields and their squares ceased to vary. The output data were the period-averaged squares (intensities) of the electric and magnetic fields.

Figures 3a–3c show two-dimensional distributions of the mean square of the electric field $\overline{E^2}$ in a PBG structure of the type described above consisting of ten periods along the x axis with the defect period $\Lambda = 5a$ for the ratio $a/\lambda = 0.454$. An analysis of the spatial distribution of the electromagnetic field shows that the light is channeled along the defect in this structure and that the field is localized at the center of the defect. In some sections of the channel the beam diameter is of the order of $\lambda/10$ (Figs. 3a–3c) which is five times smaller than the diffraction limit for an optical beam focused in air. The light intensity at the center of the defect is several hundred times higher than the intensity at the edges of the channel (Figs. 3a–3c) which indicates that the beam channeling has a high contrast. In this case, the group velocity of the light decreases substantially which indicates that the defect modes of PBG structures may be used to develop controllable optical delay lines similar to the optical delays lines using one-dimensional PBG structures developed by the authors of [8].

The results of the calculations for the superlattices with variable period Λ described above yield several important conclusions on the nature of these phenomena. First, the effect of light channeling along a defect is observed regardless of the period Λ which indicates that these light localization effects are not associated with the interference of light reflected from neighboring defects. Second, the field intensity beyond a photonic crystal far from its surface, integrated over the period Λ, remained constant regardless of the value of Λ which suggests that the light is only channeled along the particular defect. Third and finally, the field distribution in the defect mode of a photonic crystal indicates that the light intensity decreases so rapidly as a function of the y coordinate beyond the defect that the interference between neighboring defects is negligible.

An important property of the optical field localization effect in a photonic crystal is that the spatial distribution of the field in the defect mode can be controlled externally. This possibility is illustrated in Figs. 3b and 3c which show two-dimensional plots of the spatial distribution of the mean square of the electric field intensity in a two-dimensional PBG structure having the parameters described above when a transparent dielectric sphere having the refractive index 1.5 and radius $\lambda/4$ (Fig. 3b) and $\lambda/8$ (Fig. 3c) is positioned closed to the exit face of the photonic crystal. A comparison of Figs. 3b and 3c with Fig. 3a shows that the presence of a small dielectric sphere perturbs the field distribution inside the PBG structure. The physically observable effect is attributed to a change in the conditions of formation of the defect mode in the presence of the perturbing "potential" of the dielectric sphere. This effect is extremely important for possible applications of the light localization effect in photonic crystals in nonlinear optics, optical memory, and also for controlling atoms and atomic interactions in quantum calculation problems.

3.3. Formation of Evanescent Waves and Ultrahigh Near-Field Resolution

The distribution of the mean square of the electric field at the exit from a PBG structure is shown in Figs. 4a–4c. Under these conditions, the field is localized in the transverse direction on a spatial scale smaller than the wavelength ($\lambda/10$), decreasing exponentially along the x axis. The properties of this field are similar to the properties of a evanescent wave localized near the surface of the sample in near-field microscopy, which opens up the possibility of using PBG structures as elements for the efficient formation of evanescent waves in near-field optical microscopy [37]. Using the reciprocity principle, we find that PBG structures with a lattice defect can also be used to analyze the evanescent field near a sample so that an information acquisition mode can be achieved in near-field microscopy.

It is extremely unlikely that an analytic description can be obtained for the field at the exit from these PBG structures. However, it is quite clear from qualitative

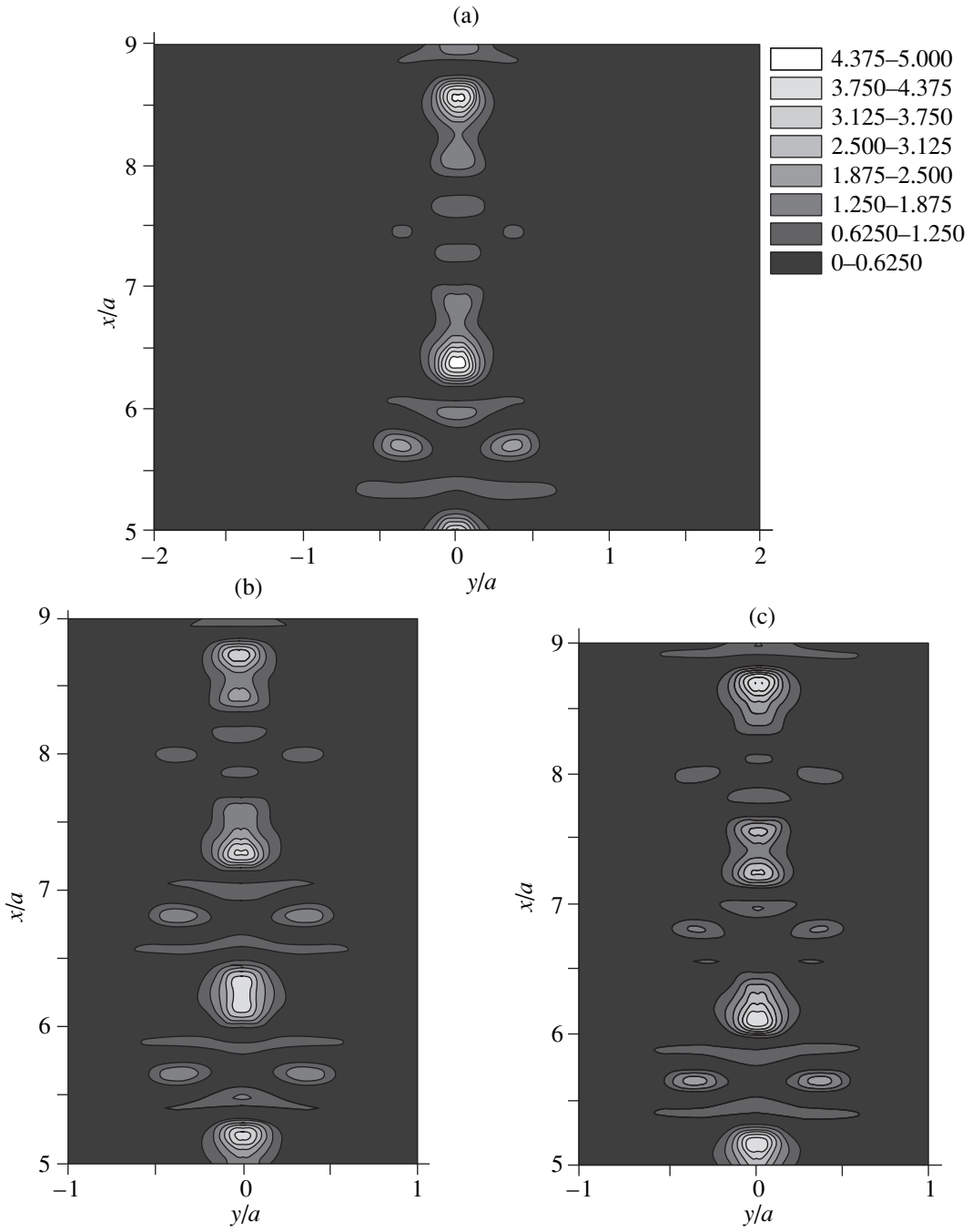

Fig. 3. Two-dimensional pattern of electric field localization in a PBG structure with a lattice defect as shown in Fig. 1b. The gray levels are the mean square of the electric field $\overline{E^2}$ for the ratio $a/\lambda = 0.454$: (a) without probe object; (b, c) with probe object of diameter $\lambda/4$ (b) and $\lambda/8$(c) whose center is positioned at distances $\lambda/8$ (b) and $\lambda/16$ (c) from the exit face of the PBG structure.

physical reasoning that an optical beam of diameter less than the wavelength, formed in a two-dimensional photonic crystal, cannot propagate in free space without changing its shape. Since this beam has a diameter substantially smaller than the wavelength and its spectrum contains higher order spatial harmonics, it exhibits properties similar to those of evanescent waves. This beam can provide high resolution in the near-field zone and decreases rapidly in the longitudinal direction (along the *x* axis).

In order to demonstrate the possibility of achieving high spatial resolution in measurements using an optical beam formed by this type of PBG structure, we calculated the electric-field distribution at the exit of a photonic crystal in the presence of a transparent dielectric sphere of refractive index 1.5 and radius $\lambda/4$ (Fig. 4b)

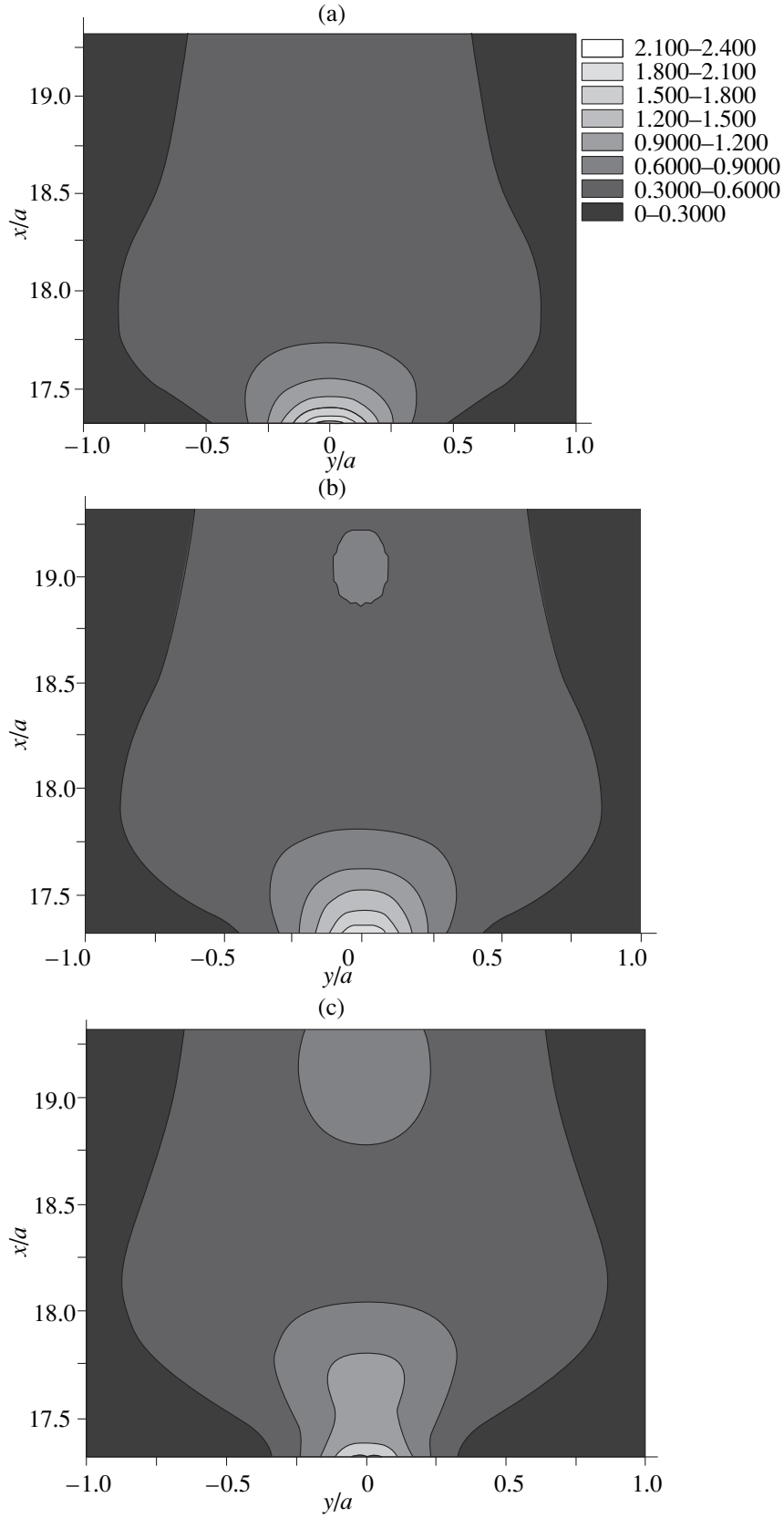

Fig. 4. Two-dimensional distribution pattern of the mean square of the electric field $\overline{E^2}$ at the exit from a PBG structure with a lattice defect as shown in Fig. 1b for the ratio $a/\lambda = 0.454$: (a) without probe object; (b, c) with probe object of diameter $\lambda/4$ (b) and $\lambda/8$ (c) whose center is positioned at distances $\lambda/8$ (b) and $\lambda/16$ (c) from the exit face of the PBG structure.

and λ/8 (Fig. 4c) positioned close to the exit face of a PBG structure. A comparison of the two-dimensional electric-field distribution patterns given in Figs. 4a–4c shows that a small dielectric probe sphere significantly changes the distribution of the radiation intensity in the near-field zone, resulting in a field "tunneling" effect as a result of the perturbation introduced by the potential of the probe sphere. The integral of the mean square of the electric field calculated in the plane $x = 19.7a$ (Figs. 4a–4c) also changes appreciably in the presence of the probe sphere. The change in this integral in the presence of probe spheres of diameters λ/4, λ/8, and λ/16 relative to the values of this integral in the absence of the probe sphere is 6, 3.2, and 0.7%, respectively. Consequently, the presence of a dielectric probe sphere of diameter less than the wavelength of the incident radiation leads to a detectable perturbation of the spatial distribution of the field at the exit of the PBG structure and also influences the characteristics of the radiation in the far field. The results of these calculations show that the evanescent waves formed in the defect modes of photonic crystals can be used for near-field optical microscopy systems (Fig. 5) with spatial resolution exceeding the wavelength of the incident radiation and a fairly high signal-to-noise ratio (cf. [38, 39]).

Although all the radiation energy in the evanescent waves only propagates along the exit surface of the PBG structure, without becoming detached from it, the high density of the optical field at the exit from this structure may give rise to physically observable phenomena in the near-field zone. This field can be used in particular for the local excitation of atoms or molecules (Fig. 6). It can also be recorded, analogous to near-field microscopy, by using a probe positioned near the surface.

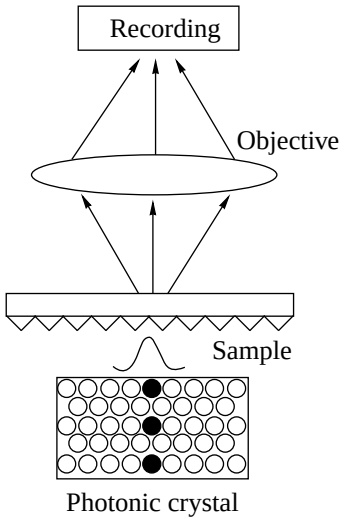

Fig. 5. Near-field optical microscopy using a evanescent wave formed in the defect mode of a photonic crystal. A PBG structure is used to excite the evanescent wave with recording in the far field.

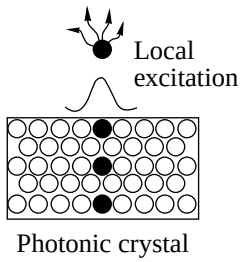

Fig. 6. Local excitation of particles by a evanescent wave formed in the defect mode of a photonic crystal.

3.4. Applications of Photonic-Crystal Defect Modes

In addition to the various near-field microscopy systems discussed briefly in Section 3.3, the optical beam propagation regimes which can be achieved using PBG structures are potentially useful for a wide range of applications. Important applications arising from the properties of the field at the exit of a PBG structure are associated with the possibilities of enhancing the spatial resolution in photolithography (as in the scheme shown in Fig. 6), increasing the data storage density in optical memory systems, and also visualizing the mode composition of the radiation in optical waveguides. The possibility of focusing an optical beam in a region smaller than the wavelength opens up new possibilities for increasing the data storage density in three-dimensional optical memory systems [40, 41]. An extremely important factor here is the absence of any interference between radiation propagating in neighboring defects in the PBG structure which has been established in the present study (Section 3.2). This factor can ensure highly local recording and reading of data and avoid any alteration of data stored in adjacent memory cells. Finally, the possibility of locally increasing the field on fairly extended spatial scales in the direction of propagation of the optical beam is potentially useful for enhancing the efficiency of nonlinear optical interactions. However, the nonlinear propagation regimes of optical beams in two-dimensional and three-dimensional photonic crystals require further study.

4. CONCLUSIONS

An analysis of the electromagnetic-field distribution in two-dimensional PBG structures with lattice defects and a study of the spectrum of the defect modes in the band gaps of these structures has revealed various important properties of field localization in photonic crystals. Under these conditions an optical field can be localized in a region smaller than the wavelength and the field at the exit of the PBG structure possesses the properties of a evanescent wave, which means that near-field spatial resolution substantially greater than the wavelength can be achieved. The presence of a dielectric probe sphere of diameter smaller than the wavelength of the incident radiation perturbs the spatial

distribution of the field both inside and beyond the PBG structure and also influences the characteristics of the radiation in the far field. These factors suggest that it is possible to control the optical field distribution in defect modes of photonic crystals and indicates that subdiffraction spatial resolution may be achieved in near-field optical microscopy systems.

The optical beam propagation regimes which can be achieved using PBG structures are potentially useful for solving a wide range of problems in near-field optical microscopy, increasing the data storage density in optical memory devices, and also for enhancing the efficiency of nonlinear optical interactions.

This work would not have been possible without many fruitful and useful discussions with the late N.I. Koroteev who initiated the authors' research into photonic crystals and actively supported this work until the very end.

ACKNOWLEDGMENTS

This research was financed by Constellation Group GmbH. The work of A.M.Z. and A.V.T. was supported in part by INTAS, grant no. 97-0369.

REFERENCES

1. E. Yablonovitch, J. Opt. Soc. Am. B **10**, 283 (1993).
2. *Photonic Band Gaps and Localization*, Ed. by C. M. Soukoulis (Plenum, New York, 1993).
3. *Photonic Band Gap Materials*, Ed. by C. Soukoulis (Kluwer, Dordrecht, 1996).
4. J. Joannopoulos, R. Meade, and J. Winn, *Photonic Crystals* (Princeton Univ., Princeton, 1995).
5. E. Yablonovitch, Phys. Rev. Lett. **58**, 2059 (1987).
6. S. John, Phys. Rev. Lett. **58**, 2486 (1987).
7. J. B. Pendry, J. Phys.: Condens. Matter **8**, 1085 (1996).
8. M. Scalora, R. J. Flynn, S. B. Reinhardt, *et al.*, Phys. Rev. E **54**, 2799 (1996).
9. S.-Y. Lin, E. Chow, V. Hietala, *et al.*, Science **282**, 274 (1998).
10. A. M. Zheltikov, S. A. Magnitskiĭ, and A. V. Tarasishin, Pis'ma Zh. Éksp. Teor. Fiz. **70**, 323 (1999) [JETP Lett. **70**, 323 (1999)].
11. N. Bloembergen and A. J. Sievers, Appl. Phys. Lett. **17**, 483 (1970).
12. J. Martorell, R. Vilaseca, and R. Corbalan, Appl. Phys. Lett. **70**, 702 (1997).
13. M. Scalora, M. J. Bloemer, A. S. Manka, *et al.*, Phys. Rev. A **56**, 3166 (1997).
14. L. A. Golovan', A. M. Zheltikov, P. K. Kashkarov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **69**, 274 (1999) [JETP Lett. **69**, 300 (1999)].
15. A. Stingl, M. Lenzner, Ch. Spielmann, *et al.*, Opt. Lett. **20**, 602 (1995); L. Xu, Ch. Spielmann, F. Krausz, and R. Szipöcs, Opt. Lett. **21**, 1259 (1996).
16. J. C. Knight, T. A. Birks, P. St. J. Russell, and D. M. Atkin, Opt. Lett. **21**, 1547 (1996).
17. M. Scalora, J. P. Dowling, C. M. Bowden, and M. J. Bloemer, Phys. Rev. Lett. **73**, 1368 (1994).
18. P. Tran, Opt. Lett. **21**, 1138 (1996).
19. H.-B. Lin, R. J. Tonucci, and A. J. Campillo, Opt. Lett. **23**, 94 (1998).
20. N. I. Koroteev, S. A. Magnitskii, A. V. Tarasishin, and A. M. Zheltikov, Opt. Commun. **159**, 191 (1999).
21. Shanhui Fan, P. R. Villeneuve, J. D. Joannopoulos, and H. A. Haus, Opt. Express **3**, 4 (1998).
22. M. D. Tocci, M. J. Bloemer, M. Scalora, *et al.*, Appl. Phys. Lett. **66**, 2324 (1995).
23. D. R. Smith, R. Dalichaouch, N. Kroll, *et al.*, J. Opt. Soc. Am. B **10**, 314 (1993); A. Figotin and A. Klein, J. Opt. Soc. Am. A **15**, 1423 (1998).
24. M. Scalora, M. J. Bloemer, A. S. Manka, *et al.*, Phys. Rev. A **56**, 3166 (1997).
25. E. Yablonovitch, J. Phys.: Condens. Matter **5**, 2443 (1993).
26. A. Taflove, *Computational Electrodynamics: The Finite-Difference Time-Domain Method* (Artech House, Norwood, 1995).
27. P. Tran, J. Opt. Soc. Am. B **16**, 70 (1999).
28. A. M. Zheltikov, N. I. Koroteev, S. A. Magnitskiĭ, and A. V. Tarasishin, Kvantovaya Élektron. (Moscow) **25**, 885 (1998).
29. S. Scholz, O. Hess, and R. Rühle, Opt. Express **3**, 28 (1998).
30. R. W. Ziolkowski and M. Tanaka, J. Opt. Soc. Am. A **16**, 930 (1999).
31. B. Engquist and A. Majda, Math. Comput. **31**, 629 (1977).
32. G. Mattei, A. Marucci, V. A. Yakovlev, and M. Pagannone, Laser Phys. **8**, 755 (1998).
33. U. Gruening, V. Lehmann, S. Ottow, and K. Busch, Appl. Phys. Lett. **68**, 747 (1996); V. V. Aristov, A. M. Zheltikov, V. V. Starkov, *et al.*, Laser Phys. **9** (6) (1999).
34. J. G. Fleming and S.-Y. Lin, Opt. Lett. **24**, 49 (1999).
35. K. Sakoda, Phys. Rev. B **52**, 8992 (1995).
36. R. D. Meade, K. D. Brommer, A. M. Rappe, and J. D. Joannopoulos, Appl. Phys. Lett. **61**, 495 (1992).
37. M. A. Paesler and P. J. Moyer, *Near-Field Optics* (Wiley, New York, 1996).
38. V. Sandoghdar, S. Wegscheider, G. Krausch, and J. Mlynek, J. Appl. Phys. **81**, 2499 (1997).
39. M. K. Lewis, P. Wolanin, A. Gafni, and D. G. Steel, Opt. Lett. **23**, 1111 (1998).
40. S. Hunter, F. Kiamilev, S. Esener, *et al.*, Appl. Opt. **29**, 2058 (1990).
41. D. A. Akimov, N. I. Koroteev, A. M. Zheltikov, *et al.*, Jpn. J. Appl. Phys., Part 1 **36**, 426 (1997).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 4, 2000, pp. 609–616.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 4, 2000, pp. 702–709.

ATOMS, SPECTRA, RADIATION

Validity of the Kramers–Henneberger Approximation

O. V. Smirnova

Moscow State University, Vorob'evy gory, Moscow, 119899 Russia
e-mail: smirnova@astra.ilc.msu.su

Received September 29, 1999

Abstract—A formal analogy between the classical method of averages and the Kramers–Henneberger approximate description of the dynamics of atomic systems in a strong monochromatic field is used to determine the limits of validity of this approximation. A unified approach is used to confirm the well-known fact that the Kramers–Henneberger approximation can be applied in the limit of superatomic frequencies and it is shown that this approximation also holds when this condition is violated, i.e., when the field frequency falls to a certain value ω_{cut}, and in this case it can be applied in the limit of superatomic fields. The results are compared with data of numerical and laboratory experiments and with the conditions for validity of the Kramers–Henneberger approximation obtained earlier. The relationship between the validity of the Kramers–Henneberger approximation and the problem of determining the adiabatic stabilization threshold is discussed. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The Kramers–Henneberger (KH) method [1, 2] of approximately describing the dynamics of atomic systems in a strong monochromatic field is interesting in the context of the stabilization of an atom in a strong field. A reduction in the probability of photoionization of an atom compared with the value predicted by the Fermi golden rule is known as the stabilization effect [3–6]. A distinction is usually made between two different types of stabilization corresponding to different manifestations of this effect and observed for different relationships between the parameters of the laser pulse and the atomic system: interference stabilization [4] and adiabatic stabilization [7, 8]. The KH method can be used to unravel the mechanism for adiabatic stabilization. The question of the limits of validity of the KH approximation is interesting from both the methodological and practical point of view. Determining the limits of validity of the KH approximation allows us to determine the thresholds for the adiabatic stabilization effect. With the appearance of the first experiments on stabilization [9, 10] it has become important first, to confirm that the experimental situation falls within the limits of validity of the KH approximation and second, to use this method to describe the experimental dependences.

The idea of the KH method is to apply to the initial Hamiltonian of the system

$$H = \frac{1}{2m}\left(\mathbf{p} - \frac{e}{c}\mathbf{A}\right)^2 + V(\mathbf{x}), \qquad (1)$$

where[1]

$$\mathbf{A} = A_0\mathbf{e}_x\sin(\omega t), \quad A_0 = -Fc/e\omega,$$

the transformation [1]

$$S_{\mathrm{KH}} = \exp\left(\frac{ie}{\hbar cm}\mathbf{p}\int_0^t \mathbf{A}(t')dt'\right)\exp\left(-\frac{ie^2}{2\hbar mc^2}\int_0^t A^2(t')dt'\right),$$

which accomplishes a transition to the oscillating coordinate system

$$\mathbf{x}_{\mathrm{KH}} = \mathbf{x} - \mathbf{e}_x a_e\cos(\omega t). \qquad (2)$$

Here, $a_e = F/m\omega^2$ is the amplitude of the free-electron oscillations in the field F, ω.

In the coordinates (2), the Hamiltonian has the following form:

$$H_{\mathrm{KH}} = i\frac{\partial S_{\mathrm{KH}}^+}{\partial t}S_{\mathrm{KH}} + S_{\mathrm{KH}}^+ H S_{\mathrm{KH}},$$
$$H_{\mathrm{KH}} = \frac{p_{\mathrm{KH}}^2}{2m} + V(\mathbf{x}_{\mathrm{KH}} + \mathbf{e}_x a_e\cos(\omega t)). \qquad (3)$$

In the KH approximation, the time-dependent potential in the Hamiltonian (3) is replaced by the period-averaged value $V_{\mathrm{KH}}(\mathbf{x}, a_e)$, i.e., the KH potential. The KH approximation is valid and productive if the influence of the corrections $\delta V = V(\mathbf{x}_{\mathrm{KH}} + \mathbf{e}_x a_e\cos(\omega t)) - V_{\mathrm{KH}}(\mathbf{x}_{\mathrm{KH}}, a_e)$ is negligible. In this case, all the necessary values such as the ionization rate and the polarizability can be calculated using perturbation theory and the energies of the steady-state levels accurately approximate the exact quasienergies of the system.

[1] Here, we shall confine our analysis to a linearly polarized field.

The dominant trend in studies of the KH approximation is numerical quantum-mechanical calculations [11–18]. So far, the properties of the KH potential, and the properties of the KH eigenfunctions and eigenstates have been comprehensively studied using numerical calculations [7, 11–14]. In particular, as the amplitude of the oscillations increases, $a_e > a$, the KH potential acquires a two-welled structure and is elongated in the direction of the electric field vector of the electromagnetic wave, its characteristic depth decreases, and the number of bound states (for the case of a short-range potential) increases. In this case, the energy of all the eigenstates tends to zero and in the laboratory system this value corresponds to the electron vibrational energy. Thus, the KH states describe effects which have the nature of corrections to the vibrational motion of a free electron in the field of an electromagnetic wave caused by the presence of a weak attractive potential. Adiabatic stabilization is such an effect: an almost-free electron does not absorb (emit) field energy. Stabilization by the KH scenario has been called adiabatic [19] because the KH states appear as a result of the adiabatic evolution of the initial unperturbed state of an atom under the action of a field.

However, although there are no fundamental difficulties involved in interpreting the results obtained using the KH approximation, at present there is no common viewpoint on the limits of validity of this approximation and the role of various parameters.

In the KH formalism, a particular role is assigned to the usual dimensionless combinations of field parameters, initial potential, and KH potential: a/a_e [2], ka_e [5], R [20], $E_{\mathrm{KH}}/\hbar\omega$ [20], and $E/\hbar\omega$ [21]. Here, a is the characteristic dimension of the initial potential, k is the electron wave vector in the continuum, $R = F^2/m\hbar\omega^3$ is the Reiss parameter,[2] E_{KH} is the bound-state energy in the KH potential, and E is the bound-state energy in the initial potential.

The characteristic parameters of the problem can be divided into two groups. We shall consider the group of parameters which controls the transition of an initial atom and "perturbation-theory regime" to a "KH atom" and "stabilization regime": a/a_e, ka_e, and R.

We shall call the limit ($a/a_e \gg 1$, $ka_e \ll 1$, $R \ll 1$) in which the KH potential is converted to the atomic potential, the perturbation-theory regime. This term can be explained by the fact that the main dynamical characteristics of an atom (the polarizability [22] and ionization rate [2]) calculated using the KH Hamiltonian agree with similar values calculated using the initial Hamiltonian (1) using standard perturbation theory [23]. We shall call the stabilization regime the opposite limit ($a/a_e \ll 1$, $ka_e \gg 1$, $R \gg 1$) in which the ionization rate calculated using the KH approximation decreases with increasing intensity [8, 24]. The parameter a/a_e determines the characteristic dimension of the KH potential and its characteristic depth. Large values of the parameter ka_e ensure that the matrix element of the transition from the bound KH state to the continuum state caused by the perturbation δV is small under the conditions of validity of an approximation which uses plane waves as the wave functions of the continuous spectrum in the Kramers system. Thus, the smallness of the parameter $(ka_e)^{-1}$ is an additional stabilizing factor. In all cases of practical interest, we find $ka_e \approx \sqrt{R}$. The Reiss parameter determines the number of harmonics [24] which are important in the stabilization regime $ka_e \gg 1$, $a/a_e \ll 1$. The ranges of field parameters corresponding to large and small values of the dimensionless parameters a/a_e, $(ka_e)^{-1}$, and R^{-1} are the same. However, it is unclear which of these parameters controls the matching of these asymptotic forms of the intensity dependence of the ionization probability. For example, it was shown in [25, 26] that for the hydrogen atom this is the parameter ka_e (or R); however, on the basis of the results of numerical experiments [26] for short-range potentials, it is difficult to unambiguously conclude which of these parameters determines the stabilization threshold. This is because for given potential parameters the thresholds calculated using each of these parameters are very similar, i.e., all three conditions ($a/a_e \ll 1$, $ka_e \gg 1$, $R \gg 1$) begin to be satisfied simultaneously.

The second group of parameters ($E_{\mathrm{KH}}/\hbar\omega$, $E/\hbar\omega$) occurs in the context of the validity of the KH method.

It follows from an analysis of the limiting cases corresponding to the various regimes described above that the validity of the KH approximation does not imply the existence of a stabilization effect, but the occurrence of a stabilization effect may be predicted in the range of field parameters in which the KH approximation is valid and the following conditions are satisfied: $a/a_e \ll 1$, $ka_e \gg 1$, $R \gg 1$.

The limits of validity of the KH approximation are usually associated with the following conditions being satisfied: $E_{\mathrm{KH}}/\hbar\omega \ll 1$ [20], $E/\hbar\omega \ll 1$ [21]. These conditions formalize the concept of the high-frequency nature of the KH approximation which arises because of the obvious analogy between the KH method and the method of averages, and they are equivalent in the case $a/a_e \gg 1$. The condition $E_{\mathrm{KH}}/\hbar\omega \ll 1$ was introduced in [20] when studying the stimulated stopping effect in a strong field in the Kramers system (2) and is usually called the Gavrila–Kaminski condition. However, this condition is not sufficient for the validity of the KH approximation in the stabilization regime. This can be confirmed using the model of a rectangular potential well with parameters V_0 and a. The dependence of the ground-state energy in the KH potential in the stabilization regime on the field parameters for this model was calculated in [27]: $E_{\mathrm{KH}} = V_0 a a_e^{-3/4}$ (in atomic units). Allowing for this dependence, the Gavrila–Kaminski

[2] The Reiss parameter is usually defined as the ratio of the ponderomotive energy to the quantum energy of the field, i.e., $R = F^2/4m\hbar\omega^3$ (see, for example [26]). In the present study, we neglect the numerical difference and retain the term "Reiss parameter" for this combination of field parameters.

condition yields the following result: the lower the field frequency, the lower the intensity at which stabilization occurs. Consequently, the condition $E_{KH}/\hbar\omega \ll 1$ is clearly a necessary but not a sufficient condition for the validity of the KH approximation.

In the context of the classical stabilization problem [6, 28–33], the analogy between the KH method and the method of averages is also used in [31, 33] to estimate the limits of validity of the KH approximation and yields the following result: $\Omega_{KH}(a_e) \ll \omega$, where $\Omega_{KH}(a_e)$ is the frequency of weak oscillations near the minimum of the KH potential. In particular, the following estimate was obtained in [33]:

$$\Omega_{KH}^2(a_e) \approx \max\left|\partial^2 U/\partial x^2\right| \approx a_e^{-1/2}.$$

However, the condition for validity of the KH approximation formulated as $\Omega_{KH}(a_e) \ll \omega$ cannot yield any unambiguous conclusion as to the dependence of the stabilization threshold on the field parameters. This is because stabilization can occur not only as a result of trajectories lying near the minimum of the KH potential but also as a result of trajectories lying above the separatrix of the KH potential and having a different dependence of the classical frequency on a_e, which determines the distance between the levels in the semiclassical limit. In this situation it is more natural to use small parameters which do not depend on the structure of the KH potential.

In the present study the limits of validity of the KH approximation in the semiclassical range are determined using a formal analogy with the classical method of averages [34–36] whose clear limits of validity are established by the Bogolyubov theorem. The attractive potential[3] $-V_0 f(x/a)$ in the field of an electromagnetic wave of frequency ω and strength F/e is used as the model system. The problem has two semiclassical parameters: $R \gg 1$ and $B \gg 1$, where R is the Reiss parameter and $B = \sqrt{2mV_0a^2/\hbar^2}$ is the Born parameter. The presence of two semiclassical parameters means that both essentially quantum ($B \approx 1$) and semiclassical ($B \gg 1$) systems can be considered from common viewpoint. The cases $R \gg 1$ for arbitrary B and $B \gg 1$ for arbitrary R are analyzed.

2. LIMITS OF VALIDITY OF THE KRAMERS–HENNEBERGER APPROXIMATION IN THE CASE $R \gg 1$ FOR ARBITRARY B

The Heisenberg equations for the coordinate and momentum operators $\hat{q}$ and $\hat{p}$ of this particular model system have the following form:

$$\frac{d\hat{q}}{dt} = \frac{\hat{p}}{m} + \frac{F}{m\omega}\sin(\omega t), \quad \frac{d\hat{p}}{dt} = \frac{V_0}{a} f'\left(\frac{\hat{q}}{a}\right). \qquad (4)$$

We write the system (4) in the dimensionless units $m = F = \omega = 1$:

$$\frac{d\hat{q}}{d\tau} = \hat{p} + \sin\tau, \quad \frac{d\hat{p}}{d\tau} = \varepsilon f'\left(\frac{\hat{q}}{\delta}\right).$$

Here, we have $\varepsilon = V_0/aF$, $\hat{q} = \hat{q}(m\omega^2/F)$, $\hat{p} = \hat{p}(\omega/F)$, and $\delta = a/a_e$ ($\hbar = 1/R$). Thus, the semiclassical situation is achieved for $R \gg 1$ and the system dynamics can be described using classical models. In the classical model, we have

$$\frac{dq}{d\tau} = p + \sin\tau, \quad \frac{dp}{d\tau} = \varepsilon f'\left(\frac{q}{\delta}\right).$$

Following the logic of the method of averages [34–36], we impose the constraint that ε is small and using the transformations

$$q_v = q + \cos\tau, \quad p_v = p,$$

$$x_1 = q_v, \quad x_2 = p_v/\sqrt{\varepsilon},$$

we reduce the system to the standard form

$$\frac{dx_1}{d\tau} = \sqrt{\varepsilon}X_1(\varepsilon, x_2, \delta), \quad \frac{dx_2}{d\tau} = \sqrt{\varepsilon}X_2(\varepsilon, x_1, \delta).$$

Here, we have

$$X_1(\varepsilon, x_2, \delta) = x_2,$$

$$X_2(\varepsilon, x_1, \delta) = f'\left(\frac{x_1 - \cos\tau}{\delta}\right).$$

We now apply the method of averages subject to the condition $\sqrt{\varepsilon} \ll 1$. The system in the first approximation has the following form:

$$\frac{du_1}{d\tau} = \sqrt{\varepsilon}u_2, \quad \frac{du_2}{d\tau} = \sqrt{\varepsilon}X_0(\varepsilon, u_1, \delta),$$

where u_1 and u_2 are the components of the column vector u which is the average solution in the first approximation. Here, we have

$$X_0(\varepsilon, u_1, \delta) = \frac{1}{2\pi}\int_0^{2\pi} f'\left(\frac{u_1 - \cos\tau}{\delta}\right)d\tau.$$

Returning to the variables q_v and p_v and the average solutions corresponding to them

$$\bar{q}_v = u_1, \quad \bar{p}_v = u_2\sqrt{\varepsilon},$$

[3] In the present study, for simplicity all the estimates are made for one-dimensional systems. The results can be generalized to the three-dimensional case. In this case, they do not change qualitatively.

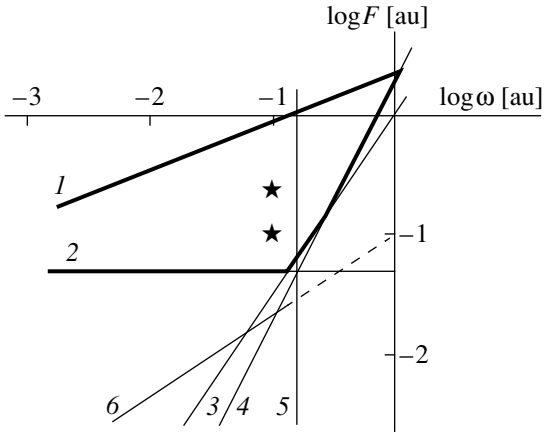

Fig. 1. Region (indicated by heavy lines) of field parameters in which the KH approximation holds in the quantum case $B \approx 0.8$ ($V_0 = 2.5$ eV, $a = 0.1$ nm). The stars indicate the values of the field parameters ($\hbar\omega = 2.5$ eV, $P_1 = 3 \times 10^{15}$, $P_2 = 10^{16}$ W/cm^2) for which stabilization was observed in the numerical experiments [24]. Notation: (*1*) the condition $V_0/a = F_L$; (*2*) $\varepsilon = 1$ ($V_0/a = F$); (*3*) $R = 1$; (*4*) $\delta = 1$; (*5*) $\omega = \Omega$; and (*6*) Gavrila–Kaminski condition $E_{\mathrm{KH}}/\hbar\omega = 1$.

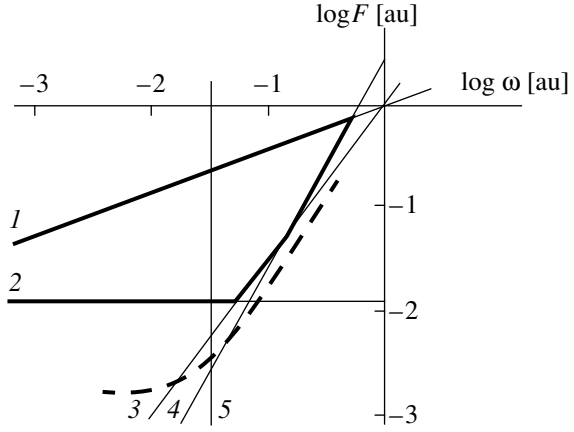

Fig. 2. Region (indicated by heavy lines) of field parameters in which the KH approximation holds in the quantum case $B \approx 0.6$ ($V_0 = 0.75$ eV, $a = 0.125$ nm). Notation: (*1*) the condition $V_0/a = F_L$; (*2*) $\varepsilon = 1$ ($V_0/a = F$); (*3*) $R = 1$; (*4*) $\delta = 1$; and (*5*) $\omega = \Omega$. The dashed line indicates the lower limit of the range of validity of the KH approximation determined numerically in [37, Fig. 10].

we obtain a system which is the classical analog of the KH approximation:

$$\frac{d\bar{q}_v}{d\tau} = \bar{p}_v, \quad \frac{d\bar{p}_v}{d\tau} = \varepsilon f'\left(\frac{\bar{q}_v - \cos\tau}{\delta}\right).$$

Thus, after reducing the initial system to the standard form, it becomes clear that the main criterion for the validity of the method of averages and therefore for the KH method in the range $R \gg 1$ is the condition $\sqrt{\varepsilon} \ll 1$. This condition begins to be satisfied when the field strength exceeds the characteristic force scale of the system, i.e., in superatomic fields.

However, the time interval in which similarity between the exact and average solutions is guaranteed is also of interest from the practical point of view. This is the condition for smallness of the classical response of a "KH atom" to the perturbation induced by KH harmonics, which determines the difference in the semiclassical limit between the exact value of the quasienergy and the corresponding value calculated using the KH approximation. The constraint $t_{\max} \gg T$, where $T = 2\pi/\omega$ is the period of the field, can introduce additional constraints on the range of permissible field parameters. This question can be answered for example using the Bogolyubov theorem which delineates the limits of validity of the method of averages. Relevant estimates are given in the Appendix. The conditions of validity of the method (in atomic units: $m = e = \hbar = 1$), supplemented by the condition for the absence of relativistic drift (the characteristic force scale is larger than the Lorenz force F_L) determine the range of field parameters in which the KH approximation is valid in the case $R \gg 1$:

(1) $F \gg \omega^{3/2}$ ($R \gg 1$);

(2) $F \gg V_0/a$ ($\varepsilon \ll 1$);

(3) $F \gg a\omega^2$ ($\delta \ll 1$);

(4) $F \ll \left(\frac{V_0}{a}\right)^{1/2} \frac{\omega^{1/2}}{\alpha^{1/2}}$ $\left(\frac{V_0}{a} \gg F_L\right)$

(α is the fine-structure constant).

The range of field parameters satisfying these four conditions is shown in Fig. 1 for a potential having the characteristic parameters[4] $V_0 = 2.5$ eV, $a = 0.1$ nm, and $B \approx 0.8$, used in the numerical experiments in [24]. We note that in this range the Gavrila–Kaminski condition $E_{KH}/\hbar\omega$ is satisfied which, for this particular model in the range $\delta \ll 11$, has the form $F \gg (V_0/a)^{4/3}\omega^{2/3}$ (see Introduction). The stars denote the field parameters corresponding to the stabilization regime in this numerical experiment. The range of field parameters satisfying the conditions given above for a potential having the characteristic parameters $V_0 = 0.75$ eV, $a = 0.125$ nm, and $B \approx 0.6$ used in the numerical experiments in [37] is shown in Fig. 2. The dashed line gives the lower limit of the range of validity of the KH approximation determined numerically in [37, Fig. 10].

[4] Here and subsequently, the ionization potential of the particular state of the unperturbed system is taken as the characteristic energy scale V_0.

3. LIMITS OF VALIDITY OF THE KRAMERS–HENNEBERGER APPROXIMATION FOR THE CASE $B \gg 1$

In this section, we use the scheme for determining the limits of validity of the KH approximation given above for the case $R \gg 1$. Writing the system (4) in dimensionless units $m = V_0 = a = 1$,

$$\frac{d\hat{q}}{d\tau} = \hat{p} + \frac{F}{\omega}\sin\frac{\tau}{\varepsilon}, \quad \frac{d\hat{p}}{d\tau} = f'(\hat{q}),$$

where $\varepsilon = \Omega/\omega$, $\Omega = \sqrt{V_0}/a$ ($\hbar = 1/B$), we find that the semiclassical case takes place for $B \gg 1$. In the classical model, we have

$$\frac{dq}{d\tau} = p + \frac{F}{\omega}\sin\frac{\tau}{\varepsilon}, \quad \frac{dp}{d\tau} = f'(q). \tag{5}$$

The system (5) is reduced to the standard form of the method of averages

$$\frac{dx_1}{d\theta} = \varepsilon X_1\left(\varepsilon, x_2, \frac{F}{\omega^2}\right),$$

$$\frac{dx_2}{d\theta} = \varepsilon X_2\left(\varepsilon, x_1, \frac{F}{\omega^2}\right)$$

by making the substitution

$$\theta = \frac{\tau}{\varepsilon}, \quad x_1 = x + \frac{F}{\omega^2}\cos\frac{\tau}{\varepsilon}, \quad x_2 = p.$$

Here, we have

$$X_1\left(\varepsilon, x_2, \frac{F}{\omega^2}\right) = x_2,$$

$$X_2\left(\varepsilon, x_1, \frac{F}{\omega^2}\right) = f'\left(x_1 - \frac{F}{\omega^2}\cos\theta\right).$$

Applying the method of averages under the condition $\varepsilon \ll 1$, we find that the system in the first approximation

$$\frac{du_1}{d\theta} = \varepsilon u_2, \quad \frac{du_2}{d\theta} = \varepsilon X_0\left(\varepsilon, u_1, \frac{F}{\omega^2}\right),$$

where

$$X_0\left(\varepsilon, u_1, \frac{F}{\omega^2}\right) = \frac{1}{2\pi}\int_0^{2\pi} f'\left(u_1 - \frac{F}{\omega^2}\cos\theta\right)d\theta,$$

is the classical analog of the KH approximation.

As in the previous case, estimating the time interval for which similarity between the exact and average solutions is guaranteed (see the Appendix), we find that the range of field parameters for which the KH approximation holds for $B \gg 1$ is limited by the following constraints:

(1) $\omega \gg \Omega \quad (\varepsilon \ll 1)$;

(2) $F \ll \left(\frac{V_0}{a}\right)^{1/2}\frac{\omega^{1/2}}{\alpha^{1/2}} \quad \left(\frac{V_0}{a} \gg F_L\right)$.

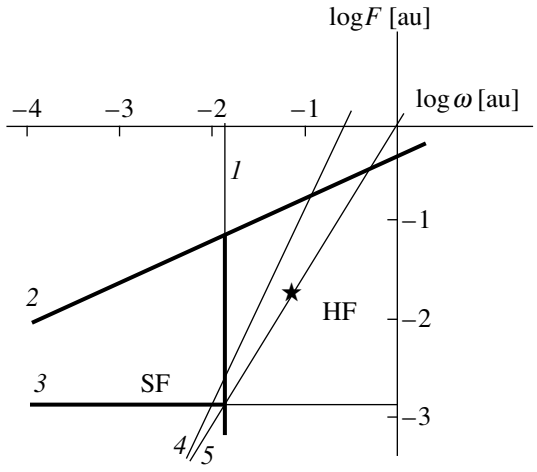

Fig. 3. Limits of validity of the KH method in the semiclassical case $B \gg 1$ ($V_0 = 0.54$ eV, $a = 1.25$ nm). The star indicates the field parameters ($\hbar\omega = 2$ eV, $P = 10^{14}$ W/cm^2) for which stabilization was observed experimentally [10]. The letters SF denote the region (bounded by lines *1*, *3*, and *4*) of field parameters in which the KH approximation holds in the strong-field limit ($F \gg V_0/a$). The letters HF indicate the region (bounded by lines *1* and *3*) of field parameters in which the KH approximation holds in the high-frequency limit ($\omega \gg \Omega$). Notation: (*1*) $\omega = \Omega$ ($\varepsilon = 1$ for HF); (*2*) $V_0/a = F_L$; (*3*) $V_0/a = F$ ($\delta = 1$ for SF); (*4*) $\delta = 1$; and (*5*) $R = 1$.

We shall consider a potential having the parameters $V_0 = 0.54$ eV, $a = 1.25$ nm ($B \approx 7$) which simulates the experimental situation [10]. The range of field parameters satisfying these two conditions is shown in Fig. 3 and is denoted by HF. Since the results of the previous section are valid for $B \gg 1$, the range of field parameters satisfying the four conditions given in Section 3 is also shown in Fig. 3 and is denoted by SF. The star indicates the parameters of an adiabatic stabilization experiment [10].

4. CONCLUSIONS

A formal analogy between the method of averages and the KH method has been used to determine the limits of validity of the latter for essentially quantum ($B \approx 1$) for $R \gg 1$ (Figs. 1, 2) and semiclassical ($B \gg 1$) (Fig. 3) systems.

A unified approach has been used to confirm the well-known fact that the KH approximation can be applied in the high-frequency limit ($\omega \gg \Omega$) and it is shown that this approximation also works when the above condition is not satisfied and the field frequency decreases to $\omega_{cut} = V_0\alpha/a$. In this case, it can be applied

in the superatomic field limit ($F \gg V_0/a$) subject to an additional constraint being satisfied for the Reiss parameter: $R \gg 1$.

This conclusion agrees with the results of numerical experiments on ionization in a superatomic low-frequency field [37]. In [37] the limits of validity of the KH approximation were analyzed numerically and it was shown that in a low-frequency field there is a threshold field intensity E_{BSI} close to the characteristic "atomic" intensity for this model ($E_{BSI} \approx V_0/7a$) and almost independent of the frequency above which the KH approximation becomes valid (see Fig. 2).

Thus, generalizing all these factors, we can conclude that the threshold of the adiabatic stabilization effect depends on the field and system parameters as follows:

$$\frac{\Omega}{\omega} \le 1, \quad \min\left\{\frac{a}{a_e}, \frac{1}{R}\right\} \le 1 \quad \text{for} \quad \frac{V_0}{aF} > 1;$$

$$\frac{V_0}{aF} \le 1, \quad \max\left\{\frac{a}{a_e}, \frac{1}{R}\right\} \le 1 \quad \text{for} \quad \frac{\Omega}{\omega} > 1, \quad \omega > \omega_{\text{cut}};$$

$$\omega_{\text{cut}} = V_0\alpha/a.$$

From the methodological point of view, it is interesting to develop the quantum–classical analogy still further in order to enrich the tools of quantum mechanics with well-developed methods of classical mechanics to solve quantum-mechanical problems in the semiclassical range. We note various aspects which arise in the course of implementing this program for the method of averages.

(1) Determining the quantum analog of the system in the second approximation of the method of averages (see, for example [36]) and using this to construct corrections to the KH approximation in perturbation theory for the case $\delta \ll 1$ when the KH potential and the correction term δV (see Section 1) are of the same order of smallness.

(2) Calculating the shift of the KH levels as a result of the corrections δV using the system in the nth approximation of the method of averages and the semiclassical quantization rules.

(3) Calculating the spacings between the KH levels with a given accuracy in terms of the small parameters $1/R$ and $1/B$ using $\hbar$ expansions of the transition quantum frequencies [38].

ACKNOWLEDGMENTS

The author thanks P.V. Elyutin for discussions on the formulation of the problem, the results, and useful comments during the work. The author is also grateful to A.M. Popov for useful discussions and for pointing out the qualitative agreement between the results of the present study and the results of numerical experiments [37]. The author also thanks R.V. Karapetyan and M.V. Fedorov for useful comments.

This work was supported by the Russian Foundation for Basic Research (project no. 98-02-16670).

APPENDIX

Estimates are made using the Bogolyubov theorem [36] in which two initial Cauchy problems are compared:

$$\frac{dx}{dt} = \varepsilon X(t, x, \varepsilon), \quad x(0) = x_0,$$

and

$$\frac{du}{dt} = \varepsilon X_0(u, \varepsilon), \quad u(0) = x_0,$$

where

$$X_0 = \frac{1}{T}\int_0^T X(t, u, \varepsilon)dt,$$

$x(t)$ is the exact solution, and $u(t)$ is the average solution of the first approximation.

If the function $X(t, x, \varepsilon)$ satisfies the physically obvious conditions of boundedness of the function and its gradient[5] $\|X(t, x, \varepsilon)\| \le M$, $\|X(t, x, \varepsilon) - X(t, x', \varepsilon)\| \le \lambda\|x - x'\|$, and is periodic with the period T, for times $0 \le t \le L/\varepsilon$, the degree of similarity of the exact and average solutions is determined by the inequality $\|x(t) - u(t)\| \le TM(L\lambda + 2)\exp(\lambda L)\varepsilon$. (A rigorous formulation of the theorem can be found, for example, in [36]).

(1) The case $R \gg 1$ for arbitrary B. In order to construct estimates, it is convenient to analyze the right-hand sides not over the entire range of determination but only in the $\rho(\varepsilon)$-vicinity of the average solution $\{u_1 \pm \rho, u_2 \pm \rho\}$. We shall first consider the case $\delta \ll 1$.

Estimating $\|X\| = \sqrt{X_1^2 + X_2^2}$, we determine the constant M:

$$M = \delta\left(\frac{2\sqrt{1+\pi}}{\pi}\right).$$

Here, it is assumed that

$$u_2^2 = \frac{p_{\text{KH}}^2}{\varepsilon} = 2\frac{V_{\text{KH}} - E_{\text{KH}}}{\varepsilon},$$

$$V_{\text{KH}} \approx \frac{2\delta V_0}{\pi E_0} = \frac{2\varepsilon\delta^2}{\pi},$$

$$|f'|^2 \le \left(\frac{2\delta}{\pi}\right)^2, \quad E_0 = \frac{F^2}{m\omega^2}.$$

It is also assumed that when the pulse is switched on, states lying above the separatrix of the KH potential are

[5] Here, we have $\|x\| = \sqrt{\sum_i x_i^2}$.

preferentially populated [24]. The fact that the norm $\|x(t) - u(t)\|$ is equal to ρ is used to determine the value of ρ together with M as a function of ε, R, δ.

Similarly to determine the constant λ we estimate

$$\left\|\frac{\partial X}{\partial x}\right\| = \sqrt{\left(\frac{\partial X_1}{\partial x_1}\right)^2 + \left(\frac{\partial X_1}{\partial x_2}\right)^2 + \left(\frac{\partial X_2}{\partial x_1}\right)^2 + \left(\frac{\partial X_2}{\partial x_2}\right)^2}.$$

We have

$$\lambda = \frac{1}{\pi}\sqrt{\pi^2 + 4}.$$

In this case, it is most convenient to set

$$L = \varepsilon^{1/4}/\lambda,$$

then subject to the condition $L\lambda \ll 1$ in the time interval

$$0 \le t \le \frac{\pi}{\varepsilon^{1/4}\sqrt{\pi^2 + 4}},$$

the Bogolyubov theorem gives

$$\|x(t) - u(t)\| \le 8\sqrt{1 + \pi}\delta\varepsilon^{1/2}.$$

For the variables in the Kramers system q_v and p_v and the corresponding average solutions

$$\bar{q}_v = u_1, \quad \bar{p}_v = u_2\varepsilon^{1/2}$$

we have the estimate

$$\sqrt{|q_v - \bar{q}_v|^2 + |p_v - \bar{p}_v|^2} \le 8\sqrt{1 + \pi}\delta\varepsilon^{1/2}.$$

Note that the region $\delta \gg 1$ cannot be analyzed adequately in the method of averages since the semiclassical parameter R is much less than unity in this range.

(2) The case $B \gg 1$ for arbitrary R. Performing similar estimates for the cases $\delta \ll 1$ and $\delta \gg 1$ we obtain

$$M = \sqrt{2 + 2\rho + \rho^2},$$

$$\lambda = \sqrt{2}.$$

We set $L = \varepsilon^{1/2}/\lambda$ and then in the time interval $0 \le \theta \le 1/\sqrt{2\varepsilon}$, we have $\|x(t) - u(t)\| \le CM\varepsilon^2$ where $C = 4\pi$. In the three-dimensional case, we have $M = \sqrt{6}$, $\lambda = \sqrt{6}$, $0 \le \theta \le 1/\sqrt{6\varepsilon}$.

REFERENCES

1. H. A. Kramers, *Les particules elementaires, Report to the Eighth Solvay Conference* (Editions Stoops, Brussels, 1950).
2. W. C. Henneberger, Phys. Rev. Lett. **21**, 838 (1968).
3. J. I. Gersten and M. N. Mittleman, Phys. Rev. A **10**, 74 (1974).
4. M. V. Fedorov, *Atomic and Free Electrons in a Strong Light Field* (Nauka, Moscow, 1991; World Sci., Singapore, 1997).
5. K. Burnnett, V. C. Reed, and P. L. Knight, J. Phys. B **26**, 561 (1993).
6. N. B. Delone and V. P. Kraĭnov, Usp. Fiz. Nauk **165**, 1295 (1995) [Phys. Usp. **38**, 1247 (1995)].
7. M. Pont, N. R. Walet, M. Gavrila, and C. W. McCrudy, Phys. Rev. Lett. **61**, 939 (1988).
8. M. Pont and M. Gavrila, Phys. Rev. Lett. **65**, 2362 (1990).
9. M. P. de Boer, J. H. Hoogenraad, R. B. Vrijen, *et al.*, Phys. Rev. A **50**, 4085 (1994).
10. N. J. van Druten, R. S. Constantinescu, J. M. Schins, *et al.*, Phys. Rev. A **55**, 622 (1997).
11. M. Pont and M. Gavrila, Phys. Lett. A **123**, 469 (1987).
12. M. Pont, Phys. Rev. A **40**, 5659 (1989).
13. M. Pont, N. R. Walet, and M. Gavrila, Phys. Rev. A **41**, 477 (1990).
14. Q. Su and J. H. Eberly, Phys. Rev. A **43**, 2474 (1991).
15. K. C. Kulander, K. J. Schafer, and J. L. Krause, Phys. Rev. Lett. **66**, 2601 (1991).
16. R. Grobe and M. V. Fedorov, J. Phys. B **26**, 1181 (1993).
17. A. M. Popov, O.V. Tikhonova, and E. A. Volkova, Laser Phys. **5**, 1184 (1995).
18. E. A. Volkova, A. M. Popov, and O. V. Tikhonova, Zh. Éksp. Teor. Fiz. **109**, 1586 (1996) [JETP **82**, 853 (1996)].
19. R. J. Vos and M. Gavrila, Phys. Rev. Lett. **68**, 170 (1992).
20. M. Gavrila and J. Z. Kaminski, Phys. Rev. Lett. **52**, 613 (1984).
21. J. I. Gersten and M. N. Mittleman, J. Phys. B **9**, 2561 (1976).
22. A. M. Popov, O. V. Tikhonova, and E. A. Volkova, Laser Phys. **10** (1), (2000) (in press).
23. N. B. Delone and V. P. Kraĭnov, *Atom in a Strong Light Field* (Énergoatomizdat, Moscow, 1984).
24. E. A. Volkova, A. M. Popov, and O. V. Smirnova, Zh. Éksp. Teor. Fiz. **106**, 1360 (1994) [JETP **79**, 736 (1994)].
25. M. Pont and R. Shakeshaft, Phys. Rev. A **44**, R4110 (1991).
26. R. M. Potvliege and P. H. G. Smith, Phys. Rev. A **48**, R46 (1993).
27. E. A. Volkova, A. M. Popov, O. V. Smirnova, and O. V. Tikhonova, Zh. Éksp. Teor. Fiz. **111**, 1194 (1997) [JETP **84**, 658 (1997)].

28. J. Grochmalicki, M. Lewenstein, and K. Rzazewski, Phys. Rev. Lett. **66**, 1038 (1991).
29. R. Grobe and C. K. Law, Phys. Rev. A **44**, 4114 (1991).
30. B. Sundaram and R. V. Jensen, Phys. Rev. A **47**, 1415 (1993).
31. F. Benvenuto, G. Casati, and D. L. Shepelyansky, Phys. Rev. A **47**, R786 (1993).
32. G. Casati, I. Guarneri, and G. Mantica, Phys. Rev. A **50**, 5018 (1994).
33. R. V. Karapetyan, submitted to Laser Phys.
34. N. N. Bogolyubov and Yu. A. Mitropol'skii, *Asymptotic Methods in the Theory of Nonlinear Oscillations* (Nauka, Moscow, 1974, 4th ed.; Gordon and Breach, New York, 1962).
35. A. H. Nayfeh, *Perturbation Methods* (Wiley, New York, 1973; Mir, Moscow, 1976).
36. V. F. Zhuravlev and D. M. Klimov, *Applied Methods in the Vibration Theory* (Nauka, Moscow, 1988).
37. A. M. Popov, O. V. Tikhonova, and E. A. Volkova, J. Phys. B **32**, 3331 (1999).
38. P. V. Elyutin and O. V. Smirnova, Teor. Mat. Fiz. **199**, 93 (1999).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 4, 2000, pp. 617–622.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 4, 2000, pp. 710–716.

FLUIDS

Electrical Conductivity of Water during Quasi-Isentropic Compression to 130 GPa

V. V. Yakushev*, V. I. Postnov, V. E. Fortov, and T. I. Yakysheva

Institute for Chemical Physics Research, Russian Academy of Sciences, Chernogolovka, Moscow oblast, 142432 Russia
**e-mail: yakushev@icp.ac.ru*

Received October 22, 1999

Abstract—The specific electrical conductivity Σ of water was measured during multistage shock loading to pressures of 12–130 GPa. At maximum pressure the density of the water was 3.2 g/cm^3. Three or four pressure discontinuities could usually be resolved experimentally and the value of Σ was determined in each of these. As the pressure was increased in this range, the value of Σ increased from 1.2 to approximately 150 S/cm. In electrochemical experiments, galvanic cells having electrodes of various metals and water as the electrolyte were subjected to dynamic compression. The characteristics of the recorded emf of these cells indicate that the high electrical conductivity of highly compressed water is of an ionic nature. *© 2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Water is one of the most abundant substances on the earth and in the universe. In particular, along with methane and ammonia it forms the main component of the inner shells of the giant planets in the solar system, Neptune and Uranus, occurring in the liquid phase at temperatures up to several thousand K and pressures up to hundreds of GPa [1–4]. It is quite possible that the electrical conductivity of highly compressed and heated water is responsible for the magnetic field of these planets [5].

The physical properties of water under these extreme conditions have recently attracted detailed study. In particular, *ab initio* calculations of its phase diagram using molecular dynamics [5] have shown that at pressures above 30 GPa and temperatures higher than 2000 K, water should be in a superionic state characterized by anomalously high proton mobility. The question of the predicted [5] transition of water to the metallic state under the action of high pressures and temperatures by analogy with molecular hydrogen [6] is also fundamentally unresolved. An experimental check of these conclusions by means of direct measurements of the electrical conductivity of water in the appropriate regions of the phase diagram is undoubtedly of interest.

It should be noted that the only source of experimental information on the properties of liquid water at high pressures and temperatures are experiments using powerful shock waves, which have already yielded unique data on its equation of state up to pressures of around 100 GPa using explosives as the energy source [7], up to 230 GPa in experiments using shock waves generated by the planar impact of a special impactor accelerated by a two-stage light-gas gun [8], and at pressures of approximately 100 GPa [9] and 1425 GPa [10] in experiments using underground nuclear explosions.

Details of the structure of shock-compressed water have been studied experimentally by spontaneous Raman scattering [11] and have also been considered theoretically [5, 12]. It was shown in particular that as the dynamic pressure increases, the fraction of hydrogen-bonded hydroxyl groups decreases [12].

Results of the first measurements of the electrical conductivity Σ of water beyond a shock wave front were published in 1960 by David and Hamann [13, 14] who observed a sharp increase in conductivity as the dynamic pressure increased from 2 to 13 GPa and attributed this to an increase in the degree of dissociation of the water molecules to form ions. At almost the same time the authors of [15] made independent measurements of Σ for water at a dynamic pressure of 10 GPa and obtained a value of 0.2 S/cm in good agreement with the data in [13, 14]. Slightly later [16] refined experimental values of the conductivity were given at dynamic pressures up to approximately 22 GPa. Measurements of the molar electrical conductivity of aqueous solutions of KCl, KOH, and HCl in the range 7–13.3 GPa [17] were used to calculate the ionic product of water at these pressures, and these authors concluded that above approximately 15–20 GPa the water beyond the shock wave front is almost completely dissociated to form ions and has an electrical conductivity of 10 S/cm. The ionic nature of the electrical conductivity of dynamically compressed water is also supported by its optical transparency, observed in [18] at pressures up to 30 GPa, and also by the good agreement between the dynamic and static [19] measurements of the electrical conductivity made at similar temperatures and pressures.

Finally, using the two-stage light-gas gun at the Lawrence Livermore National Laboratory, Mitchell

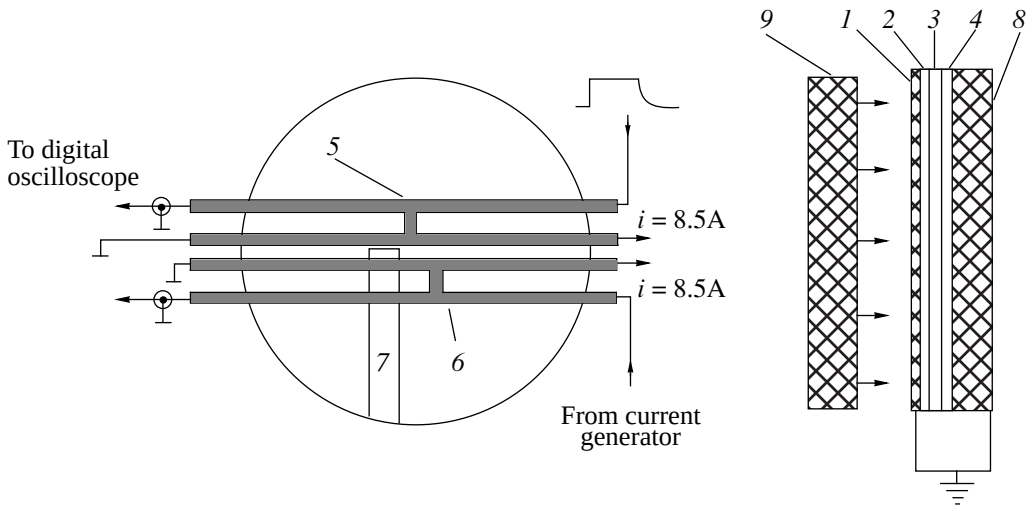

Fig. 1. Schematic of experimental apparatus: (*1*) base plate, (*2–4*) polyethylene layers, (*5*) manganin pressure gauge, (*6*) manganin gauge functioning as shunt, (*7*) water sample, (*8*) reflector plate.

and Nellis succeeded in making accurate measurements of the electrical conductivity of water at pressures of 25–59 GPa beyond the shock wave front [8].

The aim of the present study was to further substantially expand the range of pressures and compression in experiments to measure the electrical conductivity of water and this was accomplished by using a multistage dynamic loading regime in which the sample is exposed to the action of a series of successive shock waves circulating between two plane-parallel plates having a substantially higher dynamic rigidity than that of water. As a result, we succeeded in measuring Σ for water up to pressures of 130 GPa and density of 3.2 g/cm^3. The ionic nature of the electrical conductivity of water under these experimental conditions was confirmed by recording the electrochemical potentials.

2. EXPERIMENTAL RESULTS

2.1. Electrical Conductivity

Characteristic features of a dynamic experiment such as the need to have the measuring equipment some distance from the site of the experiment and the short lifetime of the substance in the compressed state (10^{-8}–10^{-6} s) necessitate using special methods to measure the electrical conductivity [20]. In the present study we used a modified dc electrical circuit [20–22].

The general experimental setup and the design of the measuring cell are shown schematically in Fig. 1. In the experiments we used two independent measuring channels supplied by a generator having symmetric floating outputs. The generator delivered rectangular current pulses of 8.5 A into low-resistance loads a few tens of microseconds before the time of entry of the shock wave into the sample. This specifically eliminated any electrolysis of the water and electrical polarization of the electrodes. The generator was triggered by a sensor located in the explosive charge.

The cell was constructed using low-density polyethylene (0.928 g/cm^3) as the electrical insulator. This material was chosen for various reasons, the main one being its fairly low electrical conductivity under the experimental conditions ($\Sigma < 10^{-5}$ S/cm) which was estimated in special preliminary experiments. Having a nonpolar structure of molecular units, polyethylene also exhibits low shock-induced electrical polarization [23] which means that a fairly low level of electrical noise can be obtained experimentally. Finally, the dynamic compressibility of polyethylene is very close to the compressibility of water [24]. Thus, the shock wave passes across the polyethylene/water interface with almost no reflections.

The cell was fabricated by successively depositing three layers of polyethylene (*2–4*), each between 0.5 and 1 mm thick, by hot pressing onto a polished base plate *1* made of 12Kh18N10T stainless steel, 120 mm in diameter and 2 mm thick. Between the first and second layers from the base we inserted two H-shaped piezoelectric pressure gauges (*5*, *6*). These were made of 35 μm thick manganin foil electroplated with a 5 μm thick layer of copper over the entire surface area, except for the bridges which formed the sensitive elements. The width of the sensor outputs was 2 mm and the distance between the outputs 8 mm. During the pressing process particular attention was paid to eliminating air gaps and bubbles from the polyethylene. The electrodes of one of the sensors (*6*) were in electrical contact with the sample *7*, as shown in Fig. 1. Thus, the manganin bridge of this sensor functioned as a shunting resistor in the measuring circuit. The geometric dimensions of the water sample were defined by a rectangular cut of

width 10 mm in the middle layer of polyethylene. In other words, the sample was positioned in a cavity inside a polyethylene insulator.

At the final assembly stage, a 5 mm thick stainless steel or copper reflector plate *8* was installed and the complete system was tightened with bolts. The sample cavity was filled with doubly distilled water immediately before the experiment.

The dynamic loading of the sample was provided by a series of shock waves circulating between the base *1* and the reflector *8* which were initiated by the planar impact of a striker plate *9* against the base. The striker plate, having a diameter of 60 or 90 mm and thickness between 2.5 and 5 mm, was also made of stainless steel. These striker plates were accelerated by standard explosives and their maximum velocity was 4.5 km/s at 2.5 mm thickness.

Positioning an electrical shunt in the immediate vicinity of the sample in the high-pressure zone minimized the influence of the stray inductance of the circuit and improved the time resolution of the method to 10–20 ns [21, 22]. In this case, the change in the resistance of the shunt with pressure was calculated using the readings from the pressure gauge *5*.

The recorded change in the resistance dR of the manganin sensor was converted to pressure p using the following formula:

$$p[\text{GPa}] = 34(dR/R_0) + 7(dR/R_0)^3,$$

where R_0 is the initial resistance of the sensor bridge. This dependence was first used to describe the results of calibration experiments carried out at pressures up to 40 GPa [25]. Its validity at pressures up to 110 GPa was checked by one of the present authors [26].

The experimentally achieved time dependence of the pressure in the samples was also calculated using a one-dimensional code based on the Hugoniot equations of state for water, polyethylene, and steel [24], the cell geometry, and the striker flight velocity.

The densities achieved during compression of the samples, which are required to calculate the resistivity, were calculated using a more accurate equation of state for water [7] constructed using experimental data on the compressibility of water and ice of different initial density in the pressure range 3–120 GPa. Note that in particular, this equation of state was used to accurately describe the results of later measurements [8] of the single-stage and two-stage compressibility of water at pressures up to 230 GPa.

As an example, Fig. 2 shows a calculated sequence of water states plotted in terms of pressure p versus mass velocity u for an experiment with an impact velocity of 3.78 km/s. It can be seen that for a first-shock amplitude of 21 GPa, the final pressure in the sample is achieved after six or seven reflections and is 110 GPa.

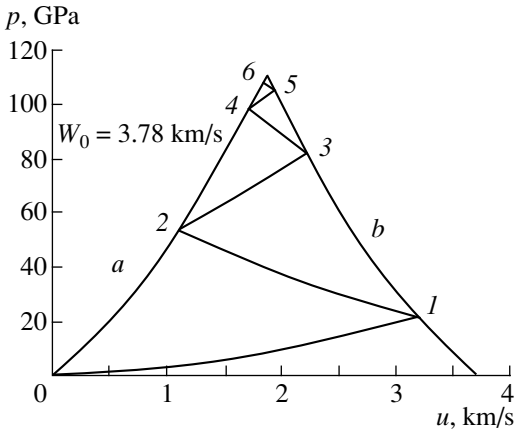

Fig. 2. Example of p–u diagram of the multistage compression of water in the experiment: (*a*, *b*) shock adiabat and unloading adiabat of the stainless steel base and reflector material; the numbers denote the sequence of water states.

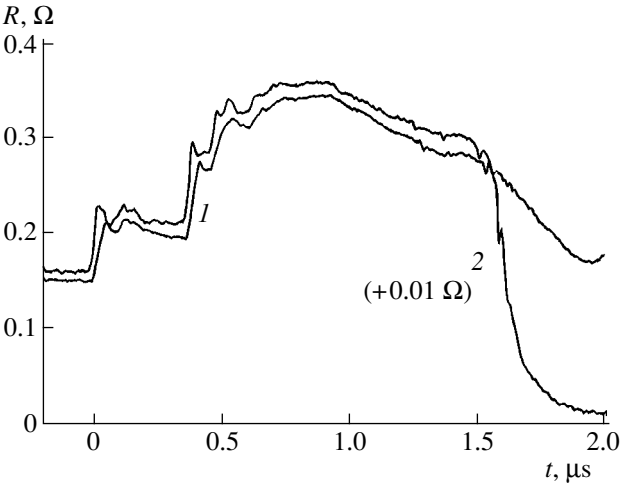

Fig. 3. Dependences $R(t)$ for eicosane experiment: (*1*) change in the resistance of the pressure gauge, (*2*) change in resistance in the sample channel.

The correspondence between the change in the electrical resistance $R(t)$ of the pressure gauge *5* and the shunt *6* was checked in a special experiment where the sample cavity was filled with a saturated hydrocarbon, eicosane, which is a good insulator under the experimental conditions. The corresponding experimental dependences obtained for a final pressure of 63 GPa are plotted in Fig. 3. As was predicted, the resistances of the pressure gauge and the shunt change fairly synchronously.

Figure 4 shows similar dependences $R(t)$ recorded in an experiment using water. It can be seen that at the first pressure jump the resistance of the water sample becomes comparable with the resistance of the shunt, which is reflected in an appreciable drop in the overall resistance of the shunt–sample system. Subsequent

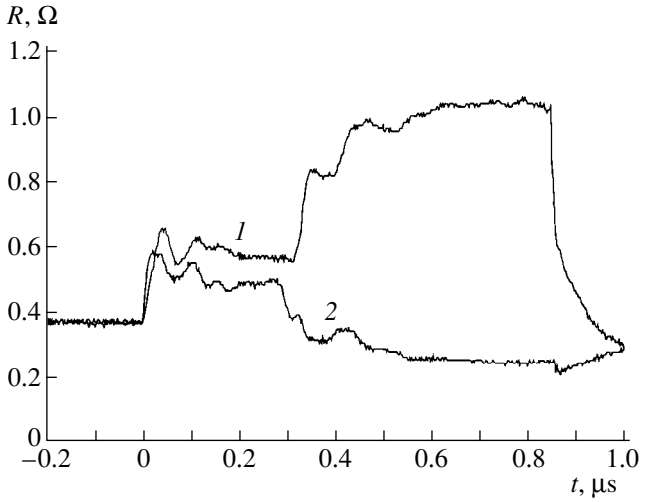

Fig. 4. Typical dependences $R(t)$ obtained in an experiment using water at the maximum pressure ~100 GPa: (*1*) change in the resistance of the pressure gauge, (*2*) change in resistance in the sample channel.

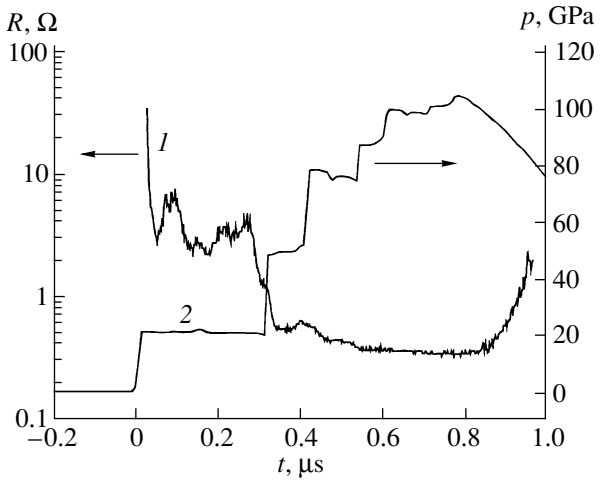

Fig. 5. Analysis of experimental results plotted in Fig. 4: (*1*) change in the resistance of the water sample during compression, (*2*) dynamic pressure profile in the sample.

pressure jumps lead to an even larger decrease in the sample resistance and cause the curves *1* and *2* to diverge. Results of an analysis of the experimental data from Fig. 4 in the form of the time dependence of the electrical resistance of the water sample and the calculated pressure profile in the water are plotted in Fig. 5. It can be seen that in this experiment we can immediately identify five points on the dependence of the sample resistance on the dynamic pressure during multistage compression. In other experiments we usually resolved between three and five pressure jumps, and therefore electrical resistances, of the samples.

Figure 6 gives the experimental results plotted as the electrical conductivity Σ of water as a function of the pressure p. Also plotted for comparison are data from earlier studies [8, 16]. It can be seen than from a dynamic pressure of ~30 GPa the value of Σ depends weakly on p, which following [16] can be attributed to the complete dissociation of water to give ions.

2.2. Electrochemical Potentials

Electrochemical phenomena in shock-compressed liquids were observed and studied experimentally in [27]. They are principally of interest because they can be used as the basis to determine the type of conductivity of a particular substance in a dynamic experiment (electronic or ionic). The idea of the experiments involves recording the emf of a galvanic cell whose electrolyte is the shock-compressed conducting substance being studied. If the substance possesses predominantly ionic conductivity, electrochemical reactions may take place at the interface between the electrodes and the sample and consequently electric double layers may form. In this case, an electrode made of a chemically active metal such as magnesium, aluminum, zinc, and so on, will become oxidized and emit its positively charged ions into the conducting liquid, becoming negatively charged. The sum of the potential jumps at the electrode interfaces with the compressed material may be measured as the emf of the experimental galvanic cell. It should be noted that any electronic component of the conductivity will reduced this emf in proportion to its fraction [28].

In the present study the electrochemical emf in shock-compressed water was recorded using a system similar to that shown in Fig. 1 where the gauge *6* in contact with the sample was replaced by two electrodes, one made of copper and the other made of ~15 μm thick

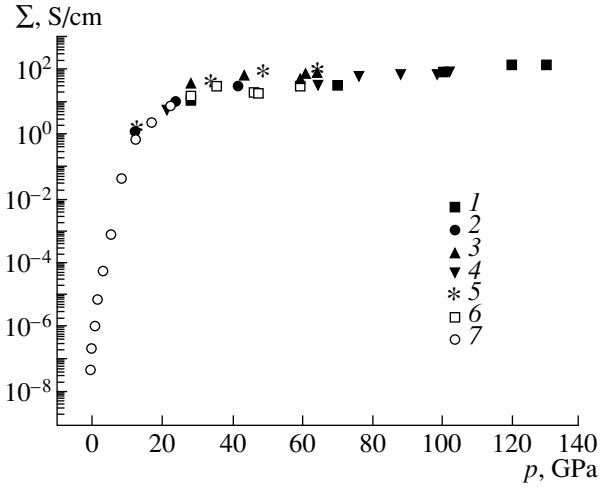

Fig. 6. Electrical conductivity of water as a function of dynamic pressure: (*1*–*5*) experimental points from the present study, (*6*) data from [8], (*7*) data from [16].

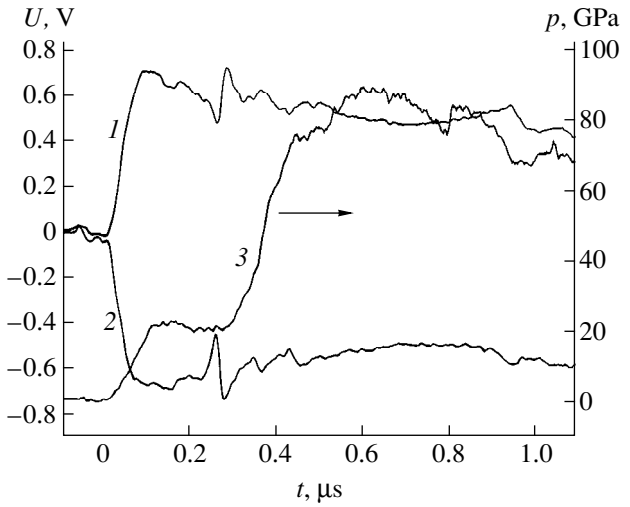

Fig. 7. Results of experiments to record the electrochemical potentials: (*1*) voltage at copper electrode, (*2*) voltage at aluminum electrode, (*3*) pressure profile in sample.

aluminum foil. No external sources of current were present in the recording channel. Each electrode was connected to the corresponding amplifier input of an oscilloscope by means of coaxial cable. In other words, a completely symmetric system with respect to ground was used to connect the electrodes to the measuring device in these experiments. In this case, the load resistance of the cell was the sum of the input impedances of the oscilloscope, 100 Ω, which is considerably higher than its internal resistance (tenths of an ohm). Figure 7 shows a typical experimental oscilloscope trace. Also shown is the dynamic pressure profile in the sample calculated using the readings from the manganin gauge. It can be seen that the potential difference between the electrodes in the experimental cell is approximately 1 V. As was predicted, the aluminum electrode becomes negatively charged relative to the less chemically active copper electrode. In accordance with [27], the amplitude and sign of the recorded emf indicate that this is of an electrochemical nature and therefore that the conductivity of the shock-compressed water is ionic in this range of dynamic pressures. In other words, highly compressed and heated water in the liquid state is essentially an ionic melt in which the majority carriers are protons whose mobility is appreciably higher than the heavier hydroxyl ions. The electrical conductivity of water is sufficient to explain the generation of the magnetic field of Uranus and Neptune by means of the dynamo effect [5].

3. CONCLUSIONS

The electrical conductivity of water has been measured in multistage (quasi-isentropic) water compression experiments in a region of the phase diagram not studied previously. The range of pressures and densities studied has been approximately doubled, being increased to 130 GPa and 3.2 g/cm^3, respectively. It has been shown that from approximately 30 GPa the electrical conductivity of water increases slowly with increasing density and reaches 150 S/cm at its maximum. Direct measurements of the electrochemical emf were used to establish that the conductivity of water is ionic under these extreme conditions.

ACKNOWLEDGMENTS

The authors thank W. Nellis for stimulating discussions.

This work was supported in part by the Russian Foundation for Basic Research (project no. 97-02-17575).

REFERENCES

1. W. B. Hubbard, Science **275**, 1279 (1997).
2. W. J. Nellis *et al.*, Science **240**, 779 (1988).
3. N. F. Ness *et al.*, Science **233**, 85 (1986).
4. N. F. Ness *et al.*, Science **246**, 1473 (1989).
5. S. Cavazzoni *et al.*, Science **283**, 44 (1999).
6. S. T. Weir, A. C. Mitchell, and W. J. Nellis, Phys. Rev. Lett. **76**, 1860 (1996).
7. A. A. Bakanova, V. N. Zubarev, Yu. N. Sutulov, and R. F. Trunin, Zh. Éksp. Teor. Fiz. **68**, 1099 (1975) [Sov. Phys. JETP **41**, 544 (1975)].
8. A. C. Mitchell and W. J. Nellis, J. Chem. Phys. **76**, 6273 (1982).
9. L. P. Volkov, N. P. Voloshin, and R. A. Mangasarov, Pis'ma Zh. Éksp. Teor. Fiz. **31**, 546 (1980) [JETP Lett. **31**, 513 (1980)].
10. M. A. Podurets, G. V. Simakov, R. F. Trunin, *et al.*, Zh. Éksp. Teor. Fiz. **62**, 710 (1972) [Sov. Phys. JETP **35**, 375 (1972)].
11. N. C. Holmes, W. J. Nellis, W. B. Graham, and G. E. Walrafen, Phys. Rev. Lett. **55**, 2433 (1985).
12. D. F. Calef and V. M. Pettitt, Chem. Phys. Lett. **139**, 129 (1987).
13. H. G. David and S. D. Hamann, Trans. Faraday Soc. **55**, 72 (1959).
14. H. G. David and S. D. Hamann, Trans. Faraday Soc. **56**, 1043 (1960).
15. A. A. Brish, M. S. Tarasov, and V. A. Tsukerman, Zh. Éksp. Teor. Fiz. **38**, 22 (1960) [Sov. Phys. JETP **11**, 15 (1960)].
16. S. D. Hamann and M. Linton, Trans. Faraday Soc. **62**, 2234 (1966).
17. S. D. Hamann and M. Linton, Trans. Faraday Soc. **65**, 2186 (1969).
18. Ya. B. Zel'dovich, S. B. Kormer, M. V. Sinitsyn, and K. B. Yushko, Dokl. Akad. Nauk SSSR **138**, 1333 (1961) [Sov. Phys. Dokl. **6**, 494 (1961)].

19. W. Holzapfel and E. U. Frank, Ber. Bunsenges. Phys. Chem. **70**, 1105 (1966).

20. V. V. Yakushev, Fiz. Goreniya Vzryva **14**, 3 (1978).

21. S. S. Nabatov, A. N. Dremin, V. I. Postnov, and V. V. Yakushev, Pis'ma Zh. Tekh. Fiz. **5**, 143 (1979) [Sov. Tech. Phys. Lett. **5**, 56 (1979)].

22. S. S. Nabatov, A. N. Dremin, V. I. Postnov, and V. V. Yakushev, Pis'ma Zh. Éksp. Teor. Fiz. **29**, 407 (1979) [JETP Lett. **29**, 369 (1979)].

23. G. E. Hauver, in *Proceedings of the Fifth International Symposium on Detonation, GPO, Washington, DC, 1970*, Ed. by S. J. Jacobs and R. Roberts, p. 387.

24. *LASL Shock Hugoniot Data*, Ed. by S. P. March (Univ. of California Press, Berkeley, 1980).

25. G. I. Kanel', G. G. Vakhitova, and A. N. Dremin, Fiz. Goreniya Vzryva **14**, 130 (1978).

26. V. I. Postnov, in *Nonstationary Problems in Hydrodynamics* (Inst. Gidrodinamiki, Novosibirsk, 1980), Vol. 48, p. 116.

27. V. V. Yakushev and A. N. Dremin, Zh. Fiz. Khim. **45**, 97 (1971).

28. C. Wagner, Z. Phys. Chem. B **21**, 42 (1933); **23**, 469 (1933).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 4, 2000, pp. 623–627.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 4, 2000, pp. 717–722.

FLUIDS

Coherent Spin Precession in a Separated Solution of ^{3}He in ^{4}He

A. I. Kuchaev and I. A. Fomin*
Kapitza Institute of Physical Problems, Russian Academy of Sciences, Moscow, 117334 Russia
**e-mail: fomin@kapitza.ras.ru*
Received November 9, 1999

Abstract—It is suggested that a coherently precessing spin structure in a normal Fermi liquid should be used to study the interface of two Fermi liquids. It is shown that the interface makes an additional contribution to the attenuation of the precessing structure. This contribution is determined by the kinetic coefficient which relates the magnetization flux across the interface to the magnetization jump at the interface. A relationship is established between this kinetic coefficient and the nature of the scattering of Fermi quasiparticles at the interface. Results of numerical calculations of the attenuation of the structure are presented for a specific object, a separated solution of ^{3}He in ^{4}He. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The interface between pure liquid ^{3}He and a saturated solution of ^{3}He in ^{4}He in a separated mixture is an ideal object for studying the properties of the interface between two different Fermi liquids. It is of considerable interest to study the spin fluxes across such an interface and the corresponding boundary conditions. In an earlier study [1] it was suggested that standing spin waves in a planar resonator filled with a separated mixture could be used for this purpose. The spin wave spectrum depends on the constraints imposed on the spin density and the spin current at the interface so that it is possible to assess the boundary conditions from data on the position and width of the natural resonator modes. However, the proposed experiment was not actually performed. In the present study we suggest using a previously studied nonlinear stable spin mode, a coherently precessing structure, as a test of the boundary conditions [2]. A coherently precessing structure has been observed separately in solutions of ^{3}He in ^{4}He and in pure liquid ^{3}He. It is formed in a weakly uniform magnetic field in the presence of large deviations of the magnetization from equilibrium and is described by the steady-state solution of the spin dynamics equations for a Fermi liquid (Leggett and Rice equations) [3]. For the case of a constant field gradient, the coherently precessing structure consists of two domains, in one of which the magnetization is parallel and in the other antiparallel to the field. The orientation of the magnetization changes over a length of domain wall perpendicular to the field gradient. The entire structure precesses at the same frequency ω_p which is equal to the Larmor frequency at the site of the domain wall, which allows the position of the wall to be monitored. A solution of the spin dynamics equations describing a coherently precessing structure in a separated mixture of ^{3}He and ^{4}He may be constructed from the solutions for each liquid by "matching" these at the interface using boundary conditions. We shall assume that the interface between the two liquids (a solution of ^{3}He in ^{4}He and pure ^{3}He) is planar and we shall direct the z axis perpendicular to the interface from the solution to the concentrated phase. We shall assume that the magnetic field $\mathbf{H}$ and the gradient of the Larmor frequency ω_L are directed along the z axis. In this particular geometry transport of the transverse projection of the magnetization relative to the field $M^{\pm} = M^x \pm iM^y$ is important. A thermodynamic analysis [4] (see also [1]) shows that the magnetization flux across the phase interface $\mathbf{J}_i$ is proportional to the jump of the effective field $\mathbf{B}_{eff} = \mathbf{M}/\chi$ at the interface (χ is the magnetic susceptibility). The boundary conditions for the transverse components $\mathbf{M}$ and $\mathbf{J}_i$ have the following form:

$$J_{12}^{+} = -J_{21}^{+} = b\left(\frac{M_1^{+}}{\chi_1} - \frac{M_2^{+}}{\chi_2}\right), \qquad (1)$$

where $M^{+} = M^x + iM^y$, the subscripts 1 and 2 refer to the different liquids (1 refers to the solution), and b is the kinetic coefficient which depends on the properties of the interface. It will be shown subsequently that the interface makes an additional contribution to the rate of relaxation of the coherently precessing structure whose magnitude for a given current is inversely proportional to b so that this structure can be used for an experimental determination of b. The value of b is in turn determined by the microscopic conditions for reflection of quasiparticles from the interface of two Fermi liquids and by knowing b, we can draw some conclusions on these conditions. A relationship between b and the probability of Fermi quasiparticles propagating across an interface was obtained in [1] for an interface having a low permeability for quasiparticles, although it was noted in this study that this simplifying assumption is

not satisfied for the interface between ^{3}He and a solution. In this context, it is convenient to reexamine the problem of microscopic boundary conditions.

2. BOUNDARY CONDITIONS

Spin waves and coherently precessing structures exist in the collisionless limit, i.e., when the characteristic frequency of the problem ω and the time between quasiparticle collisions τ_D satisfies the condition $\omega\tau_D \gg 1$. In this limit the state of the Fermi liquid is described by the density matrix

$$n_{\alpha\beta}(\mathbf{p}, \mathbf{r}, t) = n(\mathbf{p}, \mathbf{r}, t)\delta_{\alpha\beta} + \mathbf{s}(\mathbf{p}, \mathbf{r}, t) \cdot \boldsymbol{\sigma}_{\alpha\beta}, \quad (2)$$

where $\sigma_{\alpha\beta}$ are the Pauli matrices. We are only interested in the spin component $\mathbf{s}(\mathbf{p}, \mathbf{r}, t)$. In fact all the changes in the density matrix are concentrated near the Fermi surface and instead of $\mathbf{s}(\mathbf{p}, \mathbf{r}, t)$ it is convenient to use the vector

$$\mathbf{m}(\mathbf{n}, \mathbf{r}, t) = \frac{1}{2}\nu\int \mathbf{s}(\mathbf{p}, \mathbf{r}, t)d\epsilon_0$$

or the related vector

$$\boldsymbol{\rho}(\mathbf{n}, \mathbf{r}, t) = \mathbf{m}(\mathbf{n}, \mathbf{r}, t) + \int F^a(\mathbf{n}, \mathbf{n}')\mathbf{m}(\mathbf{n}', \mathbf{r}, t)\frac{d\mathbf{n}'}{4\pi},$$

where $F^a(\mathbf{n}, \mathbf{n}')$ is the spin component of the function describing the quasiparticle interaction in Landau theory, $\mathbf{n} = \mathbf{p}/|\mathbf{p}|$, ν is the density of states at the Fermi interface, and $\epsilon_0 = v_F(|\mathbf{p}| - p_F)$. For the following analysis it is convenient to express $\mathbf{m}$ in terms of $\boldsymbol{\rho}$:

$$\mathbf{m}(\mathbf{n}, \mathbf{r}, t) = \boldsymbol{\rho}(\mathbf{n}, \mathbf{r}, t) - \int G^a(\mathbf{n}, \mathbf{n}')\boldsymbol{\rho}(\mathbf{n}', \mathbf{r}, t)\frac{d\mathbf{n}'}{4\pi}.$$

The coefficients of expansion of the function $G^a(\mathbf{n}, \mathbf{n}')$ are expressed in terms of the corresponding coefficients for $F^a(\mathbf{n}, \mathbf{n}')$:

$$G_l^a = \frac{F_l^a}{1 + F_l^a/(2l+1)}.$$

The functions $\mathbf{m}(\mathbf{n}, \mathbf{r}, t)$ and $\boldsymbol{\rho}(\mathbf{n}', \mathbf{r}, t)$ are determined from the kinetic equation [5, 3] which we write in the collisionless limit:

$$\frac{\partial \mathbf{m}}{\partial t} + v_i\frac{\partial \boldsymbol{\rho}}{\partial x_i} - \mathbf{m} \times \boldsymbol{\omega}_L + \frac{4}{\hbar\nu}\mathbf{m} \times \boldsymbol{\rho} = 0. \quad (3)$$

The magnetization $\mathbf{M}$ and its flux $\mathbf{J}_i$ are expressed in terms of the zeroth and first harmonics $\mathbf{m}(\mathbf{n}, t)$ and $\boldsymbol{\rho}(\mathbf{n}, t)$, respectively:

$$\mathbf{M}(\mathbf{r}, t) = \gamma\hbar\int \mathbf{m}(\mathbf{n}, \mathbf{r}, t)\frac{d\mathbf{n}}{4\pi}, \quad (4)$$

$$\mathbf{J}_i(\mathbf{r}, t) = \gamma\hbar\int v_F n_i \boldsymbol{\rho}(\mathbf{n}, \mathbf{r}, t)\frac{d\mathbf{n}}{4\pi}. \quad (5)$$

The Leggett and Rice equations describe the long-wavelength motion of the magnetization at frequencies close to the Larmor frequency. They only contain $\mathbf{M}$ and $\mathbf{J}_i$ [3]. If the Fermi-liquid interaction is not weak, this description can be applied at distances from the interface substantially exceeding the characteristic length $l_\omega = v_f/\omega_L$. At distances from the wall shorter than l_ω all the harmonics $\mathbf{m}(\mathbf{n}, t)$ must be taken into account.

In the context of a coherently precessing structure we are interested in steady-state solutions of the kinetic equation describing the precession $\mathbf{m}(\mathbf{n}, \mathbf{r}, t)$ at frequency $\boldsymbol{\omega}_p$ which does not depend on the coordinates and is similar to the Larmor frequency:

$$\frac{\partial \mathbf{m}(\mathbf{n}, \mathbf{r}, t)}{\partial t} = \mathbf{m}(\mathbf{n}, \mathbf{r}, t) \times \boldsymbol{\omega}_p. \quad (6)$$

Assuming that the functions $\mathbf{m}$ and $\boldsymbol{\rho}$ only depend on a single coordinate z, we rewrite equation (5) in the form

$$v_z\frac{d\boldsymbol{\rho}}{dz} = \mathbf{m} \times (\boldsymbol{\omega}_L - \boldsymbol{\omega}_p) - \frac{4}{\hbar\nu}\mathbf{m} \times \boldsymbol{\rho}. \quad (7)$$

The equations for the two adjacent liquids have different values of the parameters. The boundary condition expresses the values of $\boldsymbol{\rho}$ for quasiparticles leaving the interface in terms of the values of $\boldsymbol{\rho}$ for quasiparticles approaching the interface. Because of the smallness of the perturbations, this relationship is linear and generally integral. However, if the analysis is confined to single-particle processes, i.e., those for which a quasiparticle incident at the interface is either reflected or converted into a similar quasiparticle in the other liquid, the relationship becomes algebraic. Nonsingle-particle processes have a lower probability because of the statistical weight usually found in Fermi liquids. Let us assume that θ is the angle between the direction of the quasiparticle momentum and the z axis, $\mu = \cos\theta$ (see Fig. 1). Conservation of the transverse projection of the momentum relative to the interface establishes a relationship between the values of μ in the two liquids:

$$p_{F1}\sqrt{1-\mu_1^2} = p_{F2}\sqrt{1-\mu_2^2}. \quad (8)$$

Incoming quasiparticles in liquid 1 (solution) have values of $\mu > 0$ and those in liquid 2 have values of $\mu < 0$. If we denote by $\alpha(\mu)$ the probability of a quasiparticle propagating across the interface, the boundary conditions for $\boldsymbol{\rho}_1(\mu_1)$ and $\boldsymbol{\rho}_2(\mu_2)$ may be written in the form:

$$\boldsymbol{\rho}_2(\mu_2) - \boldsymbol{\rho}_2(-\mu_2) = \frac{\nu_2}{\nu_1}\alpha(\mu_1)\boldsymbol{\rho}_1(\mu_1) - \alpha(\mu_1)\boldsymbol{\rho}_2(-\mu_2), \quad (9)$$

$$\boldsymbol{\rho}_1(\mu_1) - \boldsymbol{\rho}_1(-\mu_1) = \alpha(\mu_1)\boldsymbol{\rho}_1(\mu_1) - \frac{\nu_1}{\nu_2}\alpha(\mu_1)\boldsymbol{\rho}_2(-\mu_2). \quad (10)$$

Here μ_1 varies in the range $0 < \mu_1 < 1$ and μ_2 in accordance with (8) varies in the range $\mu^* < \mu_2 < 1$, where $\mu^* = \sqrt{1-(p_{F1}/p_{F2})^2}$. The right-hand sides of conditions (9) and (10) only contain the contributions of quasiparticles approaching the wall. In order to link the kinetic coefficient b in formula (1) with the function $\alpha(\mu)$, we consider equation (7) at distances from the interface of the order of l_ω. In this region we can neglect the first term on the right-hand side:

$$v_z \frac{d\boldsymbol{\rho}}{dz} = \frac{4}{\hbar\nu}\boldsymbol{\rho}\times\mathbf{m}. \tag{11}$$

Averaging this equation over the directions $\mathbf{n}$, we confirm that $\langle n_z\boldsymbol{\rho}\rangle \equiv \boldsymbol{\rho}_1 = \text{const}$. This equality is a consequence of spin conservation. We now rewrite the function $\boldsymbol{\rho}$ in the form

$$\boldsymbol{\rho} = \boldsymbol{\rho}_0 + 3n_z\boldsymbol{\rho}_1 + \boldsymbol{\psi}, \tag{12}$$

where the vector $\boldsymbol{\rho}_0$ is proportional to the average magnetization $\mathbf{M}$ which obeys the Leggett and Rice equations:

$$\boldsymbol{M} = \gamma\hbar\boldsymbol{\rho}_0/(1+F_0^a).$$

The variation of $\boldsymbol{\rho}_0$ over distances $\sim l_\omega$ can be neglected. The vector $\boldsymbol{\rho}_1$ is proportional to the current: $\mathbf{J}_z = \gamma\hbar v_F\boldsymbol{\rho}_1$, where the current $\mathbf{J}_z$ does not depend on z and can be calculated at any point, in particular at $z = 0$. We multiply equation (10) by $\gamma\hbar v_F\mu$ and integrate it over μ between zero and one. The left-hand side of the integrated equality is equal to $2\mathbf{J}_z/v_F$. On the right-hand side we substitute $\boldsymbol{\rho}(\mu)$ in the form (12) which gives

$$\begin{aligned}\frac{2\mathbf{J}}{v_f} &= I_1(1+F_{01}^a)\mathbf{M}_1 - \frac{\nu_1}{\nu_2}(1+F_{02}^a)\mathbf{M}_2 + I_2\mathbf{J}\\ &+ \gamma\hbar\int\mu_1\alpha(\mu_1)\boldsymbol{\psi}(\mu_1)d\mu_1\\ &- \frac{\nu_1}{\nu_2}\int\mu_1\alpha(\mu_1)\boldsymbol{\psi}(-\mu_2)d\mu_1,\end{aligned} \tag{13}$$

where

$$I_1\{\alpha(\mu)\} = \int_0^1\mu\alpha(\mu)d\mu,$$

$$I_2\{\alpha(\mu)\} = \int_0^1\mu\alpha(\mu)\times\left[\mu+\left(\frac{p_{F1}}{p_{F2}}\right)^2\sqrt{1-\left(\frac{p_{F1}}{p_{F2}}\right)^2(1-\mu)^2}d\mu\right].$$

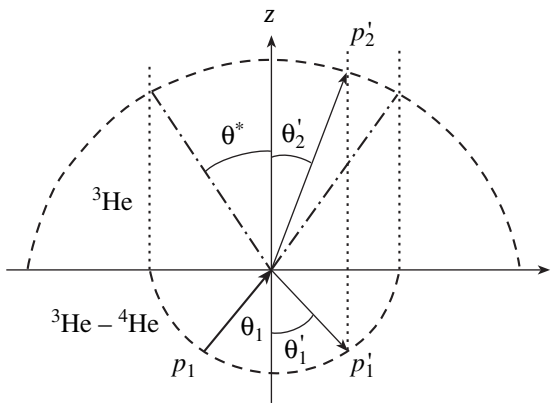

Fig. 1. Geometry of quasiparticle scattering at the interphase boundary between ^{3}He and a saturated solution of ^{3}He in ^{4}He. The dot-dash line indicates the cone of critical angles within which the momenta of quasiparticles entering the concentrated ^{3}He phase can lie.

The corrections $\boldsymbol{\psi}(\mu)$ in the distribution functions of the incoming particles can only occur as a result of the action of Fermi-liquid fields generated by quasiparticles reflected from the wall. In the absence of Fermi-liquid interaction the corrections $\boldsymbol{\psi}_1$ and $\boldsymbol{\psi}_2$ on the right-hand side of equality (13) vanish and its $(x + iy)$ projection reduces to the boundary condition (1) where

$$b = v_{F1}\nu_1\left(\frac{\hbar\gamma}{2}\right)^2\frac{I_1\{\alpha(\mu)\}}{2-3I_2\{\alpha(\mu)\}}. \tag{14}$$

The solution of the problem of particle propagation across the interface between two Fermi liquids with complete allowance for Fermi-liquid interaction is unknown to us. Note that in cases where the transparency of the interface is low ($\alpha(\mu) \ll 1$) expression (14) yields the corresponding expression from [1] whereas for cases of complete transparency ($\alpha(\mu) = 1$) we find $p_{F1} = p_{F2}$, $1/b = 0$, i.e., no $\mathbf{B}_{\text{eff}}$ jump occurs at the interface, as predicted. Using for $\alpha(\mu)$ a simple model of reflection from a rectangular potential barrier, as in [1],

$$\alpha(\mu) = \frac{4p_{1\perp}p_{2\perp}}{(p_{1\perp}+p_{2\perp})^2}, \tag{15}$$

for values of the parameters corresponding to a separated solution at zero pressure we have $b = v_F\chi(1+F_0^a)I$, where all the parameters are taken for liquid 1 (the solution) and $I = 0.320$. In order to estimate the correction to I obtained when allowance is made for Fermi-liquid interaction, we consider the equation for $\boldsymbol{\psi}(\mu)$ obtained after substituting (12) into (7):

$$v_z\frac{d\boldsymbol{\psi}}{dz} = \frac{4}{\hbar\nu}F_0^a\mathbf{m}_0\times(\boldsymbol{\psi}_0-\boldsymbol{\psi}), \tag{16}$$

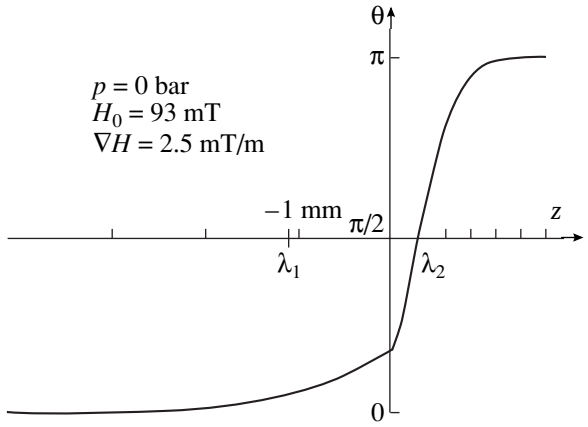

Fig. 2. Dependence of the angle θ for a coherently precessing structure near the interface between ^{3}He and a saturated solution of ^{3}He in ^{4}He; λ_1 and λ_2 are the characteristic lengths in the solution and in the concentrated ^{3}He phase, respectively.

where $\psi_0 = \langle\psi\rangle$ and the angular brackets denote averaging over angles. The solution of equation (16) for particles leaving the interface has the form

$$\psi(\mu, z) = \psi(\mu, 0)\exp\left(-\frac{iz}{\mu l_0}\right) + \frac{i}{\mu l_0}\exp\left(-\frac{iz}{\mu l_0}\right) \times \int_0^z \exp\left(\frac{iz'}{\mu l_0}\right)\psi(z')dz', \quad (17)$$

where $l_0 = 4F_0^a m_0/v_F\hbar\nu$, and for particles approaching the interface

$$\psi(\mu, z) = \frac{i}{\mu l_0}\exp\left(-\frac{iz}{\mu l_0}\right) \times \int_\infty^z \exp\left(\frac{iz'}{\mu l_0}\right)\psi(z')dz'. \quad (18)$$

Successive corrections to $\psi(\mu, z)$ may be obtained by iterations of equations (17) and (18) together with equation (13). The zeroth approximation for the incoming particles may be taken to be the solution with $\psi_0 = 0$ whereupon we return to expression (14) for *b*. For outgoing particles $\psi(\mu, 0)$ is obtained from the boundary conditions (9) and (10) after which $\psi(\mu, z)$ is obtained from equation (17) followed by ψ_0. Substituting ψ_0 into equation (18) gives the correction to $\psi(\mu, z)$ for the incoming particles which allows us to calculate the corrections to the integrals $I_1\{\alpha(\mu)\}$ and $I_2\{\alpha(\mu)\}$ in formula (14). The first iteration gives corrections to I_1 and I_2 of the order of 10%, in this case the correction to *b* only appears in the third decimal place, $I = 0.318$ is obtained instead of $I = 0.320$, which indicates that expression (14) is satisfactorily accurate for *b*. We shall subsequently use this expression in the numerical calculations.

3. ATTENUATION OF A COHERENTLY PRECESSING STRUCTURE

It was shown in [6] that the change in the angle θ between the magnetization and the *z* axis for a coherently precessing structure in a homogeneous liquid is described by a universal function of the argument $(z - z_0)/\lambda$, where

$$\lambda = \left[\frac{w^2\chi}{3\kappa\gamma M\nabla\omega_L}\right]^{1/3}$$

is the scale characterizing the thickness of the domain wall, z_0 is the coordinate of the plane perpendicular to the *z* axis on which the local Larmor frequency $\omega_L(z)$ is equal to the precession frequency ω_p of the coherently precessing structure, $w^2 = v_F^2(1 + F_0^a)(1 + F_1^a/3)$, and $\kappa = -(F_0^a - F_1^a/3)/(1 + F_0^a)$. The function $\theta(z - z_0/\lambda)$ is obtained as a solution of the equation

$$\frac{d^2\theta}{dz^2} = -\frac{z - z_0}{\lambda^3}\sin\theta, \quad (19)$$

which satisfies the boundary conditions $\theta \longrightarrow 0$ for $z \longrightarrow -\infty$ and $\theta \longrightarrow \pi$ for $z \longrightarrow \infty$. The current $\mathbf{J}_3$ is proportional to the derivative $d\theta/dz$:

$$\mathbf{J}_3^x = -\frac{w^2\chi}{3\kappa\gamma^2}\frac{d\theta}{dz}. \quad (20)$$

In order to determine the form of a coherently precessing structure which exists simultaneously in two liquids, we need to match the two solutions at the interface using the boundary conditions (1). As a result of the difference between the parameters of the two liquids, the continuity of the current leads to a jump of the derivative θ. A jump of the transverse component of **M** does not change the angle θ which thus remains continuous at the interface. Figure 2 shows an example of the matching of the angles θ for the interface between ^{3}He and a solution under typical conditions. The jump $\Delta M^-/\chi$ causes a slight distortion of the structure near the interface which is smoothed at distances of the order of λ. As a result of the smallness of the distortion, this has no significant influence on the attenuation of the coherently precessing structure of interest to us.

In order to calculate the contribution of the separation boundary to the attenuation of a coherently precessing structure, we use an equation obtained earlier (see [6], formula (4)) to describe the change in the energy of this structure. In dimensional units this has the form

$$\frac{\partial E}{\partial t} + \frac{\partial}{\partial z}\left((\mathbf{M} - \chi\mathbf{H})\cdot\frac{\mathbf{J}_3}{\chi}\right) = -\frac{3\mathbf{J}_3^2}{(1 + F_0^a)v_F^2\chi\tau}, \quad (21)$$

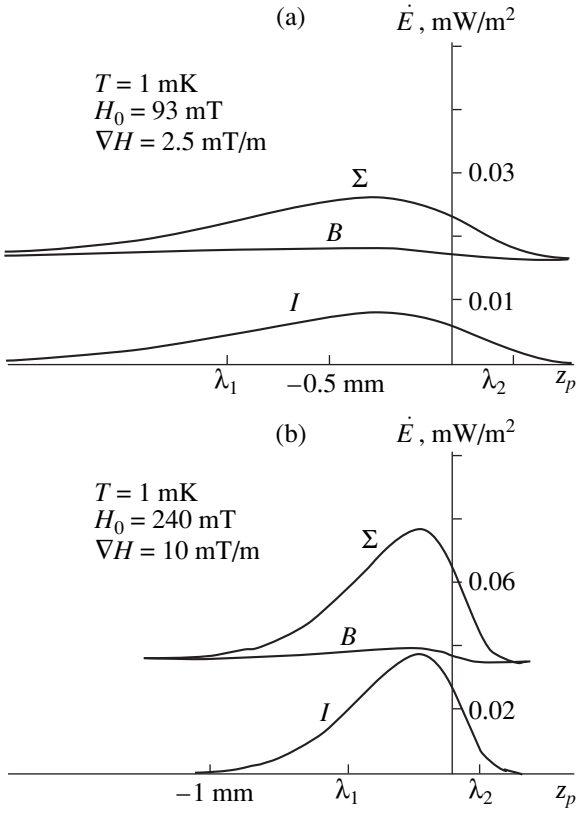

Fig. 3. Dependence of the volume (*B*) and surface (*I*) contributions to the attenuation of a coherently precessing structure and also of their sum (Σ) on the arbitrary distance between the precessing domain wall and the interface, which is defined as $z_p = (\omega_p - \omega_0)/\nabla\omega_L$, where ω_0 is the Larmor frequency at the interface, λ_1 and λ_2 are the characteristic lengths in the solution and in the concentrated ^{3}He phase, respectively.

where E is the energy density,

$$E = \frac{(\mathbf{M} - \chi\mathbf{H})^2}{2\chi} + \frac{3\mathbf{J}_3^2}{(1 + F_0^a)(1 + F_1^a/3)v_F^2\chi}.$$

Integrating equation (16) over z, we find the energy dissipation per unit area of the interface:

$$\frac{\partial}{\partial t}\left(\int E dz\right) = -\frac{3}{(1 + F_0^a)v_F^2\chi\tau}\int \mathbf{J}_3^2 dz - \frac{\mathbf{J}_3^2}{b}\bigg|_{z=0}. \quad (22)$$

The ratio of the interface contribution to the attenuation [the last term in formula (18)] to the volume contribution is of the order $((1 + F_0^a)/3)(l_\tau/\lambda)$, where $l_\tau \sim v_F\tau$ is the mean free path of the quasiparticles.

Figure 3 gives results of calculating the dependence of the attenuation of a coherently precessing structure on the distance between the precessing domain wall and the interface between ^{3}He and the solution for two values of the field and the gradient. For the calculations the current distribution determined neglecting attenuation was substituted on the right-hand side of formula (22) and expression (15) was used for the quasiparticle propagation probability $\alpha(\mu)$. In both cases considered, the interface makes a significant contribution to the attenuation of the coherently precessing structure and this contribution increases as the field and its gradient increase. The results for H_0 = 93 mT and ∇H = 2.5 mT/m differ from that given in our previous study [7] where we made a calculation error. This error influenced the numerical values of the coefficients in the formula for the reflection coefficient of the spin waves for small wave vectors and the correct expression for the reflection coefficient in this limit has the form $R \approx 0.0145–13.12(kv_F/\omega_L)$.

4. CONCLUSIONS

These results show that coherently precessing structures may be used as a tool to study the interface between ^{3}He and a solution. The most convenient experimental setup is that corresponding to continuous NMR [8] when the energy dissipation of the coherently precessing structure is compensated by the resonant rf field. The field frequency determines the position of the domain wall which can thus be monitored. This measurement scheme has many advantages over that proposed in [1] which is based on the observation of spin-wave modes in a resonator.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project no. 98-02-167444) and INTAS (grant no. 96-0610).

REFERENCES

1. A. Heff, D. Candela, D. O. Edwards, and S. Kumar, Europhys. Lett. **4**, 1043 (1987).
2. V. V. Dmitriev and I. A. Fomin, Pis'ma Zh. Éksp. Teor. Fiz. **59**, 352 (1994) [JETP Lett. **59**, 378 (1994)].
3. A. J. Leggett, J. Phys. C **3**, 448 (1970).
4. L. D. Fleshner, D. R. Fredkin, and S. Schultz, Solid State Commun. **18**, 207 (1976).
5. V. P. Silin, Zh. Éksp. Teor. Fiz. **33**, 1227 (1957).
6. I. A. Fomin, Physica B **210**, 373 (1995).
7. I. A. Fomin and A. I. Kouchayev, submitted to Physica B (1999).
8. V. V. Dmitriev, Czech. J. Phys. **46**, Suppl. S6, 3011 (1996).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 4, 2000, pp. 628–638.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 4, 2000, pp. 723–734.

SOLIDS
Structure

Differential Elastic Moduli of an Anisotropic Medium under Pressure

A. P. Kochkin

Vereshchagin Institute of High-Pressure Physics, Russian Academy of Sciences, Troitsk, Moscow oblast, 142430 Russia
e-mail: kochkin@hppi.troitsk.ru

Received June 30, 1998; in final form, September 26, 1999

Abstract—Natural strain theory is used to determine the tensor of the differential elastic moduli for strains of arbitrary magnitude and its relationship is established with values measured in acoustic and mechanostatic experiments in which an additional small load is applied. For illustration purposes an expression for this tensor is given for an object which is isotropic before application of the load, and formulas to determine this tensor from the results of measurements are calculated for an anisotropic object under pressure to within the second order in terms of the shear produced as a result of its hydrostatic compression. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

A method of analyzing the finite elastic strains of an isotropic medium using the natural strain tensor was developed in [1], its principal values having being introduced in [2] in calculations of the virtual deformation work when the increment of the strain tensor is coaxial with this strain. The problem of the uniaxial expansion of an isotropic object subjected to preliminary compression of arbitrary magnitude in a high-pressure chamber was discussed in [3] and the concepts of Young's modulus and the Poisson ratio under pressure were correctly introduced in terms of nonlinear natural strain theory. This was possible because of the simplicity of the equation of state when the stress tensor is coaxial with the strain tensor so that all the behavior characteristics of the material are determined only by the dependences of the scalar coefficients on the invariants of the strain tensor with direction tensors (the first three degrees of the strain tensor).

In analyses of additional small deformations of a preliminarily loaded object, the results can be meaningfully expressed in terms of the differential elastic moduli, while for an anisotropic object this is frequently the most reasonable method because the equation of state for low-symmetry crystals may be fairly cumbersome.

In [4] natural strain theory was extended to anisotropic media. The mathematical tools described in this study can be applied to calculate the differential elastic moduli of an arbitrary homogeneous medium, which is the aim of the present study. The notation used here is exactly the same as that introduced in [4]; some of the points noted here are clarified in [4] and we shall not specifically discuss these in the text for reasons of space. References to [4], for example, formulas (12b), (C.12b), and Appendix C will be given as [4 (12b)], [4 (C.12b)], and [4, C], respectively.

We also add that as a result of the additivity and commutativity of the volume and shear strains in natural strain theory, expansions in powers of the shear, i.e., the distance from the hydrostatic curve (and not the origin as in all previously known formulations of nonlinear strain theory), could be constructed for any value of these strains.

2. FUNDAMENTAL QUANTITIES IN STRAIN THEORY FOR AN ANISOTROPIC OBJECT

Let us assume that deformation transfers an arbitrary point ξ_i of an unstressed object to the point x_i as a result of a displacement by the vector u_i: $x_i = \xi_i + u_i$. The initial quantities describing the deformation in [1, 3] are the matrices μ and its inverse λ, where their elements are determined as follows (β is the distortion tensor):

$$\mu_{ij} = \frac{\partial \xi_i}{\partial x_j}, \quad \lambda_{ij} = \frac{\partial x_i}{\partial \xi_j},$$
$$\mu = E - \beta^{\mathrm{T}}, \quad \beta_{ij} = \frac{\partial u_j}{\partial x_i}. \tag{1}$$

We then introduce the natural strain tensor

$$s = -\frac{1}{2}\ln\gamma, \quad \gamma = \mu\mu^{\mathrm{T}}. \tag{2}$$

Thus, the quantity γ introduced differs from $\gamma = \mu^{\mathrm{T}}\mu$ used in [1, 3] but, as explained in [4], they are both

orthogonally similar: if γ from (2) is denoted as γ_0 and similar notation is introduced for s, we obtain

$$\gamma = O\gamma_0 O^{\mathrm{T}}, \quad s = Os_0 O^{\mathrm{T}}, \tag{2a}$$

where

$$\mu = O^{\mathrm{T}} S = S_0 O^{\mathrm{T}}, \quad \gamma = S^2, \quad \gamma_0 = S_0^2, \tag{2b}$$

where S, S_0 are symmetric matrices and O is an orthogonal matrix [5].

Since the thermodynamic variables for the anisotropic case are s_0, γ_0, and S_0, the values s, γ, and S cannot appear in this study. Thus, for conciseness the subscript "0" is subsequently [as in (2)] omitted from the strain variables (apart from Section 9 where the difference between the values with and without the subscript "0" needs to be recalled).

3. MEASUREMENT OF STRAIN USING THE DIFFRACTION PATTERN OF ELASTIC SCATTERING FOR A SINGLE CRYSTAL

The vectors of the affine (generally speaking) basis of the Bravais crystal lattice are denoted by $\mathbf{a}_i$ following conventional notation and the contravariant vectors are denoted by $\mathbf{b}^i$ (vectors of the reciprocal lattice).

We introduce the matrix ν of the transition from the Cartesian basis $\{\mathbf{i}_k\}$ to the affine basis $\{\mathbf{a}_i\}$:

$$\mathbf{a}_i = \nu_i^k \mathbf{i}_k, \quad \nu_i^k = \mathbf{i}_k \cdot \mathbf{a}_i, \tag{3}$$

[$\mathbf{a} \cdot \mathbf{b}$ is the scalar product in $E(3)$] and the inverse matrices ζ which are used to determine the contravariant (and biorthogonal to $\{\mathbf{a}_i\}$) basis $\{\mathbf{b}^i\}$ using the formula

$$\mathbf{b}^i = \zeta_k^i \mathbf{i}_k, \quad \zeta_k^i = \mathbf{b}^i \cdot \mathbf{i}_k.$$

The metric tensor in the affine basis is

$$g_{ij} = \mathbf{a}_i \cdot \mathbf{a}_j = \nu_i^k \nu_j^k = (\nu^{\mathrm{T}}\nu)_{ij}, \tag{4}$$

and its inverse (contravariant) is

$$g^{ij} = \mathbf{b}^i \cdot \mathbf{b}^j = \zeta_k^i \zeta_k^j = (\zeta\zeta^{\mathrm{T}})_{ij} \tag{4a}$$

with the usual rules for raising and lowering the indices so that if (but only in this section) the coordinates of the tensors in the Cartesian basis are denoted by the letters ($\hat{c}_i$, $\hat{\mu}_{ij}$):

$$\begin{aligned} \mathbf{c} &= \hat{c}_k \mathbf{i}_k = c_i \mathbf{b}^i = c^i \mathbf{a}_i, \\ \mu &= \hat{\mu}_{kl}(\mathbf{i}_k \otimes \mathbf{i}_l) = \mu^{i\cdot}_{\cdot j}(\mathbf{a}_i \otimes \mathbf{b}^j) \\ &= \mu^{ij}(\mathbf{a}_i \otimes \mathbf{a}_j) = \ldots, \end{aligned} \tag{5}$$

a relationship between the components in the different bases follows directly from (3), (4), and (4a), for example

$$\hat{\mu}_{ij} = \nu_k^i \zeta_j^l \mu^{k\cdot}_{\cdot l}, \quad \mu^{i\cdot}_{\cdot j} = \zeta_k^i \nu_j^l \hat{\mu}_{kl} \tag{5a}$$

[in (5) the symbol $\otimes$ indicates the Kronecker product of the vectors from $E(3)$ which determines the basis in the space $E(3) \otimes E(3)$ of the second-rank tensors].

We shall assume that as a result of continuous (i.e., not accompanied by phase transition) deformation, the vectors $\mathbf{a}_i$ become

$$\mathbf{a}'_i = T_i^j \mathbf{a}_j. \tag{6}$$

Then after deformation the point

$$\boldsymbol{\xi} = \xi^i \mathbf{a}_i$$

is converted into the point

$$\mathbf{r} = \xi^i \mathbf{a}'_i = \xi^i T_i^j \mathbf{a}_j,$$

i.e.,

$$x^i = T_j^i \xi^j.$$

In [4] we showed that in curvilinear coordinates the components $\mu^{i\cdot}_{\cdot j}$ of the tensor μ are absolute derivatives of the coordinates ξ_i of the point before deformation with respect to its coordinates x_j after deformation and since in affine coordinates the Kronecker symbols are zero, these derivatives reduce to ordinary partial derivatives (see [6] or [7]):

$$\mu^{i\cdot}_{\cdot j} = \partial\xi^i/\partial x^j = (T^{-1})^i_j. \tag{7}$$

It is now obvious that since $\mathbf{b}'^i \cdot \mathbf{a}'_j = \mathbf{b}^i \cdot \mathbf{a}_j = \delta^i_j$,

$$\mathbf{b}'^i = \mu^{i\cdot}_{\cdot j} \mathbf{b}^j \tag{8}$$

or, introducing the no-less-traditional vectors $\mathbf{g}^i = 2\pi\mathbf{b}^i$,

$$\mathbf{g}'^i = \mu^{i\cdot}_{\cdot j} \mathbf{g}^j.$$

The vectors $\mathbf{g}^i$ are measured directly in a structural experiment using the Laue method (after determining their absolute values by the Bragg method or directly if a spectrum analyzer is available). For the selected reflex (with the vector $\mathbf{g}$ before deformation and $\mathbf{g}'$ after deformation) we can then write (m_i are integers, $\mathbf{k}$ is the wave vector of the incident wave)

$$\mathbf{g} = k\{\cos\varphi\sin\vartheta, \sin\varphi\sin\vartheta, \cos\vartheta - 1\} = m_i \mathbf{g}^i, \tag{9}$$

$$\begin{aligned} \mathbf{g}' &= k\{\cos\varphi'\sin\vartheta', \sin\varphi'\sin\vartheta', \cos\vartheta' - 1\} \\ &= m_i \mathbf{g}'^i = m'_i \mathbf{g}^i, \end{aligned} \tag{9a}$$

where

$$m'_i = m_j \mu^{j\cdot}_{\cdot i} = \frac{1}{2\pi}\mathbf{g}' \cdot \mathbf{a}_i. \tag{9b}$$

Thus, three independent reflexes ($\alpha = 1, 2, 3$)

$$\mathbf{g}_\alpha = m_{\alpha i}\mathbf{g}^i, \quad \mathbf{g}'_\alpha = m'_{\alpha i}\mathbf{g}^i \tag{10}$$

completely determine the matrix μ:

$$\mu^{i\cdot}_{\cdot j} = (m^{-1})^{i\alpha} m'_{\alpha j}. \tag{11}$$

Now, if the initial orientation of the crystal, i.e., the matrix ν (or ζ) is known, formulas (5) and (5a) can be used to calculate the components of the matrix μ in the Cartesian basis:

$$\hat{\mu}_{ij} = \nu^i_k \zeta^l_j \mu^{k\cdot}_{\cdot l},$$

after which [see (2b)] we find the matrix S, and then the rotation matrix O:

$$S = \sqrt{\hat{\mu}\hat{\mu}^{\mathrm{T}}}, \quad O = \hat{\mu}S^{-1} = S^{-1}\hat{\mu}. \tag{12}$$

The accuracy of determining the strain by this method depends on the size of the diffraction spots, i.e., for an ideal crystal it is determined by the properties of the lattice sum for a three-dimensional diffraction grating (although it can also be estimated from expressions for the total scattering cross section in the spot σ and the maximum differential cross section $((d\sigma/do)_{\max}$ in the Laue method, see [8]), i.e., since $\delta o \sim (a/L)^2$, it follows from (9), (9a), and (11) that

$$\Delta s \sim \Delta m \sim \Delta\varphi \gg \delta\varphi \sim \sqrt{\delta o} \sim a/L \sim 10^{-8}\text{–}10^{-7} \tag{13}$$

(here the symbol Δ refers to the measurable change of an observable quantity, a is the lattice constant, and L is the sample dimension).

This estimate is also valid in dynamic scattering theory [9] (obviously for asymptotic values of the scattering amplitude, i.e., for intensity measurements at distances from the sample $r \gg L$).

Broadening of the diffraction spot is caused [10] by the presence of defects whose length is comparable with its dimensions (for example, dislocation lines or randomly distributed overlap errors in the crystal) which leads to a loss of accuracy in the strain measurements. By increasing the diffusion background and lowering the intensity of the principal maximum, finite defects keep the line width constant although they may lead to distortion of the average (over the crystal) unit cell [10]. If in this last case, the strain does not cause any redistribution of defects which influences this distortion of the cell, this estimate of the accuracy still applies. The experimental errors mainly arising from the inaccurate determination of k appreciably exceed (13) so that this accuracy estimate is the maximum possible.

4. ISOTHERMAL DIFFERENTIAL TENSOR OF THE EFFECTIVE ELASTIC MODULI AND THERMODYNAMIC STRAIN SUSCEPTIBILITY OF A PRELOADED OBJECT

1. As in [4], the tensor of the stresses σ of an anisotropic medium satisfying the equilibrium conditions

$$\partial\sigma_{ij}/\partial x_j + f_i = 0, \tag{14}$$

is a function of the asymmetric tensor μ and the flux

$$\delta\sigma_{ij} = \frac{\partial\sigma_{ij}}{\partial\mu_{kl}}\delta\mu_{kl}. \tag{15}$$

However, it is meaningless to introduce the tensor of the differential elastic moduli as a derivative of σ with respect to μ because the deformation of the rotation, by increasing the number of its different components (for this derivative this number is 54), introduces no new information on the physical properties of the medium. In addition, the tensor thus determined depends on the choice of origin for the strain.

Assuming that the deformation consists of a preliminary displacement which converts ξ_i to $\bar{x}_i$ and a subsequent displacement by u'_i:

$$\bar{x}_i = \xi_i + \bar{u}_i, \quad x_i = \bar{x}_i + u'_i, \tag{16}$$

bearing in mind that $\bar{\mu}_{ij} = \partial\xi_i/\partial\bar{x}_j$ and having determined by analogy with (1)

$$\mu'_{ij} = \frac{\partial\bar{x}_i}{\partial x_j} = \delta_{ij} - \frac{\partial u'_i}{\partial x_j}, \tag{17}$$

we quite clearly obtain

$$\mu_{ij} = \bar{\mu}_{ik}\mu'_{kj}. \tag{17a}$$

It is then clear that

$$\mu_{qi}\frac{\partial}{\partial\mu_{qj}} = \mu'_{qi}\frac{\partial}{\partial\mu'_{qj}}, \tag{18}$$

and this implies that applying this operator to a quantity which does not depend on the choice of origin for the strain again gives a quantity which does not depend on this choice. In particular, if the point of instantaneous strain (which we shall denote by "c") is taken as the new origin, where $\mu' = E$, we obtain

$$\mu_{qi}\frac{\partial}{\partial\mu_{qi}} = \left.\frac{\partial}{\partial\mu'_{qj}}\right|_c.$$

Thus, instead of the components of the derivatives from (15) mentioned above, it is more logical to consider the quantities

$$N_{ijkl} = -\frac{\partial\sigma_{ij}}{\partial\mu_{qk}}\mu_{ql} \equiv -\left.\frac{\partial\sigma_{ij}}{\partial\mu'_{lk}}\right|_c, \tag{19}$$

in terms of which these components can easily be calculated:

$$\partial\sigma_{ij}/\partial\mu_{lk} = -N_{ijkm}\lambda_{ml}. \tag{20}$$

We shall call the tensor $\mathcal{N}$ thus introduced the differential tensor of the effective elastic moduli (isothermal). Using its components, the equilibrium equations (14) for arbitrary strains are written in the form

$$N_{ijkm}\lambda_{ml}\frac{\partial^2 u_l}{\partial x_j \partial x_k} + f_i = 0. \tag{21}$$

Exactly three of these equations are obtained for the components of the displacement vectors since it is unnecessary to use additional conditions that the tensor of the curvature of the Euclidean space is zero, which are a generalization of the St. Venant conditions (as in [1]).

2. We shall now calculate the tensor $\mathcal{N}$, expressing its components in terms of the components of the tensor $\mathcal{Y}$ introduced in [4] which we shall call the thermodynamic strain susceptibility:

$$Y_{ijkl} = (\partial\sigma_{0ij}/\partial s_{kl})_T. \tag{22}$$

Here σ_0 is the "thermodynamic" stress tensor obtained from the equation of state (p is the pressure, ρ is the density, and τ_0 is the shear stress):

$$\sigma_0 = \rho(\partial f/\partial s)_T = -pE + \tau_0, \tag{23}$$

f is the free energy per unit mass of the material, the natural strain tensor of an anisotropic medium s was determined in (2), and σ_0 and σ are related by

$$\begin{gathered}\sigma = \Sigma\sigma_0, \quad \Sigma = -2\mathcal{M}^{\mathrm{T}}\mathcal{D}, \\ \mathcal{M} = (\mu\otimes\mu), \quad \mathcal{D} \equiv \mathcal{D}^s_\gamma = \partial s/\partial\gamma\end{gathered} \tag{24}$$

[here $\otimes$ is the sign of the Kronecker product of the operators from $E(3)$].

Strain rotation is evidently eliminated by using these quantities but they are generally asymmetric with respect to interchange of the pairs $(ij) \rightleftharpoons (kl)$ so that $\mathcal{Y}$ generally has 36 different components. Bearing in mind that

$$\begin{gathered}\frac{\partial\gamma_{ij}}{\partial\mu_{kl}} = \frac{\partial(\mu_{im}\mu_{jm})}{\partial\mu_{kl}} \\ = \delta_{ik}\delta_{ml}\mu_{jm} + \mu_{im}\delta_{jk}\delta_{ml} = \delta_{ik}\mu_{jl} + \mu_{il}\delta_{jk},\end{gathered} \tag{25}$$

and denoting the second derivative of s with respect to γ by $\mathcal{D}^{(2)}$ (see Appendix in [4]), taking into account (24), [4 (A.3a)], and the symmetry of γ, we obtain

$$\begin{gathered}\frac{\partial\sigma_{ij}}{\partial\mu_{qk}} = -2\delta_{ik}(D\sigma_0)_{qn}\mu_{nj} - 2\mu_{mi}(D\sigma_0)_{mq}\delta_{jk} \\ -4\mu_{mi}\frac{\partial(D\sigma_0)_{mn}}{\partial\gamma_{uq}}\mu_{nj}\mu_{uk},\end{gathered} \tag{26}$$

where

$$\begin{gathered}\frac{\partial(D\sigma_0)_{mn}}{\partial\gamma_{uq}} \equiv \frac{\partial(D_{mnrs}\sigma^0_{rs})}{\partial\gamma_{uq}} \\ = D^{(2)}_{mnrsuq}\sigma^0_{rs} + D_{mnrs}Y_{rstv}D_{tvuq}.\end{gathered}$$

Again recalling (24) and [4 (A.3a)], we obtain

$$\begin{gathered}N_{ijkl} = -\delta_{ik}\sigma_{jl} - \delta_{jk}\sigma_{il} + 4\mu_{mi}\mu_{nj}D^{(2)}_{mnrsuq}\sigma^0_{rs}\mu_{uk}\mu_{ql} \\ + 4\mu_{mi}\mu_{nj}D_{mnrs}Y_{rstv}D_{tvuq}\mu_{uk}\mu_{ql}.\end{gathered} \tag{27}$$

Also taking into account [4 (A.7)], we write this expression in the form

$$\begin{gathered}\mathcal{N} = -2\Pi_+(E\otimes\sigma) \\ + 4\mathcal{M}^{\mathrm{T}}\{\mathcal{D}^{(2)}[\sigma_0] + \mathcal{D}\mathcal{Y}\mathcal{D}\}\mathcal{M}\end{gathered} \tag{27a}$$

(the last term can evidently be written in the form $\Sigma\mathcal{Y}\Sigma^{\mathrm{T}}$).

We now consider how $\mathcal{N}$ can be measured experimentally for static loading. We envisage a preloaded (in particular, hydrodynamically compressed in a pressure chamber) single-crystal sample to which an additional load is applied at the preloading point to measure the moduli. This additional load induces a change in the stress tensor $\delta\sigma_{ij}$ and we then study the displacement of the X-ray reflexes caused by the additional deformation. It follows from (15), (17a), and (20) that

$$\begin{gathered}\delta\sigma_{ij} = \frac{\partial\sigma_{ij}}{\partial\bar{\mu}_{nk}}(\mu_{nk} - \bar{\mu}_{nk}) \\ = \frac{\partial\sigma_{ij}}{\partial\bar{\mu}_{nk}}\bar{\mu}_{nl}(\mu'_{lk} - \delta_{lk}) = -N_{ijkl}\delta\mu'_{lk}.\end{gathered} \tag{28}$$

If quantities referring to the unloaded state are denoted by "0" and those referring to the preloaded state are denoted by a bar, from (10) and (9b) for an arbitrary set of three reflexes ($m^0_{\alpha i}$ are integers), we have

$$\mathbf{g}_{\alpha 0} = m^0_{\alpha i}\mathbf{g}^i_0, \quad \bar{\mathbf{g}}_\alpha = \bar{m}_{\alpha i}\mathbf{g}^i_0, \quad \mathbf{g}_\alpha = m_{\alpha i}\mathbf{g}^i_0, \tag{29}$$

where

$$\begin{gathered}\bar{m}_{\alpha i} = \frac{1}{2\pi}\bar{\mathbf{g}}_\alpha\cdot\mathbf{a}^0_i = m^0_{\alpha j}\bar{\mu}^{j\cdot}_{\cdot i}, \\ m_{\alpha i} = \frac{1}{2\pi}\mathbf{g}_\alpha\cdot\mathbf{a}^0_i = m^0_{\alpha j}\mu^{j\cdot}_{\cdot i}.\end{gathered} \tag{30}$$

Having determined the inverse matrix to $\bar{m}_{\alpha i}$ [see (11)],

$$(\overline{m^{-1}})^{i\alpha} = \bar{\lambda}^{i\cdot}_{\cdot j}(m^{-1}_0)^{j\alpha}, \tag{31}$$

we obtain

$$(\overline{m^{-1}})^{i\alpha}m_{\alpha j} = \bar{\lambda}^{i\cdot}_{\cdot k}\mu^{k\cdot}_{\cdot j} = \mu'^{i\cdot}_{\cdot j}.$$

Thus, we can obtain the tensor μ' from structural data and by measuring the tensor σ (for various orientations of the sample) we can determine $\mathcal{N}$ in the crystallographic coordinates.

For example, for the uniaxial deformation of an arbitrarily oriented, preloaded sample, assuming that ω_{ij} is the matrix of the operator for rotation of the sample from the laboratory to the crystallographic axes, and its real orientation written in the laboratory basis is:

$$e_i^{(cr)} = \omega e_i = \omega_{ji} e_j, \tag{32}$$

we have

$$N_{ijkl} = \omega_{ii'}\omega_{jj'}\omega_{kk'}\omega_{ll'}N^{(cr)}_{i'j'k'm'}. \tag{33}$$

Assuming that uniaxial expansion takes place along the z axis in the laboratory system:

$$\delta\sigma = tb^{33}, \quad (b^{pq})_{ij} = \delta_{ip}\delta_{jq}, \tag{34}$$

we obtain from (28)

$$t\omega_{3i}\omega_{3j} = N^{(cr)}_{ijkl}\delta\mu'^{(cr)}_{lk}, \tag{35}$$

where

$$\delta\mu'^{(cr)}_{lk} = \omega_{ml}\omega_{nk}\delta\mu'_{mn}. \tag{35a}$$

Since the number of components of the tensor $\mathcal{N}$ is too large to make excessive measurements, we must bear in mind that the aim is to determine the symmetric tensor $\mathcal{K}$ [see the following section, formulas (37), (37a), and (35b)] which is expressed in terms of $\mathcal{N}$ using (39) and (39a). For specific calculations using these formulas see Section 8 below.

5. COMPLETELY SYMMETRIC TENSOR OF THE DIFFERENTIAL ELASTIC MODULI $\mathcal{K}$ IN TERMS OF INDICES

1. We shall describe tensors possessing interchange symmetry both within the first and second pairs of indices and with respect to the interchange of these pairs as being completely symmetric. This symmetry is clearly obtained for the quantities (the T sign indicates at constant temperature)

$$H^T_{ijkl} = \rho\left(\frac{\partial^2 f}{\partial s_{ij}\partial s_{kl}}\right)_T = \rho\left(\frac{\partial(\sigma^0_{ij}/\rho)}{\partial s_{kl}}\right)_T = Y^T_{ijkl} + \sigma^0_{ij}\delta_{kl}. \tag{36}$$

A similar construction is obtained for the adiabatic tensors $\mathcal{Y}^\eta$ and $\mathcal{H}^\eta$ (η is the entropy) where f is replaced by the energy ε per unit mass. Only at this point is a difference observed in the calculated values, all the other transformations are geometric and the same for the adiabatic and isothermal moduli. Thus, in the following formulas (where there is no need) we shall simply talk about the moduli, dropping the "T" or "η" signs.

Choosing $\mathcal{H}$ as the elastic modulus would be unsuccessful because, as can be seen from (36), this contains a linear term with respect to stress which is not explicitly related to the state of the object. An isothermal symmetric tensor closest to the susceptibility may be introduced using the formula

$$K^T_{ijkl} = \frac{1}{2}(Y_{ijkl} + Y_{klij}) = H^T_{ijkl} - \frac{1}{2}(\sigma^0_{ij}\delta_{kl} + \sigma^0_{kl}\delta_{ij}), \tag{37}$$

or, more concisely

$$\mathcal{K} = \mathcal{H} - \frac{1}{2}(\sigma_0\rangle\langle E + E\rangle\langle\sigma_0), \tag{37a}$$

where $(a\rangle\langle b)_{ijkl} = a_{ij}b_{kl}$ and then (36) gives

$$\mathcal{Y} = \mathcal{H} - \sigma_0\rangle\langle E = \mathcal{K} + \frac{1}{2}(E\rangle\langle\tau_0 - \tau_0\rangle\langle E). \tag{38}$$

It can be seen that for a hydrostatically compressed object the moduli $\mathcal{Y}$ and $\mathcal{K}$ are the same so that in this case, the tensor $\mathcal{Y}$ is symmetric with respect to interchange of the first and second pairs of coordinates.

Substituting (38) into (27a) we obtain

$$\mathcal{N} = \mathcal{N}' + \mathcal{R}, \tag{39}$$

where

$$\mathcal{N}' = -2\Pi_+(E\otimes\sigma) + \frac{1}{2}(E\rangle\langle\tau - \tau\rangle\langle E) + 4\mathcal{M}^{\mathrm{T}}\mathcal{D}^{(2)}[\sigma_0]\mathcal{M}, \tag{39a}$$
$$\mathcal{R} = \Sigma\mathcal{K}\Sigma^{\mathrm{T}}$$

(the form of the term containing τ can be understood from (24) and [4 (36a)]). The components of the tensor $\mathcal{N}'$ are known as soon as the stress and strain tensors are known, the last term in $\mathcal{N}'$ is a calculable quantity as is Σ (see Section 8) and only $\mathcal{K}$ depends on the properties of the material.

Since unlike $\mathcal{N}$, the tensor $\mathcal{K}$ (and therefore $\mathcal{R}$) possesses the same symmetry with respect to interchange of the indices as the elasticity tensor in linear theory, it has a maximum of 21 different components. However, since the free energy f is invariant with respect to intrinsic rotations of the symmetry group of an undeformed crystal, the true number of different components of the tensor $\mathcal{K}$ in a crystal with any specific symmetry group is the same as that given in [11] for linear theory.

Equation (35) may be written in the form

$$R^{(cr)}_{ijkl}\delta\mu'^{(cr)}_{lk} = t\omega_{3i}\omega_{3j} - N'^{(cr)}_{ijkl}\delta\mu'^{(cr)}_{lk}, \tag{35b}$$

which can reduce the number of different crystal orientations in experiments for cases of the most general (triclinic) symmetry to six (a larger number of these could be used to enhance the measurement accuracy). From these experiments we find $\mathcal{R}^{(cr)}$ and the $\mathcal{K}^{(cr)}$ using formulas (39) and (39a).

2. For illustration purposes we calculate the tensor of the elastic moduli of an isotropic object using (37). In this case we have [1, 3]

$$\sigma_0 = -pE + 2\mu\Delta + \nu\Delta_2, \tag{40}$$

whence

$$\begin{aligned}\mathcal{H} &= \frac{\partial\sigma_0}{\partial s} + \frac{1}{2}((2\mu\Delta + \nu\Delta_2))\rangle\langle E - E\rangle\langle(2\mu\Delta + \nu\Delta_2)) \\ &= K^{ab}\mathcal{P}_{ab} + 2\mu\Pi_+ + 2\nu\mathcal{K}_{12},\end{aligned} \tag{41}$$

where the symmetric matrix K^{ab} has the elements

$$\begin{aligned}K^{11} &= K - \frac{2\mu}{3}, \quad K^{12} = \mu + 2\frac{\partial\mu}{\partial I_{1s}} - \frac{2\nu}{3}, \\ K^{13} &= \frac{\nu}{2} + \frac{\partial\nu}{2I_{1s}}, \\ K^{22} &= 2\frac{\partial\mu}{\partial k_2}, \quad K^{23} = 2\frac{\partial\nu}{\partial k_2}, \quad K^{33} = \frac{\partial\nu}{\partial k_3},\end{aligned} \tag{41a}$$

where

$$K = -\frac{\partial p}{\partial I_{1s}} = -V\frac{\partial p}{\partial V}, \tag{41b}$$

$I_{1s} = \mathrm{Sp}s = \ln(\rho_0/\rho)$, and ρ_0 is the density of the undeformed material [4 (10)].

The operator Π_+ (see [4, A]) is the space unit of the second-rank symmetric tensors $T_+(2)$ (and the projector on it). Here the following operators appear

$$\mathcal{P}_{ab} = e_a\rangle\langle e_b, \quad \mathcal{K}_{ab} = (e_a \otimes e_b)_+, \tag{41c}$$

where $[(a \otimes b)_+]_{ijkl} = (a_{ik}b_{jl} + b_{il}a_{jk} + b_{ik}a_{jl} + a_{il}b_{jk})/4$ and e_a are the vectors of the basis $\mathbf{B}_2(s) = \{E, \Delta, \Delta_2\}$ [4 (B.7a)].

On the hydrostatic axis (where $\Delta = \Delta_2 = 0$) we obtain

$$\mathcal{H} = K^{11}\mathcal{P}_{11} + 2\mu\Pi_+ = \left(K - \frac{2\mu}{3}\right)E\rangle\langle E + 2\mu\Pi_+, \tag{41d}$$

or in index notation (see [4 (A.3a), (A.7)])

$$K_{ijkl} = K\delta_{ij}\delta_{kl} + \mu\left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk} - \frac{2}{3}\delta_{ij}\delta_{kl}\right), \tag{41e}$$

from which it can be seen that this is exactly the same as the usual expression in [11] but with the bulk modulus K and the shear modulus μ determined under pressure as in [12] and [3].

6. ADIABATIC DIFFERENTIAL ELASTIC MODULI OF A PRELOADED OBJECT

In order to calculate the adiabatic moduli from the results of ultrasound experiments, we consider the propagation of sound in a uniformly loaded elastic medium. By measuring the strain in the acoustic wave field from the equilibrium value $\bar{x}_i$ for given σ_0 and T, and denoting the displacement in the acoustic wave field by u_i', we obtain using (16)–(17a):

$$\frac{\partial\mu_{kl}}{\partial x_j} = \bar{\mu}_{kp}\frac{\partial\mu'_{pl}}{\partial x_j} = -\bar{\mu}_{kp}\frac{\partial^2 u'_p}{\partial x_l\partial x_j}. \tag{42}$$

We thereby assume that a small amount of heat is released as a result of the dissipation of acoustic energy. For small u_i' we can write $dv_i/dt \approx \partial^2 u'/\partial t^2$ so that the sound propagation equation has the form

$$\bar{\rho}\frac{\partial^2 u_i'}{\partial t^2} = \frac{\partial\sigma_{ij}}{\partial\mu_{qk}}\frac{\partial\mu_{qk}}{\partial x_j} = -\frac{\partial\sigma_{ij}}{\partial\mu_{qk}}\bar{\mu}_{ql}\frac{\partial^2 u_i'}{\partial x_j\partial x_k}. \tag{43}$$

Thus, this equation contains the tensor $\mathcal{N}$ already analyzed by us (but adiabatic) which plays the same role here as the linear-theory tensor λ_{ijkl} in formula (23.1) from [11]:

$$\bar{\rho}\frac{\partial^2 u_i'}{\partial t^2} = \bar{N}_{ijkl}\frac{\partial^2 u_l'}{\partial x_j\partial x_k}. \tag{43a}$$

After substituting $\mathbf{u}(\mathbf{r}) = \mathbf{u}_0\exp i(\mathbf{kr} - \omega t)$ and dividing the resulting equation by k^2, the equation to determine the natural modes and velocities of sound has the form

$$\rho c^2 u_i = N_{ij}(\mathbf{n})u_j, \quad \mathbf{n} = \mathbf{k}/k, \tag{44}$$

where the operator $N(\mathbf{n})$ acting in $E(3)$ with the matrix elements

$$N_{ij}(\mathbf{n}) = N_{iklj}n_k n_l \tag{44a}$$

(and similarly for the operators $T(\mathbf{n})$ and $R(\mathbf{n})$ is (see (39), (39a), and [4 (A.7)])

$$\begin{aligned}N_{ij}(\mathbf{n}) &= p\delta_{ij} + pn_i n_j \\ &- \frac{1}{2}(n_i\tau_{jk}n_k + \tau_{ik}n_k n_j) - \tau_{ij} + R'_{ij}(\sigma_0, \mathbf{n}),\end{aligned} \tag{44b}$$

where $\mathcal{R}' = \mathcal{R} + \mathcal{T}$

$$\mathcal{T}[\sigma_0] = 4\mathcal{M}^{\mathrm{T}}\mathcal{D}^{(2)}[\sigma_0]\mathcal{M} \tag{44c}$$

and thus

$$\mathcal{R}'_{ij}(\sigma_0, \mathbf{n}) = T_{ij}(\sigma_0, \mathbf{n}) + R_{ij}(\mathbf{n}). \tag{44d}$$

Since the number of different components of the operator $\mathcal{R}'$ acting in $E(3) \otimes E(3)$is the same as in linear theory, it becomes clear that determining the value of $\mathcal{R}'$ experimentally, and then $\mathcal{R}$ (and thus $\mathcal{H}$) with the matrix elements R_{iklj} for a known stress tensor σ is not much more difficult (neglecting the specific characteristics of experiments using stressed objects) than obtaining the ordinary tensor of the elastic moduli λ from [11]. In order to determine the moduli in the general case of triclinic crystal system, it is clearly necessary to measure the propagation velocities of all three acoustic modes for no less than seven differently oriented samples.

7. LOCAL DIFFERENTIAL ELASTIC MODULI

It was noted in Section 4 that some quantities (including $\mathcal{K}$) depend on the choice of strain origin (see also [13]) which is caused by the non-interchangeability of the elements $T(2) = E(3) \otimes E(3)$. Thus, as in [4], we select two variants of this choice: natural, or absolute, when the point at which $\sigma = 0$ is taken as the origin, and tangential when the origin is the point of instantaneous deformation $\bar{s}$, where the calculated values are denoted by the index "c" in this case. Quite clearly, the choice of strain origin has no influence on the observable quantities [$\mathcal{N}$ or, as follows from (39), $\mathcal{R}$).

As a result of fairly lengthy calculations using expressions for the matrix derivatives of the matrix-valued function $\ln\gamma$ [4, F] with the transformation from the basis $\mathbf{B}_2(\gamma)$ to the basis $\mathbf{B}_1(\gamma) = \{E, \gamma, \gamma^2\}$ [4, B] and going to the limit $\mu' \longrightarrow 0$ [see (17a)], we obtain from (39a)

$$\mathcal{N}' = 2\Pi_+(\tau \otimes E)\Pi_- + \frac{1}{2}(E\rangle\langle\tau - \tau\rangle\langle E), \qquad (45)$$

and since $\mathcal{R} = \mathcal{K}^c$, the expression for $\mathcal{N}$ in terms of $\mathcal{K}^c$ from (39) has the form

$$\mathcal{N} = 2\Pi_+(\tau \otimes E)\Pi_- + \frac{1}{2}(E\rangle\langle\tau - \tau\rangle\langle E) + \mathcal{K}^c, \quad (46)$$

or in terms of the components

$$N_{ijkl} = \frac{1}{2}(\tau_{ik}\delta_{jl} - \delta_{ik}\tau_{jl} + \delta_{il}\tau_{jk} - \tau_{il}\delta_{jk}) + \frac{1}{2}(\delta_{ij}\tau_{kl} - \tau_{ij}\delta_{kl}) + \mathcal{K}^c_{ijkl}.$$

The number of different components of the tensors $\mathcal{K}$ and $\mathcal{K}^c$ may differ: the first is related to the initial symmetry of the crystal (before deformation) and the second is related to the symmetry of the strain-distorted lattice.

It can be seen that even in a hydrostatically loaded anisotropic object, the values of $\mathcal{N}$ and $\mathcal{K}^c$ are particularly simply related:

$$\mathcal{N} = \mathcal{K}^c.$$

8. CALCULATION OF EXPRESSIONS FOR THE DIFFERENTIAL ELASTIC MODULI OF AN OBJECT UNDER PRESSURE

1. Since we are subsequently interested in measurements under hydrostatic pressure, we assume [4]

$$\sigma = \sigma_0 = -pE, \quad \mathcal{D}^{(2)}[-pE] = -p\mathcal{D}^{(2)}[E], \quad \mathcal{N}' = 2p\Pi_+ - p\mathcal{T}[E]. \qquad (47)$$

Then formula (44b), which is used to extract R_{ijkl} from the experimental data, becomes

$$N_{ij}(\mathbf{n}) = p\delta_{ij} + pn_in_j + R'_{ij}(\mathbf{n}), \qquad (47a)$$

where [see (44c), (44d)]

$$R'_{ij}(\mathbf{n}) = -pT_{ij}(E, \mathbf{n}) + R_{ij}(\mathbf{n}). \qquad (47b)$$

The tensor of the elastic moduli $\mathcal{K}$ is calculated from (39) and (39a) after determining the tensor $\mathcal{R}$ experimentally: in the static case (isothermal moduli) from (35b) and in the dynamic case (adiabatic moduli) from (47a). Thus, we must analyze all the quantities appearing in (39).

It can be seen from (24), (2b), and [4 (A.6a)] that it is possible to write

$$\mathcal{M} = (S \otimes S)\Omega^{\mathrm{T}}, \ \Sigma = \Omega\mathcal{B}, \qquad (48)$$

where

$$\Omega = (O \otimes O), \quad \mathcal{B} = -2(S \otimes S)\mathcal{D}. \qquad (48a)$$

It is then easy to see from (39a) that

$$\mathcal{K} = \Sigma^{-1}\mathcal{R}(\Sigma^{\mathrm{T}})^{-1} \equiv \mathcal{B}^{-1}\tilde{\mathcal{R}}\mathcal{B}^{-1} = \mathcal{B}^{-1}\tilde{\mathcal{N}}\mathcal{B}^{-1} - \Sigma^{-1}\mathcal{N}'(\Sigma^{\mathrm{T}})^{-1}, \qquad (49)$$

where

$$\tilde{\mathcal{R}} = \Omega^{\mathrm{T}}\mathcal{R}\Omega, \quad \tilde{\mathcal{N}} = \Omega^{\mathrm{T}}\mathcal{N}\Omega, \qquad (49a)$$

and

$$\Sigma^{-1}\mathcal{N}'(\Sigma^{\mathrm{T}})^{-1} = 2p(\mathcal{B}^{-1})^2 - p\mathcal{D}^{-1}\mathcal{D}^{(2)}[E]\mathcal{D}^{-1}. \qquad (49b)$$

As is clear from (48a), the operator $\mathcal{B}^{-1}$ may be written as

$$\mathcal{B}^{-1} = -\frac{1}{2}\mathcal{D}^{-1}(S^{-1} \otimes S^{-1}), \qquad (50)$$

where $\mathcal{D}^{-1}$ is the matrix derivative of γ with respect to s, the function $\gamma = \exp(-2s)$ is the particular case of the function

$$\exp(-ts) = E^a_t e_a = z^{t/2}\tilde{E}^a_t e_a, \quad z = \exp(-2I_{1s}/3) = (\rho/\rho_0)^{2/3} \qquad (51)$$

(see [4 (10)], and the correspondence between this notation and that in [3] [formulas (C.6)–(C.6b) is given by

$$\tilde{E}^1_t = a_t/3, \quad \tilde{E}^2_t = b_t, \quad \tilde{E}^3_t = c_t. \qquad (51a)$$

Calculations of $\mathcal{D}^{-1}$ in the basis $\mathbf{B}_2(s)$ [see (41b)] give (summing over a between 1 and 2):

$$\mathcal{D}^{-1} = -2(\gamma \otimes E)_+\mathcal{P} + aE_2^{a+1}\mathcal{L}_{1a}. \qquad (52)$$

Here we introduce (see (41c), [4 (A.2), (A.3)])

$$\mathcal{P} = \mathcal{P}_{ab}g^{ab} = e_a\rangle\langle e^a,$$
$$\mathcal{Q} = \mathcal{I} - \mathcal{P}, \quad \mathcal{L}_{ab} = \mathcal{K}_{ab}\mathcal{Q}, \tag{52a}$$

where $\mathcal{P}$ is the projection on the half-space L_s (formed by all powers of s), $g^{ab} = \langle e^a e^b\rangle$ is the contravariant metric tensor in L_s, $\mathcal{Q}$ is the projection on the orthogonal complementary minor $L_\perp$ to L_s as far as $T_+(2)$, $\mathcal{I}$ is the unit operator in $T(2)$ [4 (A.4)], the values of $\mathcal{K}_{ab}$ are determined in (41c), and all $\mathcal{L}$ are calculated, as [4 (A.3), (A.6), (B.15)–(B.17)]

$$\mathcal{L}_{ab} = \mathcal{K}_{ab} - \mathcal{Y}_{ab},$$
$$\mathcal{Y}_{ab} = \mathcal{K}_{ab}\mathcal{P} = q^c_{ab}q^{df}_c\mathcal{P}_{df}, \quad q^c_{ab} = \langle e^c e_a e_b\rangle. \tag{52b}$$

The expression for $\mathcal{D}^{(2)}[E]$ in the basis $\mathbf{B}_1(\gamma)$ [see (45)] is most easily obtained using formula [4 (C.12)]. Setting $f(\gamma) \equiv S(\gamma) = -(1/2)\ln\gamma$ and referring to (A.7) from [3], we conclude that since $S'(\gamma) = -\gamma^{-1}/2$,

$$(S')^2 = \frac{J^\gamma_1}{2I^\gamma_3}, \quad (S')^3 = -\frac{1}{2I^\gamma_3}, \tag{53}$$

and as is clear from (51) and [4 (B.7a)]

$$J^\gamma_p = 3E^1_{2p}. \tag{54}$$

Then, recalling the definition of the invariants in [1, 3] and their expression in terms of each other, we obtain

$$I^\gamma_3 = \frac{1}{3}J^\gamma_3 - \frac{1}{2}J^\gamma_1J^\gamma_2 + \frac{1}{6}(J^\gamma_1)^3$$
$$= E^1_6 - \frac{9}{2}E^1_2E^1_4 + \frac{9}{2}(E^1_2)^3. \tag{54a}$$

It is now easy to conclude that

$$\mathcal{D}^{(2)}[E] = \frac{1}{2}(\gamma^{-2}\otimes E)_+\mathcal{P} + C^a\mathcal{L}_{1a},$$
$$C^a = \delta^a_1(S')^2 + 2(S')^3E^a_2. \tag{55}$$

Multiplying (52) and (55) taking into account the commutativity of $\mathcal{K}_{ab}$ with $\mathcal{P}$ [4], the rules for multiplication of the operators $\mathcal{L}_{ab}$ given there [4 (34a)], and the identities

$$(\gamma^n\otimes\gamma^{-n})_+\mathcal{P} = E^a_{2n}E^b_{-2n}(e_a\otimes e_b)_+\mathcal{P}$$
$$= E^a_{2n}E^b_{-2n}e_ae_be_c\rangle\langle e^c = E^a_0q^b_{ac}e_b\rangle\langle e^c = e_c\rangle\langle e^c = \mathcal{P}, \tag{56}$$

obtained from the relationships

$$E^a_{t+t'} = E^b_tE^c_{t'}q^a_{bc}, \tag{56a}$$

derived by multiplying the corresponding exponential functions, having noted that the part of the unknown operator acting in L_s consists of terms having the form (56) with $n = 0, 1, 2$, we obtain after simple calculations:

$$\mathcal{D}^{-1}\mathcal{D}^{(2)}[E]\mathcal{D}^{-1} = 2\mathcal{P} + T^{ab}\mathcal{L}_{ab}, \tag{57}$$

where

$$T^{ab} = \frac{1}{4}cdE^{c+1}_2E^{d+1}_2$$
$$\times C^f(\delta^a_1q^g_{cd}q^b_{gf} + \delta^a_cq^b_{df} + \delta^a_dq^b_{cf} + \delta^a_fq^b_{cd}). \tag{57a}$$

The final result has the form

$$\mathcal{K} = \mathcal{B}^{-1}\tilde{\mathcal{N}}\mathcal{B}^{-1} + p[T^{ab} - 2U^{ab}]\mathcal{L}_{ab},$$
$$\tilde{\mathcal{N}} = \Omega^{\mathrm{T}}\mathcal{N}\Omega, \tag{58}$$

where, as can be established from (50), (52), and [4 (A.6)–(A.6b)],

$$\mathcal{B}^{-1} = \mathcal{P} - B^{ab}\mathcal{L}_{ab}, \quad B^{ab} = \frac{1}{2}cE^{c+1}_2E^a_{-1}E^d_{-1}q^b_{cd},$$
$$U^{ab} = B^{cd}_+B^{fg}_+q^a_{cf}q^b_{dg}, \quad B^{ab}_+ = \frac{1}{2}(B^{ab} + B^{ba}).$$

The rotation Ω as a result of the hydrostatic loading of an anisotropic sample is expressed in terms of the distortion using (12), where it depends on the choice of initial crystallographic plane fixed under loading, i.e., on the orientation and method of attaching the sample in the high-pressure chamber.

The given values of the moduli become functions of pressure as soon as the values of the shears in (58) $e_2 = \Delta(p)$ and $e_3 = \Delta_2(p)$ are expressed in terms of pressure using the equation of state (23).

2. The formulas obtained so far are exact and can be used, for example, to analyze strains in anisotropic rubber but to calculate $\mathcal{H}$ we require (in addition to the values of Ω and $\mathcal{R}$ obtained experimentally) the values of E^a_b determined (in a form parametrized by the eigenvalues of the shear tensor) using formulas (C.7)–(C.7b) from [3].

We shall now make calculations to within the second order with respect to the shear, i.e., neglecting terms $\sim|\Delta|^3$ and higher, where

$$|\Delta| = \sqrt{\langle\Delta^2\rangle} = \sqrt{2k_2},$$

but assuming that the compression (and the pressure) are arbitrary (i.e., particularly having in mind a solid). From the expansions (C.8), (C.8a) from [3] and (51), (51a) we obtain

$$E^1_t = z^{t/2}\left(1 + \frac{t^2}{3}k_2\right),$$
$$E^2_t = z^{t/2}\left(-t - \frac{t^3}{6}k_2\right), \tag{59}$$
$$E^3_t = z^{t/2}\left(\frac{t^2}{2} + \frac{t^4}{4!}k_2\right),$$

and (54a) gives

$$I_3^\gamma = z^3, \tag{60}$$

whence

$$(S')^2 = \frac{1}{z^2}\left(\frac{3}{2} + 2k_2\right), \quad (S')^3 = -\frac{1}{2z^3} \tag{61}$$

and then for the coefficients in (55)

$$C^1 = \frac{1}{2z^2}\left(1 + \frac{4}{3}k_2\right), \quad C^2 = \frac{2}{z^2}\left(1 + \frac{2}{3}k_2\right),$$
$$C^3 = -\frac{2}{z^2}\left(1 + \frac{k_2}{3}\right). \tag{62}$$

From (57), (57a) using one of the relationships proven in [4] for the values of $\mathcal{L}$ [4 (35a)] from (49) we now obtain

$$\mathcal{K} = \tilde{\mathcal{N}} + \frac{k_2}{3}\{\tilde{\mathcal{N}}, \mathcal{Q}\}_+ + \{\tilde{\mathcal{N}}, \mathcal{L}_{13}\}_+$$
$$= \left(1 + \frac{2}{3}k_2\right)\tilde{\mathcal{N}} - \frac{k_2}{3}\{\tilde{\mathcal{N}}, \mathcal{P}\}_+ + \{\tilde{\mathcal{N}}, \mathcal{K}_{13}\}_+ - \{\tilde{\mathcal{N}}, \mathcal{Y}_{13}\}_+, \tag{63}$$

where $\{a, b\}_+ = ab + ba$, it is clear that $\mathcal{K}_{11} = \mathcal{I}$, $\mathcal{Y}_{11} = \mathcal{P}$, and $\mathcal{L}_{11} = \mathcal{Q}$.

As we have noted, the values of $\mathcal{L}_{ab}$ are calculated using (52b) and we have the following representations for $\mathcal{Y}_{13}$

$$\mathcal{Y}_{13} = \frac{1}{3}(\mathcal{P}_{13} + \mathcal{P}_{31}) + a\mathcal{P}_{22} + b(\mathcal{P}_{23} + \mathcal{P}_{32})$$
$$- c\mathcal{P}_{33} = \frac{k_2}{3}(\hat{\mathcal{P}}_{22} - \hat{\mathcal{P}}_{33}). \tag{63a}$$

Here we have

$$a = \frac{1}{3}\left(1 - \frac{1}{2}\operatorname{cosec}^2\varphi\right),$$
$$b = \frac{3k_2k_3}{|g|}, \quad c = \frac{2k_2^2}{|g|}, \tag{63b}$$

and for the definition of $|g|$ and φ see [4 (B.8a)]. The values of $\hat{\mathcal{P}}_{ab}$ are the corresponding projection operators (41c) constructed using the vectors of the orthonormalized basis [4 (B.20)] so that the last expression in (63a) contains no singularities as explained in [4.B]; on the other hand, the coefficients of $\mathcal{P}_{ab}$in (63a) must be calculated exactly, some approximation in terms of $|\Delta|$ was of no interest to us.

The representation of $\mathcal{Y}_{13}$ in an orthonormalized basis is useful for computer calculations near the singularities $|g|$.

It can be seen from (52a) and [4 (B.8a)] that

$$\mathcal{P} = \hat{\mathcal{P}}_{11} + \hat{\mathcal{P}}_{22} + \hat{\mathcal{P}}_{33} = \frac{1}{3}\mathcal{P}_{11}$$
$$+ \frac{1}{|g|}(2k_2^2\mathcal{P}_{22} - 9k_3(\mathcal{P}_{23} + \mathcal{P}_{32}) + 6k_2\mathcal{P}_{33}), \tag{64}$$

and as we have noted, no term can be neglected in the last expression.

3. Hydrostatic compression leads to volume and shear deformation of the sample [1, 3]:

$$x_i = \xi_i/\alpha + u_i', \quad \mu = \alpha\mu', \quad \mu' = E - \beta', \tag{65}$$

and if $\det\mu' = 1$, i.e., $\alpha^2 = z$ (51), the rotation O, as can be seen from (12), is expressed as follows in terms of the "deviator" contribution to the distortion β':

$$O = E + \frac{1}{2}(\beta'^{\mathrm{T}} - \beta')$$
$$+ \frac{1}{8}(3\beta'^{\mathrm{T}2} - \beta'^{q} - \beta'\beta'^{\mathrm{T}} - \beta'^{\mathrm{T}}\beta'), \tag{66}$$

which must be substituted into (49a) and (63) to calculate $\tilde{\mathcal{N}}$. We shall not give the relevant expressions because of their cumbersome nature and the absence of any fundamental difficulties.

4. The apparatus available to experimentalist can usually determine either the isothermal or the adiabatic tensors of the elastic moduli. For example, if the adiabatic tensor of the elastic moduli $\mathcal{K}^\eta$ is obtained from ultrasound measurements, by means of a standard procedure [12] using thermodynamic identities, we can obtain the isothermal tensor which, as we can easily see, can be written in one of the forms:

$$\mathcal{K}^T = \mathcal{K}^\eta - \frac{\rho c_s}{T}\left(\frac{\partial T}{\partial s}\right)_\eta \Big\rangle\Big\langle \left(\frac{\partial T}{\partial s}\right)_\eta$$
$$= \mathcal{K}^\eta - \frac{T}{\rho c_s}\left(\frac{\partial \sigma_0}{\partial T}\right)_s \Big\rangle\Big\langle \left(\frac{\partial \sigma_0}{\partial T}\right)_s, \tag{67}$$

where c_s is the specific heat per unit mass of material for fixed strain and

$$\left(\frac{\partial \sigma_0}{\partial T}\right)_s = -\left(\frac{\partial p}{\partial T}\right)_s E + \left(\frac{\partial \tau_0}{\partial T}\right)_s. \tag{67a}$$

An expression for $\mathcal{N}$ can easily be obtained in an isotropic (before deformation) arbitrarily strained object from (41) which may be of interest for rough orientation during an experiment.

9. HOW DOES THE PROPOSED FORMULATION OF THE THEORY RELATE TO PREVIOUS ONES?

It is appropriate to make some observations on the correspondence between these results and those obtained ear-

lier and discussed widely (see, for example [14] and the literature cited therein).

1. First we assume that only Euler variables are adequate for the strain problem since only in terms of these variables (and only using the concept of natural strain) can the independent quantities of bulk compression and shear be separated (for details see [1, 3]).

2. Then, as explained in [4], the thermodynamic variable in the anisotropic case, similar to the Lagrange tensor for an isotropic medium, is the tensor u_0 determined from

$$\gamma_0 = E - 2u_0 \tag{68}$$

with the components

$$u^0_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} - \frac{\partial u_i}{\partial x_k}\frac{\partial u_j}{\partial x_k}\right) \tag{68a}$$

[a similar construction for an isotopic medium, i.e., for the tensor $\gamma = \mu^T\mu$ (see (2a)) yields an ordinary Lagrange tensor [11] in which the quadratic term with respect to the displacement **u** has the form $-(\partial u_k/\partial x_i)(\partial u_k/\partial x_j)$: the minus sign is attributed to the use of Euler variables]. The derivative of the thermodynamic potential with respect to u_0 in the Murnaghan–Birch equation should be taken to find the stress tensor in an anisotropic solid (the procedure for determining σ for this variant of the theory is described in [4]).

3. The adiabatic tensor of the elastic moduli $\mathscr{C}$ determined in [14] is related to that calculated in the present study (we obviously simplify the index notation and subsequently drop the index "0" for the tensor u: u and u' denote different components of the tensor u) by:

$$C^\eta_{uu'} = \rho\frac{\partial^2\varepsilon}{\partial u\partial u'} = D^s_u H^\eta_{ss'} D^{s'}_{u'} + D^{(2)s}_{uu'}[\sigma_0], \tag{69}$$

and since, as is easy to see from (68),

$$\mathscr{D}^s_u = -2\mathscr{D}, \quad \mathscr{D}^{(2)s}_{uu'} = 4\mathscr{D}^{(2)}, \tag{70}$$

where $\mathscr{D}$ and $\mathscr{D}^{(2)}$ are the quantities used earlier, the relationship between this tensor and $\mathscr{R}$, as is easily seen from (39a), (24), (48), and (50) after taking into account the equality $\mathscr{D}E = -0.5\gamma^{-1}$ derived from [4 (C.5)] is

$$\begin{aligned}\mathscr{C}^\eta &= (S^{-1}\otimes S^{-1})\tilde{\mathscr{R}}(S^{-1}\otimes S^{-1})\\ &\quad - (\mathscr{D}\sigma_0\rangle\langle\gamma^{-1} + \gamma^{-1}\rangle\langle\sigma_0\mathscr{D}),\end{aligned} \tag{69a}$$

which after substituting all the already-known quantities, can then be expanded by a normal procedure in powers of the deviator of the natural strain. Then, the tensor s must be expressed in terms of u as $-0.5\ln(E - 2u)$ using formulas from [4, F] where, as can be seen from (65), $E - 2u = \alpha^2(E - 2u')$. However, without having recourse to the natural strain, it would be necessary to express the result as a function of $u = 0.5(1 - \alpha^2)E + \alpha^2 u'$ and if u' is considered to be a pure deviator contribution determined by the condition $\det(E - 2u') = 1$, it becomes clear that the deviator component is scaled by hydrostatic compression $\alpha^2 = (\rho/\rho_0)^{2/3}$ so that an independent expansion in terms of the hydrostatic and deviator components of the strain becomes impossible which we consider to be a major achievement of natural strain theory.

4. We shall now consider the fluctuations of the strain under pressure. As in [12], we determine the coefficients $C'_{uu'}$ of the quadratic form in terms of deviations from the equilibrium values δu_0 from [4 (64)], this being the expression for the Gibbs potential of an anisotropic object under pressure, $\varphi = f + p/\rho$

$$C'_{uu'} = \rho\frac{\partial^2\varphi}{\partial u\partial u'} = \rho\left(D^s_u\frac{\partial^2\varphi}{\partial s\partial s'}D^{s'}_{u'} + \frac{\partial\varphi}{\partial s}D^{(2)s}_{uu'}\right), \tag{71}$$

whence we obtain

$$\begin{aligned}C'_{uu'} &= D^s_u(H_{ss'} + pE\rangle\langle E)D^{s'}_{u'}\\ &\quad + (\sigma + pE)D^{(2)s}_{uu'}.\end{aligned} \tag{71a}$$

Thus, since in equilibrium $\sigma + pE = 0$, it is found that $\mathscr{C}'$ and $\mathscr{K} = \mathscr{H} + pE\rangle\langle E$ are similar so that the fluctuations of s are related to the matrix $\mathscr{K}$ in the same way that the fluctuations of u_0 are related to $\mathscr{C}'$ and consequently for a hydrostatically compressed sample we can use the stability criteria formulated in [14], replacing the matrix $\mathscr{C}'$ (which must be determined in terms of the derivatives with respect to u_0 and not u) by $\mathscr{K}$.

10. CONCLUSIONS

We have shown that of all the possible variants of determining the tensor of the elastic moduli of an anisotropic object in nonlinear macroscopic theory, only the completely symmetric tensor $\mathscr{K}$ possesses the following properties simultaneously: first, for small strains it yields the tensor of the elastic modulus in linear theory, second, for arbitrary strains it has a minimum number of different components (which is the same as the number of components of the tensor of the elastic moduli in linear theory for each crystal system of an unstrained crystal [11]), and third it can be calculated if we know the dependence of the free energy (or, for its adiabatic value, the energy) on the strain.

All other moduli (which occur in analyses of particular problems): the tensor of the effective elastic moduli $\mathscr{N}$, the thermodynamic susceptibility tensor $\mathscr{Y}$, and the local modification $\mathscr{K}^c$ of the completely symmetric tensor $\mathscr{K}$ are expressed in terms of the tensor $\mathscr{K}$ which is calculated in terms of the tensor $\mathscr{H}$ of the second derivatives with respect to the strain of the free energy per unit mass of the object f (or, for adiabatic coefficients, the energy per unit mass ε).

It has been shown that on the hydrostatic strain line all these quantities possess complete interchange sym-

metry with respect to the indices and have a maximum of 21 different components, where $\mathcal{Y} = \mathcal{K}$, $\mathcal{N} = \mathcal{K}^c$ and the difference between $\mathcal{K}$ and $\mathcal{N}$ [see (63)] is observed in cases where the sample is arbitrarily clamped in the first order with respect to the deviator component of the distortion (as a result of the rotation of an anisotropic sample under pressure) whereas when the sample is clamped symmetrically, eliminating any rotation, it is observed in the second order with respect to the strain deviator, which is nonzero in an anisotropic object even in the hydrostatic case.

Near the hydrostatic strain line the difference between $\mathcal{Y}$ and $\mathcal{K}$ [see (38)], exactly as between $\mathcal{N}$ and $\mathcal{K}^c$ [see (46)] is determined by the shear stress, i.e., is observed in the first order with respect to the resulting additional shear strain, increasing, albeit trivially, the number of different components of the tensors $\mathcal{Y}$ and $\mathcal{N}$.

Bearing all these factors in mind, we find that the value of $\mathcal{K}$ determined by expressions (36)–(37a) and related to the other quantities discussed here by relationships (38) and (39) can reasonably be taken as the tensor of the elastic moduli in nonlinear theory.

ACKNOWLEDGMENTS

The author thanks the management of the Institute for allowing this work to be carried out over such an unreasonable time for our science. The author also thanks S.M. Stishov for impartial criticism of the initial text which acted as the stimulus for writing the present work.

REFERENCES

1. A. P. Kochkin, Indian J. Pure Appl. Math. **17**, 564 (1986).
2. H. Hencky, Z. Tech. Phys. (Leipzig) **9**, 215 (1928).
3. A. P. Kochkin, Zh. Éksp. Teor. Fiz. **109**, 1823 (1996) [JETP **82**, 983 (1996)].
4. A. P. Kochkin, Available from VINITI No. 2120-B99 (June 29, 1999).
5. F. R. Gantmacher, *The Theory of Matrices* (Fizmatgiz, Moscow, 1967, 3rd ed.; Chelsea, New York, 1959).
6. S. P. Demidov, *Theory of Elasticity* (Vysshaya Shkola, Moscow, 1979).
7. L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields* (Nauka, Moscow, 1973, 6th ed.; Pergamon Press, Oxford, 1975, 4th ed.).
8. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics. Vol. 8. Electrodynamics of Continuous Media* (Nauka, Moscow, 1982; Pergamon Press, New York, 1984).
9. Z. G. Pinsker, *Dynamical X-ray Scattering in Ideal Crystalls* (Nauka, Moscow, 1974).
10. M. A. Krivoglaz, *Theory of X-ray and Thermal Neutron Scattering by Real Crystals* (Nauka, Moscow, 1967; Plenum Press, New York, 1969).
11. L. D. Landau and E. M. Lifshitz, *Theory of Elasticity* (Nauka, Moscow, 1987, 4th ed.; Pergamon Press, Oxford, 1986, 3rd ed.).
12. L. D. Landau and E. M. Lifshitz, *Statistical Physics* (Nauka, Moscow, 1976, 3rd ed.; Pergamon Press, Oxford, 1980, 3rd ed.), Part 1.
13. Kwang Yul Kim, Phys. Rev. B **54**, 6245 (1996).
14. Z. Zhou and B. Joos, Phys. Rev. B **54**, 3841 (1996).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 4, 2000, pp. 639–645.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 4, 2000, pp. 735–742.

SOLIDS
Structure

Localized Phonon States at the Edge of the Continuous Spectrum

L. A. Falkovsky

Landau Institute of Theoretical Physics, Russian Academy of Sciences, Moscow, 117334 Russia
Groupe d'Etudes des Semiconducteurs, cc074 UM2-CNRS 34095, Montpelier, France
e-mail: falk@ipt.ac.ru
Received November 16, 1999

Abstract—An analysis is made of the influence of defects on the averaged Green's function of optical phonons whose imaginary part is proportional to the cross section for Raman one-phonon light scattering, which depends on the frequency transfer. A variant of the "cross" technique which can take into account the localized states at defects is used. The defects are assumed to be two-dimensional (of the dislocation type) at which localized states exist near the edge of the continuous spectrum interacting weakly with the defect. As a result of this interaction, which depends on the defect concentration, the states of the continuous spectrum are shifted and broadened, the inhomogeneous broadening depending strongly on the frequency transfer (phonon density of states effect), which leads to asymmetry of the Raman line. The defect concentration also influences the localized states which interact mainly via band states. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Since the appearance of the first studies on localized states [1,2], numerous theoretical and experimental investigations have addressed this topic. The simplest problem which can be solved exactly is the problem of states at a single isotopic defect [1, 3]. Unfortunately, in this example the frequency of the localized state is separated from the edge of the continuous spectrum by a finite frequency interval and if the difference between the masses of the matrix and defect atoms is small, a localized state may not exist. Thus, the state at the defect is determined by the integral characteristics of the entire phonon spectrum and in addition to interaction between the defect and the matrix, it is also necessary to know the complete density of states of the initial unperturbed problem.

Of particular interest is the case of a low but finite concentration of defects when the interaction of states at neighboring defects cannot be neglected [4]. It is then necessary to solve the problem of phonon scattering in all its aspects, taking into account the possible formation of a band of localized states and a transition to extended states [5–11]. The situation is reminiscent of the problem of impurity levels in semiconductors. An important difference, however, is that as a result of the low effective mass of the carriers and the large dielectric constant in semiconductors the impurity levels are shallow and we can usually talk about the density-of-states tails in the band gap. The states at an isotopic defect are always deep.

Finally, another characteristic of the observed phonon spectra is the necessary presence of so-called natural phonon width even at zero temperature and for ideal lattices. This width is associated with the possible decay of phonons (of frequency ω_0) into pairs and is described by third- and higher-order anharmonicities. The relative magnitude of the natural width Γ^{nat}/ω_0 is of the order of the square of the ratio of the amplitude of the atomic vibrations to the interatomic spacing and is usually around 10^{-2}.

In the present study we attempt to cope with these difficulties by considering localized states at anisotropic defects and specifically at two-dimensional dislocation defects. In this case, the localized states exist at an arbitrarily weak defect potential. As we know from quantum mechanics, the energy of a state in a two-dimensional quantum well is exponentially close to the edge of the continuous spectrum. Consequently, the localized state at a two-dimensional defect is determined by the immediate vicinity of the phonon branch extremum. Interaction of these defects among themselves and with the continuous spectrum should be considered in terms of a unified scheme in a suitable order of the defect concentration.

Significantly, the interaction of phonons with defects is elastic unlike their interaction with each other. For this reason the contribution of defects to the phonon width depends on the behavior of the density of states near the edge of the continuous spectrum and is therefore a particular function of the frequency variable. This can easily be confirmed by writing the probability of phonon scattering by a defect potential according to the "golden rule" and considering the $\delta(\omega^2 - \omega_0^2 + s^2k^2)$ function which takes into account the energy conservation law for a phonon near, say, the maximum of the ω_0

branch. For the case of scattering by a point defect, the phonon width goes to zero proportionately as $\sqrt{\omega_0^2 - \omega^2}$, for a linear defect we have a finite jump, and for a planar defect we have a root divergence. At the same time, the natural width Γ^{nat} is a smooth function of ω (almost constant in a small range of variation of ω) and since we are interested in the small vicinity of ω_0, this width can be taken into account (in the phonon spectrum or in the expression for the retarded Green's function) by adding the imaginary constant $i\Gamma^{nat}$ to the frequency variable ω.

It can be seen from the formula containing the δ-function given above that we are interested in distances on the scale $1/k \sim s/\sqrt{\omega_0^2 - \omega^2} \sim s/\sqrt{\omega_0 \Gamma^{nat}}$. Since the dispersion parameter s has the order of the velocity of sound and for optical phonons $\omega_0 \sim \pi s/a$, where a is the lattice parameter, the characteristic values $1/k \sim a\sqrt{\omega_0/\Gamma^{nat}}$ are large on the atomic scale. At present no experimental information is available on the distances within which the defect potential should fall. Calculations of the phonon spectra are usually made using a scheme which assumes that the non-Coulomb part of the atomic interaction may be taken into account using an approximation of several nearest neighbors. This is equivalent to assuming short-range interaction. In addition, it has been established that a considerable fraction of the perturbation caused by the lattice mismatch in heterojunctions also relaxes over atomic distances. For this reason we shall consider short-range defects, i.e., we shall assume that their radius r_0 is small compared with the characteristic value $1/k$ ($r_0 \ll a\sqrt{\omega_0/\Gamma^{nat}}$) and take the Fourier component of the interaction potential as being constant in the range of k of interest to us. In this case, an isotopic defect only differs in dimension from a linear or planar defect.

Finally, let us address the experimental side of the problem. So far the most accurate information on the phonon spectra has been obtained using Raman light scattering, in which a frequency resolution of the order of 1 K has been achieved. This means that not the density of states but the Raman scattering line itself can be studied under the influence of various factors, such as pressure, ion implantation, defects, and so on (see, for example [12]). A study of the role of defects is of particular interest in semiconductor applications where a micro-Raman scattering technique is employed, using laser beams having a spot size of the order of the wavelength.

Our aim is to calculate the Green's functions of optical phonons (whose imaginary part directly gives the one-phonon Raman scattering cross section) using a model of disorder induced by two-dimensional short-range defects.

2. CONTRIBUTION OF LOCALIZED VIBRATIONS TO PHONON SCATTERING

The existing "cross" technique [13] is usually used to study scattering of band quasiparticles at randomly distributed impurities. In our case, this should be modified to allow for localized states. We shall consider the equation for the retarded Green's function

$$(H - i\omega\Gamma^{nat} + U(\mathbf{r}) - \omega^2)D(\mathbf{r}, \mathbf{r}', \omega) = \delta(\mathbf{r} - \mathbf{r}'), \quad (1)$$

where the matrix

$$H_{ij} = \omega_0^2\delta_{ij} + \mu_{ijlm}\frac{\partial^2}{\partial x_1 \partial x_m}$$

describes the long-wavelength expansion of the dynamic matrix near the branch extremum, the natural width Γ^{nat} allows for phonon interaction with each other, and the perturbation

$$U_{ij}(\mathbf{r}) = \sum_n u_{ij}(\mathbf{r} - \mathbf{r}_n) \quad (2)$$

is the interaction with defects positioned at points $\mathbf{r}_n$. For linear defects the vector $\mathbf{r} - \mathbf{r}_n$ is a two-dimensional vector in the plane perpendicular to the defect axis. The subscript j allows for any degeneracy of the phonon branch which may be substantial for Raman phonons at the center of the Brillouin zone.

We shall expand the solution of equation (1) in powers of the perturbation U, which we write in terms of the Fourier components

$$U_{ij}(\mathbf{r}) = \sum_{n,q} u_{ij}(\mathbf{q})\exp[i\mathbf{q}\cdot(\mathbf{r} - \mathbf{r}_n)], \quad (3)$$

and at each step we shall average, i.e., integrate the resulting expression

$$\frac{1}{V}\int d\mathbf{r}_n \ldots,$$

over the defect configuration. The Green's function in the zeroth approximation depends on the difference of coordinates, and its Fourier component has the form

$$D_{ij}^0(k, \omega) = \delta_{ij}(\omega_0^2 - s^2k^2 - i\omega\Gamma^{nat} - \omega^2)^{-1}, \quad (4)$$

where, to simplify the notation, we neglect the dependence of the dispersion parameter on direction.

The first-order correction is $-cD_0(k, \omega)u_0D_0(k, \omega)$, where $u_0 = u(q = 0)$ and c is the defect concentration; for two-dimensional defects this is calculated per unit surface area perpendicular to the defect axis. Here we conserved the sequence of the functions D and u since they are matrices. However, we shall not observe this in the subsequent intermediate calculations.

In the second order, we have, first, the contribution $c^2u_0^2D^3(k, \omega)$ (Fig. 1a) which is obtained if the terms corresponding to different defects are taken in both

sums over the defects and, second, the contribution from the same defect (Fig. 1b):

$$cu_0^2 D_0^2(k, \omega)\sum_q D_0(q, \omega).$$

We also assume that the number of defects is large in a sample. In addition, as we noted in the Introduction, small $q \sim \sqrt{\Gamma^{nat}/\omega_0}/a$ are significant when integrating over q. Thus, when calculating the second contribution

$$\sum_q |u(\mathbf{k}-\mathbf{q})|^2 D_0(q, \omega) = u_0^2 \sum_q D_0(q, \omega),$$

the factor $|u(\mathbf{k}-\mathbf{q})|^2$ is taken the zero argument and the remaining logarithmic integral may be cut off at large q, of the order of the reciprocal defect radius. Note that the result, i.e., the correction to the Green's function, is of interest to us for even smaller $k \ll \sqrt{\Gamma^{nat}/\omega_0}/a$ and specifically for $k \sim 10^{-3}/a$, which corresponds to the optical range, since this is the range in which Raman light scattering is studied experimentally.

Simple calculations yield the following result for the integral:

$$\sum_q D_0(q, \omega) = -\frac{1}{4\pi s^2}\left[\frac{1}{2}\ln\frac{s^4/r_0^4}{(\omega^2-\omega_0^2)^2+(\omega\Gamma^{nat})^2} - i\left(\frac{\pi}{2}\mathrm{sgn}(\omega) - \arctan\frac{\omega^2-\omega_0^2}{\omega\Gamma^{nat}}\right)\right]. \quad (5)$$

We note that both contributions (Figs. 1a, 1b) can be of the same order despite the fact that at low concentration c the first contains an extra power of c. This is because for $\omega \longrightarrow \omega_0$ this contribution has an additional pole whereas the second is only logarithmically large.

Four types of terms appear in the third order. The contribution

$$-c^3 u_0^3 D^4(k, \omega),$$

shown in Fig. 2a is obtained if the terms corresponding to different defects are taken in all the sums over defects. The contribution from the same defect is shown in Fig. 2b:

$$-cu_0^3 D_0^2(k, \omega)\left(\sum_q D_0(q, \omega)\right)^2.$$

The "mixed" contribution (Fig. 2c) has the form

$$-2c^2 u_0^3 D_0^3(k, \omega)\sum_q D_0(q, \omega),$$

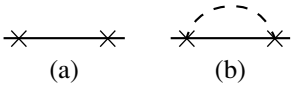

Fig. 1. Second-order diagrams. The solid lines show the Green's functions of an optical phonon, the crosses denote interaction with defects, and the dashes show the transfer of momentum; integration over the momentum transfer is performed in the next Green's function.

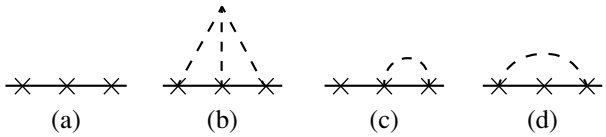

Fig. 2. Third-order diagrams for interaction with a defect. The connected dashed lines indicate conservation of total momentum.

where the coefficient 2 allows for the number of these diagrams, and the "cross" term (Fig. 2d) is given by

$$-c^2 u_0^3 D_0^2(k, \omega)\sum_q D_0^2(q, \omega).$$

The last two terms are of the same order in terms of the concentration c although, as we can see, the integral in the cross term gives a pole of the type $D_0(k, \omega)$ and both terms have an equal number of poles whereas the mixed term has an additional large logarithm $\ln(\omega_0/\Gamma^{nat})$ which has appeared as a result of integrating over q. For this reason we drop the cross term.

In the fourth order with respect to the interaction we have four terms of different order in terms of concentration: the fourth-order term

$$c^4 u_0^4 D^5(k, \omega),$$

the first-order term

$$cu_0^4 D_0^2(k, \omega)\left(\sum_q D_0(q, \omega)\right)^3,$$

several diagrams of order c^3 among which the large logarithm only contains the mixed contribution,

$$3c^3 u_0^4 D_0^4(k, \omega)\sum_q D_0(q, \omega),$$

and finally, diagrams of order c^2:

$$c^2 u_0^4 D_0^3(k, \omega)\left(\sum_q D_0(q, \omega)\right)^2,$$

$$c^2 u_0^4 D_0^3(k, \omega)\sum_q D_0^2(q, \omega),$$

which contain the square of the large logarithm.

The following orders in terms of defect interaction can then be analyzed similarly. In each order in terms of concentration the cross terms are small compared with this large logarithm. The remaining terms, allowing for coefficients of 2 in the diagram in Fig. 2c, become a geometric progression. Summing this yields the following result for the Fourier component of the averaged Green's function:

$$D(k,\omega) = D_0(k,\omega) + cD_0^2(k,\omega) \times \sum_{n=0}^{\infty}(-u_0)^{n+1}\left(cD_0(k,\omega) + \sum_q D_0(q,\omega)\right)^n = \left(\frac{1}{D_0(k,\omega)} + \frac{cu_0}{\left(1 + u_0\sum_q D_0(q,\omega)\right)}\right)^{-1}. \quad (6)$$

An expression of the type (6) was obtained in [14] for scattering by point isotopic defects by closing the equation for the scattering matrix. The difference in this case is that first, we are considering two-dimensional defects and second, we are taking into account natural phonon damping. This allows us to estimate the order of the omitted diagrams (in terms of $\ln(\omega_0/\Gamma^{nat})$) and also to consider interaction between localized and extended states.

3. OPTICAL MODE AND LOCALIZED STATES IN WEAK COUPLING REGIME

At low defect potential u_0 the Green's function (6) has two obvious singularities. One, near the pole $D_0(k, \omega)$, corresponds to the band states of a crystal containing defects and the other, near the zero of the expression $1 + u_0\sum_q D_0(q, \omega)$ occurs for $u_0 > 0$ and describes states localized at defects. At a fairly strong potential u_0 the frequency spacing between these states is large compared with their width and the corresponding contributions may be separated.

3.1. Influence of Defects on Shift and Damping of the Optical Mode

The presence of defects leads to a shift and additional width of the optical mode. Near the pole D_0 the slowly varying real part of the expression $1 + u_0\sum_q D_0(q, \omega)$ [see (5)] may be taken for $\omega = \omega_0$ and the Green's function (6) has the form

$$D(k,\omega) = (\Omega^2 - s^2k^2 - i\omega\Gamma - \omega^2)^{-1}, \quad (7)$$

where the new values of the branch edge Ω and the damping Γ are given by

$$\Omega^2 = \omega_0^2 + \frac{n_c u\omega_0^2}{1 - u\ln(s^2/r_0^2\omega_0\Gamma^{nat})}, \quad (8)$$

$$\Gamma = \Gamma^{nat} + \left(\frac{\pi}{2} - \arctan\frac{\omega^2 - \omega_0^2}{\omega_0\Gamma^{nat}}\right) \times \frac{n_c u^2\omega_0}{[1 - u\ln(s^2/r_0^2\omega_0\Gamma^{nat})]^2}, \quad (9)$$

where we introduce the dimensionless coupling constant with a defect $u = u_0/4\pi s^2$ and the number of defects $n_c = 4\pi cs^2/\omega_0^2$ in a region of dimensions s/ω_0 of the order of the defect radius r_0. Note that under the conditions of validity of the formulas given above, and specifically far from the frequency of the localized level, the denominators in (8) and (9) do not vanish.

It can be seen from expressions (8) and (9) that the branch edge and the width vary by a value of the order of Γ^{nat} (the scale observed experimentally) if the defect concentration is $n_c \sim \Gamma^{nat}/\omega_0$. For small u the linear term with respect to u is the influence of the average defect field on the phonon frequency shift and the quadratic term describes the inhomogeneous shift and mode broadening as a result of phonon scattering at defect fluctuations. These were obtained earlier [15] as an expansion in terms of u using the usual diagram technique and were observed experimentally using Raman light scattering near heterojunctions [12]. The fluctuation contribution to the damping is a rapidly varying function of the frequency variable ω near ω_0 on an interval of the order of Γ^{nat}; the factor in brackets in (9) then varies between π and zero. Thus, a singularity appears in the phonon density of states near the edge (maximum) of the continuous spectrum: phonons can only be scattered at defects for $\omega^2 < \omega_0^2$ (in the limit $\Gamma^{nat} \longrightarrow 0$). The influence of defects can be observed experimentally from the increase in the Raman scattering line width and, as we can see, from the appearance of asymmetry in the spectrum. For the maximum of the phonon branch the low-frequency line wing is more gently sloping. This asymmetry may be observed on the curves (Figs. 3, 4) constructed using the exact expression (6). For lines corresponding to the branch minimum the asymmetry will be the reverse. For the case $u > 0$ as the magnitude of the interaction increases the dependences (8) and (9) have a resonant character for $u \simeq 1/\ln(s^2/r_0^2\omega_0\Gamma^{nat})$ whereas for larger u the shift and width do not depend on the interaction. The resulting resonant dependence reflects the possibility that localized states may exist.

3.2. Localized States near the Optical Mode Edge

For $u_0 > 0$ a localized state forms at a single impurity near the edge of the continuous spectrum. This is determined by the equation

$$\mathrm{Re}\left(1 + u_0 \sum_q D_0(q, \omega)\right) = 0, \qquad (10)$$

whose solution allowing for (5) may be written in the form

$$\omega_l^2 = \omega_0^2 \pm [(s/r_0)^4 e^{-2/u} - (\omega_0 \Gamma^{\mathrm{nat}})^2]^{1/2}. \qquad (11)$$

The upper sign in this formula corresponds to the localized state. For the lower sign the frequency lies in the region of the continuous spectrum: in formula (5) even in the limit $\Gamma^{\mathrm{nat}} \longrightarrow 0$ a finite imaginary part $i\pi/2$ exists and this state is in fact quasi-local.

If as before we assume that the interaction u is so strong that the distance between the edge of the continuous spectrum ω_0 and the frequency ω_l is large compared with the damping, the contribution of the localized state to the Green's function may be expressed in the form

$$D(k, \omega) = -n_c u \omega_0^2 (\omega_0^2 - \omega_l^2)^{-2} \times \left[1 + u_0 \sum_q D_0(q, \omega) + n_c u \omega_0^2 (\omega_0^2 - \omega_l^2 - i\omega\Gamma^{\mathrm{nat}})^{-1}\right]^{-1}. \qquad (12)$$

Comparing (12) with (7), we can see that the contribution of the localized states is proportional to the defect concentration, as predicted. The defect concentration appears in the denominator of this last expression which leads to a shift and width of the localized state, such is the effect of interaction with the continuous spectrum. Having expanded the denominator in powers of $\omega^2 - \omega_l^2$, we find the corresponding corrections:

$$\Omega_l^2 = \omega_l^2 + n_c \omega_0^2, \qquad (13)$$

$$\Gamma_l = \Gamma^{\mathrm{nat}} + n_c \omega_0^2 \Gamma^{\mathrm{nat}} / (\omega_l^2 - \omega_0^2). \qquad (14)$$

Curiously, the potential of the interaction with the defect only appears here by way of ω_l [see (11)].

Note that for a quasi-local state the frequency is $\omega^2 < \omega_0^2$ and the imaginary part in (5) is close to π. Then, on the right-hand side of expression (14) the first term Γ^{nat} is replaced by $2\pi|\omega_0 - \omega_l|$, i.e., a quantity larger than the interval between the quasi-local vibration and the edge of the continuous spectrum. This implies that the width of the quasi-local vibrations is comparable to the distance from the edge of the continuous spectrum.

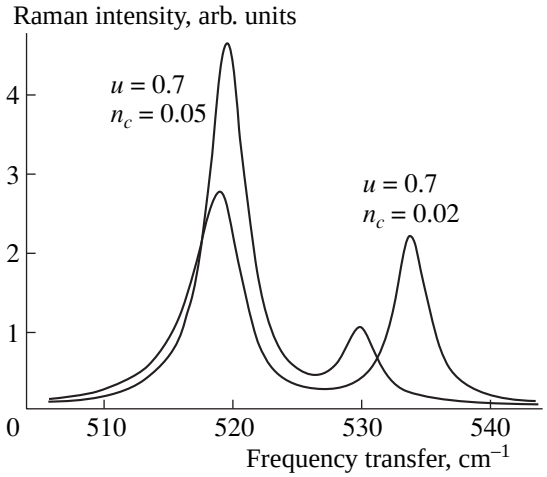

Fig. 3. Dependence of the imaginary part of the phonon Green's function averaged over defects on the frequency transfer ω. The curves were plotted using formula (6) for various defect concentrations and the same phonon–defect interaction. The initial parameters, 520 cm^{-1} line center and 3.2 cm^{-1} natural width, correspond to pure silicon. It is assumed that this line corresponds to the maximum of the optical phonon branch at the center of the Brillouin zone. The peaks at high transferred frequencies reflect a localized state.

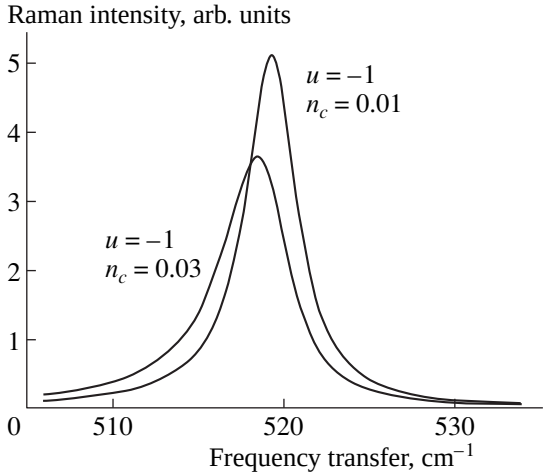

Fig. 4. Frequency dependence of the imaginary part of the Green's function for the case where the possible existence of a localized state is eliminated ($u < 0$).

4. RESONANT COUPLING OF AN OPTICAL MODE WITH LOCALIZED STATES

The small denominators in formulas (8), (9), and (14) are a reflection of the resonant interaction between the band and localized states. When the frequency spacing between these states is small but large compared with the damping, their interaction can be analyzed by expanding the logarithm (5) in the general formula (6) in powers of $\omega^2 - \omega_l^2$, confining ourselves to the first-

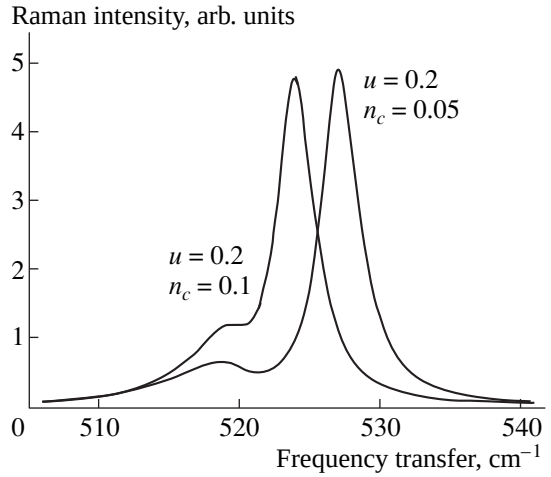

Fig. 5. Frequency dependence of the imaginary part of the Green's function at high defect concentration.

order term. We find that the poles of the Green's function (6) are related to the zeros of the expression

$$D_0^{-1}(k = 0, \omega)\left(1 + u_0 \sum_q D_0(q, \omega)\right) + cu_0 \propto (\omega^2 - \omega_l^2)(\omega_0^2 - \omega^2) + n_c\omega_0^2(\omega_l^2 - \omega_0^2) - i(\omega^2 - \omega_0^2)\omega\Gamma(\omega), \quad (15)$$

where, as before, ω_l is determined by expression (11). The frequencies

$$\omega_{1,2}^2 = \frac{\omega_0^2 + \omega_l^2}{2} \pm \left[\frac{(\omega_0^2 - \omega_l^2)^2}{4} + n_c\omega_0^2(\omega_l^2 - \omega_0^2)\right]^{1/2}, \quad (16)$$

at which the real part of this expression vanishes give the positions of the band and localized states for $k = 0$ and thus the Raman scattering maxima. The condition for validity of the expansion of the logarithm given above is that the second term under the root in (16) should be small compared with the first and then this formula is reduced to the form

$$\omega_1^2 = \omega_l^2 + n_c\omega_0^2, \quad \omega_2^2 = \omega_0^2 - n_c\omega_0^2. \quad (17)$$

The first formula (17) is the same as (13) while the second differs from (8) in the limit of large u by the absence of a logarithm.

The value of the imaginary part in (15) depends strongly on frequency. Near the value of ω_1 describing the shift of the localized states, the damping is $\Gamma(\omega_1) = \Gamma^{\text{nat}}$. However, for the value of ω_2 determining the new edge of the initial continuous spectrum we find $\omega\Gamma(\omega_2) = \pi(\omega_l^2 - \omega_0^2)\operatorname{sgn}\omega$.

Using expressions (15) and (16) we find the imaginary part of the Green's function:

$$\operatorname{Im}D(k = 0, \omega) = \operatorname{Im}\frac{\omega^2 - \omega_l^2 + i\omega\Gamma^{\text{nat}}}{(\omega_1^2 - \omega^2)(\omega^2 - \omega_2^2) - i(\omega^2 - \omega_0^2)\omega\Gamma(\omega)}, \quad (18)$$

which gives the Raman scattering cross section.

Unlike (7) and (12), expression (18) simultaneously describes band and localized states at low defect concentration. The concentration dependence of their position can be seen from formula (17). We note a difference in their width. Whereas the width of the localized states is mainly determined by their natural width Γ^{nat}, scattering at defects makes a significant contribution to the width of the band states which is given by $\Gamma = \pi n_c\omega_0$.

5. DISCUSSION OF RESULTS

Figures 3–5 show dependences of the Raman scattering cross section, i.e. the imaginary part of the Green's function of the optical phonons, on the frequency transfer ω. The graphs are plotted using the exact expression (6) for values of the frequency $\omega_0 = 320$ cm^{-1} at the center of the Brillouin zone and the natural width $\Gamma^{\text{nat}} = 3.2$ cm^{-1} determined experimentally for silicon. The defect parameters were varied, i.e., the dimensionless concentration n_c, the number of defects in a region of the order of the interaction radius (which in (5) was assumed to be $r_0 = \pi s/\omega_0$), and the dimensionless coupling constant u. For $u < 0$ no localized states exist and an increase in interaction or concentration leads to a shift and broadening of the edge of the continuous spectrum in accordance with formulas (8) and (9). For $u > 0$ a localized state appears outside the continuous spectrum (for $\omega^2 > \omega_0^2$) for a maximum at the band center. For a minimum of the optical branch at the center of the Brillouin zone, the localized state appears for $u < 0$ and its frequency lies below the edge of the branch. As the interaction increases, the localized state becomes separated exponentially rapidly from the continuous spectrum in accordance with (13) whereas the edge of the continuous spectrum is displaced nonmonotonically (8).

The mutual influence of localized states and states of the continuous spectrum on their width is particularly large when the distance between them is relatively small [see (9) and (14)]. We note that the dependence of the width of the band and localized states on the natural damping differs significantly for $\Gamma^{\text{nat}} \longrightarrow 0$; for the band state a weakly varying contribution remains as a result of scattering at defects, as given by the second term in (9) whereas the width (14) of the localized state vanishes.

Figure 5 shows the Raman spectrum at high defect concentration when the band contribution is barely

discernible from the series with a sharp peak created by the localized states.

To conclude we note that we have considered the case of the lowest possible defect concentration c: in the denominator of expression (6) all corrections linear in c were summed. The next step involves a self-consistent consideration, i.e., replacing $D_0(q, \omega)$ with $D(q, \omega)$ after the summation sign in the last equality in (6). This so-called renormalized approximation was investigated in [16].

ACKNOWLEDGMENTS

A considerable proportion of this work was carried out during the author's stay at GES, University of Montpellier II, France. The author thanks J. Camassel for cooperation and hospitality and would also like to thank S. Meshkov and V. Fateev for discussions. This work was supported by the Russian Foundation for Basic Research.

REFERENCES

1. I. M. Lifshits, Zh. Éksp. Teor. Fiz. **17**, 1017 (1947); I. M. Lifshits, Nuovo Cimento Suppl. **3**, 716 (1956).
2. F. J. Dyson, Phys. Rev. **92**, 133 (1953).
3. P. G. Dauber and R. J. Elliott, Proc. R. Soc. London **273**, 222 (1963).
4. I. M. Lifshits, Usp. Fiz. Nauk **83**, 617 (1964) [Sov. Phys. Usp. **7**, 549 (1965)].
5. I. M. Lifshits and A. M. Kosevich, Rep. Prog. Phys. **29** (1), 217 (1966).
6. I. M. Lifshits, S. A. Gredeskul, and L. A. Pastur, J. Stat. Phys. **38**, 37 (1985).
7. Yu. M. Kagan and Ya. A. Ioselevskiĭ, Zh. Éksp. Teor. Fiz. **45**, 819 (1963) [Sov. Phys. JETP **18**, 562 (1963)].
8. T. R. Kirkpatrick, Phys. Rev. B **31**, 5746 (1985).
9. M. A. Ivanov and Yu. G. Pogorelov, Zh. Éksp. Teor. Fiz. **72**, 2198 (1977) [Sov. Phys. JETP **45**, 1155 (1977)].
10. H. Böttger, *Principles of the Theory of Lattice Dynamics* (Physik-Verlag, Weinheim, 1983; Mir, Moscow, 1986), Chap 2.
11. I. Ya. Polishchuk, A. L. Burin, and L. A. Maksimov, Pis'ma Zh. Éksp. Teor. Fiz. **51**, 644 (1990) [JETP Lett. **51**, 730 (1990)].
12. L. A. Falkovsky, J. M. Bluet, and J. Camassel, Phys. Rev. B **55**, R14697 (1997); **57**, 11283 (1998).
13. A. A. Abrikosov, L. P. Gor'kov, and I. E. Dzyaloshinskiĭ, *Methods of Quantum Field Theory in Statistical Physics* (Nauka, Moscow 1963).
14. D. V. Taylor, in *Dynamical Properties of Solids*, Ed. by G. K. Horton and A. A. Maradudin (North-Holland, Amsterdam, 1975), Vol. 2, p. 285.
15. L. A. Falkovsky, Pis'ma Zh. Éksp. Teor. Fiz. **66**, 817 (1997) [JETP Lett. **66**, 860 (1997)].
16. L. A. Falkovsky, Pis'ma Zh. Éksp. Teor. Fiz. **71**, 225 (2000) [JETP Lett. **71**, 155 (2000)].

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 4, 2000, pp. 646–654.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 4, 2000, pp. 743–752.

SOLIDS
Electronic Properties

Quantum Hall Effect in an Antidot Lattice: Macroscopic Limit

É. M. Baskin and M. V. Éntin*
Institute of Semiconductor Physics, Siberian Division, Russian Academy of Sciences, Novosibirsk, 630090 Russia
** e-mail: Entin@isp.nsc.ru*
Received February 8, 1999

Abstract—The quantum Hall effect in a 2D system with antidots is studied. The antidots are assumed to be large compared with the quantum and relaxation lengths. In this approximation the electric field in the system can be described by the continuity equation. It is found that the electric field in a system without conducting boundaries can be expressed in terms of the same system without a magnetic field. Specific problems of the electric field and current in structures containing one or two antidots and in a circular disk with point contacts are solved. The effective Hall and longitudinal conductivities in a sample containing a large number of randomly distributed antidots are found. In the limit of zero local longitudinal conductivity, the effective longitudinal conductivity also vanishes, and the Hall conductivity is equal to the local conductivity. The corrections to the conductivity tensor which are due to the finiteness of the local conductivity are obtained. Breakdown of the quantum Hall effect in a lattice of antidots is studied on the basis of the assumption that a high current density in narrow locations of the system results in overheating of the electrons. Local and nonlocal models of overheating are studied. The high-frequency effective conductivity of a system with antidots and the shift of the cyclotron resonance frequency are found. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

It is known that various microscopic approaches lead to quantization of the Hall conductivity and vanishing of the drift conductivity. These approaches include the microscopic approximation, based on a local relation between the electric field and the current density, and an approach based on edge currents which attributes the quantum Hall effect (QHE) to currents flowing along the boundaries.

The standard microscopic theories of the QHE are based on the derivation of an expression for the conductivity tensor in a uniform electric field and neglect the spatial fluctuations of the external electric field. In contrast to them, the quantum Hall conductivity is considered in some recent works to be a macroscopic phenomenon, using a quantum Hall conductivity and zero drift conductivity of an ideal system as the zeroth approximation, and the problem of current flow in a mixture of a Hall conductor and an ideal metal [1] or in Hall conductors on various Hall plateaus [2, 3] is solved. This approach leads to a so-called "semicircle relation" between the Hall and drift components that does not contain Planck's constant. This approach correctly describes the experimental data for the integral and fractional Hall effects.

The macroscopic approach expands the range of applicability of microscopic theories by taking account of the spatial variations of the applied electric field. Thus, the main result "survives" independently of the presence of boundaries, mixing of the Hall phases, and so on. Nonetheless, the question of the limits of the Hall quantization remains.

The purpose of the present paper is to investigate the electric fields and currents in a nonuniform two-dimensional system in the quantum Hall effect regime. We shall consider this system to be a mixture of a uniform Hall conductor and an insulating phase, neglecting the screening length. This approach is applicable, for example, to a QH system with strong doping and compensation, where the insulating regions are formed as a result of a randomly decreased donor density. Specifically, in the quantum limit of a single partially filled Landau level the system can be treated as a mixture of insulating regions, where the first Landau level is empty, and a Hall insulator, where the first Landau level is completely filled. Another example are artificially structured systems: antidot lattices, where the problem of the current density distribution is important in itself, specifically, in the problem of breakdown of QHE [4].

The method is based on mapping this system into an analogous conductor without a magnetic field. It will be shown that the spatial distribution of the electric field in a conductor with and without a magnetic field with the same geometry is identical. The longitudinal and transverse effective conductivities can be expressed in terms of the effective conductivity of a system without a magnetic field. If the Hall phase is connected, then the

effective conductivity also vanishes in the limit of an ideal Hall insulator, $\sigma_{xx} \longrightarrow 0$.

It is known that a high current density results in breakdown of QHE. We shall find the breakdown threshold, studied experimentally in [4], in an antidot lattice.

The formulas obtained for the conductivity are also applicable in the limit of a high-frequency field. The absorption of an electromagnetic field in an antidot lattice will be studied and the shift of the cyclotron resonance line will be found.

2. IDEAL HALL INSULATOR WITH INSULATING INCLUSIONS

We shall examine first a mixture of a Hall insulator and an ordinary insulator. In the Hall phase the expression for the current has the form

$$\mathbf{j} = \sigma_{xx}\nabla\phi + \sigma_{xy}\mathbf{e_B} \times \nabla\phi, \tag{1}$$

where $\mathbf{e_B}$ is the magnetic-field unit vector. In the limit of a Hall insulator $\sigma_{xx} \longrightarrow 0$. Nonetheless, we retain σ_{xx} in equation (1) in order to determine the potential ϕ. For example, the longitudinal conductivity can be assumed small but finite because of the finiteness of the temperature.

It follows from the continuity equation $\nabla\mathbf{j} = 0$ that

$$\sigma_{xx}\Delta\phi = 0. \tag{2}$$

The boundary conditions at the boundary of the Hall and ordinary insulators with the normal $\mathbf{n}$ on the side of the Hall insulator have the form $\mathbf{jn} = 0$. In the limit $\sigma_{xx} \longrightarrow 0$ the boundary conditions give

$$\mathbf{n} \times \nabla\phi_0 = 0. \tag{3}$$

The last expression does not contain σ_{xx}. The potential is a harmonic function with a zero tangential field at the boundary of the insulator. Therefore an ordinary insulator in a Hall insulator can be treated as an ideal conductor embedded in a conducting medium.

We shall now consider a random, on the average uniform, infinite medium consisting of a Hall insulator plus an inclusion consisting of an insulator with an average field directed along the y axis $\overline{\nabla\phi_0} = (0, \overline{E_y})$. (The overbar denotes averaging over an area: $\bar{A} = \int A\, dS/S$.)

The effective Hall conductivity is defined as

$$\sigma_{xy}^{\text{eff}} = \frac{\bar{j}_x}{\overline{\nabla_y\phi_0}} = \sigma_{xy}\frac{\int_{\text{Hall}} \nabla_y\phi_0 dS}{\int \nabla_y\phi_0 dS}. \tag{4}$$

The index "Hall" means that the integration extends only over the Hall phase. Let us assume that the insulating phase consists of closed disconnected regions. Since the potential on their boundaries is constant, the integral of the field over the insulator vanishes:

$$\int_{ins} \nabla_y\phi_0 dS = \int dx \int_{y_1(x)}^{y_2(x)} dy \nabla_y\phi_0 = \int dx(\phi_0(y_2(x)) - \phi_0(y_1(x))) = 0. \tag{5}$$

Therefore the numerator and denominator in equation (4) are identical and $\sigma_{xy}^{\text{eff}} = \sigma_{xy}$. This proves that the Hall plateau is stable with respect to the appearance of insulating inclusions: The effective conductivity does not change, while the insulating phase does not form an infinite connected region, breaking up the Hall phase into disconnected regions.

3. HALL CONDUCTOR WITH INSULATING INCLUSIONS

We shall study below the problem of finite σ_{xx}. Let us consider the analytic function $w = \phi_0 + i\psi_0$, where ϕ_0 and ψ_0 satisfy the Cauchy–Riemann equations:

$$\nabla_x\phi_0 = \nabla_y\psi_0, \quad \nabla_y\phi_0 = -\nabla_x\psi_0. \tag{6}$$

In terms of the dual potential ψ_0

$$\Delta\psi_0 = 0, \quad \mathbf{n}\nabla\psi_0 = 0, \quad \int \nabla_y\psi_0 dS = 0. \tag{7}$$

If the relation (7) is applied to the expression for the current $\mathbf{j} = \sigma\nabla\psi$, they correspond to the longitudinal conductivity of the medium with a local longitudinal conductivity σ, containing insulating inclusions in the absence of a magnetic field.

We shall now express the electric field and current density in terms of the solution of the problem without a magnetic field. Let the function ϕ be determined by the equation

$$\phi = \phi_0 \sin\alpha + \psi_0 \cos\alpha = \text{Im}(e^{i\alpha}w). \tag{8}$$

The function ϕ satisfies the boundary conditions $\mathbf{nj} = 0$ with $\mathbf{j}$ from equation (1), if the angle α is

$$\cos\alpha = \frac{\sigma_{xx}}{\sqrt{\sigma_{xx}^2 + \sigma_{xy}^2}}, \quad 0 < \alpha < \pi/2. \tag{9}$$

As a result, the solution ϕ of the boundary-value problem with the magnetic field B differs from the solution in the absence of a magnetic field, ψ_0, by a rotation of the electric field at any point by an angle α:

$$\mathbf{E}(\mathbf{r}, B) = \hat{U}(\alpha)\mathbf{E}(\mathbf{r}, 0),$$

where $\hat{U}(\alpha)$ is the matrix giving a rotation by the angle α.

Let us now consider the boundary-value problem with a prescribed electric field at infinity $\mathbf{E}(\infty)$. The electric field at the point $\mathbf{r}$ is related with $\mathbf{E}(\infty)$ by a linear transformation $\hat{G}$:

$$\mathbf{E}(\mathbf{r}) = \hat{G}\mathbf{E}(\infty).$$

Transforming the field at the point $\mathbf{r}$ and at infinity, we find that

$$\hat{G}\big|_B = \hat{U}(\alpha)\hat{G}\big|_{B=0}\hat{U}(-\alpha), \tag{10}$$

i.e., the exterior boundary conditions must be rotated by an angle $-\alpha$, after which the equation must be solved in the absence of a magnetic field and the electric field obtained must be rotated by the angle α.

The formulation of the problem with the electric field prescribed at infinity assumes that the sample has no boundaries, and all nonuniformities lie in a bounded region. In the problem of the effective properties of the medium, prescribing the average field plays the role of a boundary condition at infinity. Therefore the relation (10) also determines a relation between the local and average fields.

It is easy to see that the current density is a potential field with the potential $\sqrt{\sigma_{xx}^2 + \sigma_{xy}^2}\,\psi_0$. Indeed,

$$\begin{aligned}\mathbf{j}(\mathbf{r}) &= \sqrt{\sigma_{xx}^2 + \sigma_{xy}^2}\hat{U}(-\alpha)\mathbf{E}(\mathbf{r}, B)\\ &= \sqrt{\sigma_{xx}^2 + \sigma_{xy}^2}\hat{U}(-\alpha)\hat{U}(\alpha)\mathbf{E}(\mathbf{r}, 0)\\ &= \sqrt{\sigma_{xx}^2 + \sigma_{xy}^2}\nabla\psi_0.\end{aligned} \tag{11}$$

The equation (11) and the boundary conditions for the current at the boundary of an insulator do not contain the magnetic field. Therefore the magnetic field does not affect the local current density in problems with (a) a prescribed current density at infinity and (b) a fixed average current in a sample with no boundaries.

We note that these results must be used with care. For example, they cannot be used for a sample with metallic contacts that violate the conditions of the theorem.

The case (a) can be used to describe current flow around a limited number of insulating inclusions in a Hall conductor. The case (b) can be used to describe current in microstructures with point contacts and for a large statistically homogeneous system containing insulating inclusions.

4. ELECTRIC FIELD AROUND ONE AND TWO ANTIDOTS

We shall discuss below various consequences of the method proposed above. The projection of the problem of the current and electric field in the QHE regime on the problem in a zero magnetic field makes it possible to find the field and current in various geometries, which can be solved using conformal mappings. For example, let us consider the distribution of fields around one and two antidots with radii a. In the first case the electric field

$$E(z) = E_x + iE_y$$

is determined by the expression

$$E^*(z) = E^*(\infty) - E(\infty)e^{-2i\alpha}a^2/z^2. \tag{12}$$

The case of two close antidots is important for describing breakdown of QHE in antidot lattices [4], which is attributed to concentration of current density between antidots. The solution can be found similarly to [5]. A conformal mapping of two circles with centers at $\pm(a+\delta)$ onto the ring $1/R < |w| < R$ in the complex w plane,

$$R = \frac{\sqrt{2a+\delta} + \sqrt{\delta}}{\sqrt{2a+\delta} - \sqrt{\delta}},$$

is given by the expression

$$w = \frac{z+b}{z-b}, \quad b = \sqrt{2a\delta + \delta^2}. \tag{13}$$

The boundary condition at infinity for the electric field transforms into the potential of a dipole at the point $w = 1$. The electric field is determined by the expression

$$\begin{aligned}E^* = 4b^2\Bigg[&E^*(\infty)\sum_{n=\infty}^{\infty}\frac{R^{4n}}{\left(R^{4n}(z-b)-z-b\right)^2}\\ &-E(\infty)e^{-2i\alpha}\sum_{n=\infty}^{\infty}\frac{R^{4n+2}}{\left(R^{4n+2}(z-b)-z-b\right)^2}\Bigg].\end{aligned} \tag{14}$$

In the situation of close antidots, $\delta \ll a$, the strongest field and highest current density are reached near the point $z = 0$. At this point

$$E(0) = \sqrt{a/2\delta}(E^*(\infty) - e^{-2i\alpha}E(\infty))$$

diverges as $\delta \longrightarrow 0$.

Figures 1 and 2 show equipotential lines and current lines around one and two antidots for the cases where $\alpha = 0$, $\alpha = \pi/4$, and $\alpha = \pi/2$. For the QHE problem it can be assumed that in the Hall plateau regime $\sigma_{xx} \longrightarrow 0$, while at the maximum $\sigma_{xx} \sim \sigma_{xy}$, i.e., the angle α changes from zero to $\sim\pi/4$ and then to 0 as the maximum is crossed.

Another easily solvable problem is the problem of the current and field distributions in a large quantum dot with small tunneling contacts (Fig. 3). Tunneling occurs in the thinnest parts of the barriers, so that the contacts can be assumed to be point sources of current. We note that point contacts do not violate the condi-

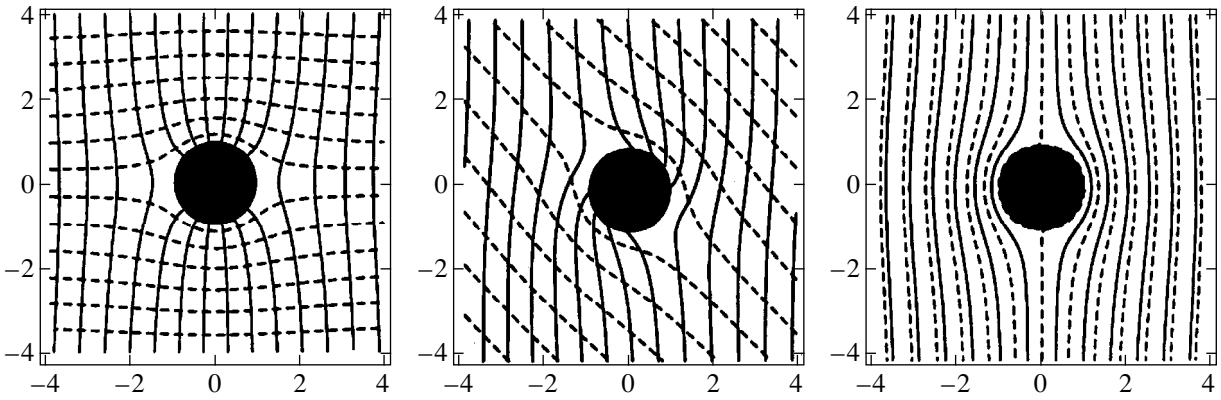

Fig. 1. Equipotential lines (solid lines) and current lines (dashed lines) around an antidot (filled circle). The average electric field is directed along x. Hall angles from left to right: 0, $\pi/2$, and $\pi/4$ (Hall plateau).

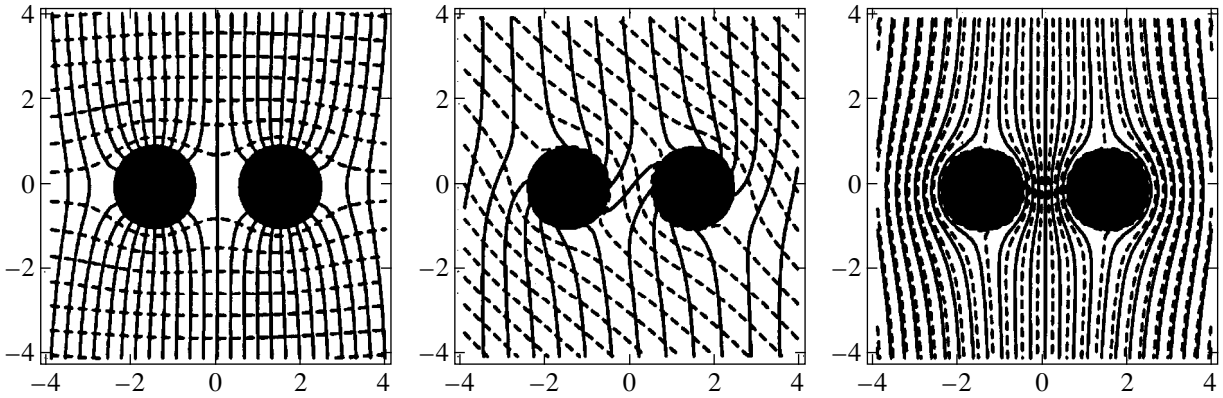

Fig. 2. Same as in Fig. 1, but with two antidots.

tions of the theorem and, in consequence, a magnetic field has no effect on the current distribution, in contrast to contacts of finite size.

The current density and the potential in this case are given by the expressions

$$j_x - ij_y = \frac{2aJ}{\pi}\frac{1}{a^2 - z^2},$$
$$\phi = \frac{J}{\pi\sqrt{\sigma_{xx}^2 + \sigma_{xy}^2}}\mathrm{Re}\left(\exp(-i\alpha)\log\frac{a-z}{a+z}\right), \quad (15)$$

where J is the total current. Figure 3 shows the equipotential lines and the current lines in this case.

5. EFFECTIVE CONDUCTIVITY OF AN ANTIDOT LATTICE

The effective conductivity tensor is given by the formulas

$$\sigma_{xx}^{\mathrm{eff}} = \frac{\overline{\mathbf{j}\mathbf{E}}}{\overline{\mathbf{E}}^2}, \quad \sigma_{xy}^{\mathrm{eff}} = \frac{(\mathbf{e_B}[\overline{\mathbf{E}} \times \overline{\mathbf{j}}])}{\overline{\mathbf{E}}^2}. \quad (16)$$

The average current in a disordered system, containing insulating inclusions, in the absence of a magnetic field is characterized by the effective conductivity σ^{eff}:

$$\bar{\mathbf{j}} = \sigma^{\mathrm{eff}}\bar{\mathbf{E}} = f\sigma\bar{\mathbf{E}}.$$

The ratio f of the conductivities is determined by the structure of the medium. Specifically, if the inclusions are circular holes with radius a and density n, then f depends only on the dimensionless density $x = na^2$.

Let us consider antidots randomly and uniformly distributed in a plane. If the antidot density is low, $f(x) = 1 - 2x$. Near the percolation threshold $x \longrightarrow x_c$

$$f(x) = (x - x_c)^t$$

(where $t = 1.3$ [6]) irrespective of the geometry of the system [7].

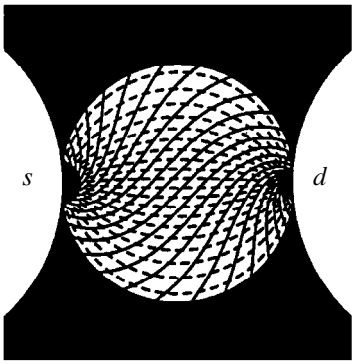

Fig. 3. Distribution of the current and field in a large circular quantum dot with two tunneling contacts. The Hall angle is $\pi/4$.

In the case of a quadratic periodic antidot lattice with period d, the quantity f in limiting cases is

$$f = 1 - 2\pi(a/d)^2,$$

if $a/d \ll 1$, and

$$f = \pi^{-1}\sqrt{(d-2a)/a},$$

if $d - 2a \ll d$.

To find the effective conductivity in a magnetic field, all integrals in equation (16) must be expressed in terms of a single quantity, for example, $\int \nabla_x \psi_0 dS$. Using equation (8) and

$$\sigma \int_{\text{Hall}} \nabla\psi_0 dS = \sigma^{\text{eff}} \int \nabla\psi_0 dS,$$

we obtain

$$\bar{\mathbf{j}} = \sqrt{\sigma_{xx}^2 + \sigma_{xy}^2}\,\overline{\nabla\psi_0}, \tag{17}$$

$$\overline{\nabla\phi} = \overline{\nabla_x\psi_0}(f^{-1}\cos\alpha, -\sin\alpha). \tag{18}$$

As a result

$$\sigma_{xx}^{\text{eff}} = f\frac{\sigma_{xx}(\sigma_{xx}^2 + \sigma_{xy}^2)}{\sigma_{xx}^2 + f^2\sigma_{xy}^2},$$
$$\sigma_{xy}^{\text{eff}} = f^2\frac{\sigma_{xy}(\sigma_{xx}^2 + \sigma_{xy}^2)}{\sigma_{xx}^2 + f^2\sigma_{xy}^2}. \tag{19}$$

Similar formulas were obtained in [8] by a different method for a standard macroscopic medium with inclusions and were not analyzed in the QHE limit.

The equations (19) make it possible to apply to the QHE problem all results of the theory of conductivity of a nonuniform medium. This theory can be used to find the effective conductivity from the local conductivity of a uniform sample and to solve the inverse problem of determining the local conductivity according to a measured effective parameter.

In the limit of an ideal Hall insulator $\sigma_{xx} \longrightarrow 0$, equation (19) gives $\sigma_{xx}^{\text{eff}} \longrightarrow 0$ and $\sigma_{xy}^2 \longrightarrow \sigma_{xy}$. The function f can be eliminated from equations (19), and this leads to a semicircle relation (2), which was found in the case of a mixture of different Hall phases:

$$(\sigma_{xx}^{\text{eff}})^2 + \left(\sigma_{xy}^{\text{eff}} - \frac{\sigma_{xy}^2 + \sigma_{xx}^2}{2\sigma_{xy}}\right)^2 = \left(\frac{\sigma_{xy}^2 + \sigma_{xx}^2}{2\sigma_{xy}}\right)^2. \tag{20}$$

In the limit $\sigma_{xx} \longrightarrow 0$ equation (20) does not contain σ_{xx}:

$$(\sigma_{xx}^{\text{eff}})^2 + \left(\sigma_{xy}^{\text{eff}} - \frac{\sigma_{xy}}{2}\right)^2 = \left(\frac{\sigma_{xy}}{2}\right)^2.$$

This relation permits σ_{xx}^{eff} to be finite for small σ_{xx}, i.e., finite global losses in a nondissipative medium. This apparent paradox is resolved by taking into account the fact that the semicircle relation arises when the indeterminate form is expanded in the limit $\sigma_{xx}/\sigma_{xy} = H \longrightarrow 0$ and $f \longrightarrow 0$. As the ratio H/f varies, the components of the effective conductivity tensor run through all points of a semicircle. For small H/f, a finite σ_{xx}^{eff} arises because of the small value of f. In other words, finite dissipation near the percolation threshold arises because of competition of dissipation in the Hall phase and in narrow necks.

In the limit $H \ll f$

$$\sigma_{xy}^{\text{eff}} = \sigma_{xy}, \quad \sigma_{xx}^{\text{eff}} = \sigma_{xx}/f.$$

If $H \gg f$,

$$\sigma_{xy}^{\text{eff}} = \frac{f^2\sigma_{xy}^3}{\sigma_{xx}^2}, \quad \sigma_{xx}^{\text{eff}} = \frac{\sigma_{xy}^2 f}{\sigma_{xx}}.$$

The paradoxical result is that after holes are cut out of the $2D$ plane, σ_{xx}^{eff} can increase (and the losses with a fixed electric field can decrease). The explanation is that for small σ_{xx} the current is approximately perpendicular to the electric field. The field in narrow necks between antidots, determining the entire current, is large compared with the average field. A finite σ_{xx} "shorts" the field in the necks, and all the more strongly, the narrower the necks are. Therefore, the smaller f, the larger σ_{xx}^{eff} is.

In the limit $H \longrightarrow 0$, $f \longrightarrow 0$, and $H \gg f$ the finiteness of σ_{xx} gives rise to losses concentrated in narrow necks, playing the role of "hot spots."

6. BREAKDOWN OF QHE IN AN ANTID OT LATTICE

Breakdown of QHE consists in the appearance of finite σ_{xx}^{eff} at the minima of σ_{xx} with a high current density. Measurements of breakdown of QHE in regular

and random antidot lattices [4] show that the threshold decreases compared with the sample with no pattern.

We shall discuss the breakdown of QHE in a disordered antidot lattice using the model of [4] as a basis. According to this model, if the current density in narrow necks between a pair of close antidots is higher than a critical value j_c^0, the electron gas in the necks becomes overheated and the Hall insulator near these necks becomes a conductor. Breakdown of QHE in the entire sample means that the regions of normal metal overlap (and the Hall phase breaks up).

The radius of the zone of influence of overheating is determined by a phenomenological electronic cooling length l_e. The scenarios of breakdown differ, depending on the ratio of l_e to the distance between the antidots and their size.

We shall discuss the nonlocal case $l_e \gg a$ first. We note that even an individual antidot decreases the critical current. According to equation (12), the maximum current density near an antidot is two times higher than the average current density. Therefore if the average current density is greater than $j_c^0/2$, the neighborhood of an antidot of radius l_e becomes a conductor. This value determines the critical current in the entire sample, if the resistive phase which appears becomes connected. The condition for connectedness of the resistive phase is $4\pi n l_e^2 = B_c = 4.4$ [6].

If $4\pi n l_e^2 < B_c$ and $j_c > j > j_c^0/2$, the resistive phases around isolated antidots are torn up. Current flows around a solitary normal region within a Hall phase, if the diagonal conductivity of the resistive phase is much higher than the diagonal conductivity of the Hall phase. As a result, the thermal overheating within resistive regions can maintain only a moderate increase in conductivity within the region up to a value of the order of the conductivity σ_{xx} of the Hall phase.

If the current density is sufficiently low, the neighborhood of isolated antidots remains in the Hall phase, while the necks between close pairs of antidots become resistive. Even though the size l_e of the resistive phase is greater than the antidot radius, insulation of the resistive region from the current limits overheating inside, if these regions do not overlap.

If the regions of the resistive phase around close pairs overlap, we can find a percolation threshold, estimating the current maintaining the resistive state.

The density n_c of critical pairs, determined by the same condition $4\pi n_c l_e^2 = B_c$, can be found from the inequality expressing the fact that the current $j(0)$ in a neck between antidots is greater than j_c^0. The value of $j(0)$ is $\sqrt{2a/\delta}\, j(\infty)|\cos\chi|$, where χ is the angle between the intercenter direction and the direction of the average current. Hence the critical pair density is

$$n_c = 8\pi(na)^2(j_c/j_c^0)^2.$$

As a result

$$j_c = j_c^0\sqrt{\frac{B_c}{2}\frac{1}{4\pi n a l_e}}. \tag{21}$$

The last formula is valid if the antidot density is sufficiently high, $n \gg B_c(16\pi a l_e)^{-1}$. If $n_c < n \ll B_c(16\pi a l_e)^{-1}$, the critical current determined by isolated antidots is $j_c^0/2$.

7. BREAKDOWN OF QHE IN AN ANTIDOT LATTICE. LOCAL MODEL

Here we shall discuss the lower breakdown threshold of QHE in a random antidot lattice, based on the assumptions that $na^2 \ll 1$ and the critical current produces self-maintaining local overheating in a narrow strip dividing the sample into two QHE conductors adjoining the current contacts.

First, we apply a strong current to a Hall strip of variable width with no antidots. In this narrow part the current density is maximum and the threshold for breakdown is lowest. Near the narrowest section, the strip passes into the resistive state.

At the lowest threshold the current density elsewhere is below threshold, and these locations remain in the Hall insulator phase, where the voltage drop vanishes. For this reason, the critical location in the sample is the narrowest part of the sample. The width of the resistive strip vanishes as the average current density approaches its threshold value. The local width of the resistive phase and the current density depend on the voltage between the Hall phases and therefore along the edge of the Hall phases.

We shall apply these considerations to a sample with antidots. Based on them, it is natural to expect that the lowest breakdown threshold should be determined by the shortest section of the Hall phase, dividing the sample into two disconnected regions adjoining the contacts (Fig. 4). A resistive phase appears along this section. To a first approximation, the current density is bounded by this resistive phase and does not depend on the coordinates.

For this reason, the problem of finding the threshold reduces to the mathematical problem of finding the length of the shortest section of a rectangle, containing circular openings separating the rectangle into disconnected regions. For $na^2 \ll 1$ the length in the zeroth approximation is equal to the width W of the rectangle. Corrections arising in this quantity as a result of the

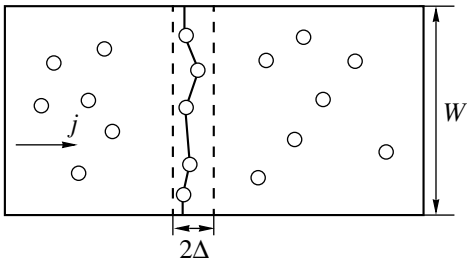

Fig. 4. Breakdown of QHE in an antidot lattice: local model. The resistive phase appears along the shortest cut connecting the antidots.

finiteness of na^2 can be found as done in [9]. We must minimize the length of the cut

$$\mathcal{L} = \sum_i (|\mathbf{r}_i - \mathbf{r}_{i+1}| - 2a)$$

between the antidots with positions $\mathbf{r}_i$. The sum includes a subset of antidots, minimizing the length $\mathcal{L}$.

At first glance the corrections are determined by the fraction of the straight line separating the sample and falling on an antidot, $\sim -Wna^2$. Actually, this line includes only antidots whose centers are located not more than $2a$ away from it. To minimize the length of the cut, it must be made as straight as possible, but the maximum number of antidots must be included in it. This can be done by including in it all antidots in a strip of width 2Δ, such that $a \ll 2\Delta \ll n^{-1/2}$. The length of the cut is determined by the number N of antidots which it intersects. The number N is of the order of the total number of antidots in the strip 2Δ: $N = 2nW\Delta$. The typical distance between the centers of the antidots within a strip is $l = \sqrt{(W/N)^2 + \Delta^2} \sim W/N + \Delta^2 N/(2W)$. For this reason, the minimum length of a cut passing along the Hall phase is

$$\mathcal{L} = W + \frac{\Delta^2 N^2}{2W} - 2aN = W + 2n^2 W\Delta^4 - 4anW\Delta. \quad (22)$$

The expression (22) must be minimized with respect to Δ. This gives $\Delta = (a/(2n))^{1/3}$ and $\mathcal{L} \sim W(1 - 3 \times 2^{-1/3}(na^2)^{2/3})$. This value of Δ is much less than $n^{-1/2}$, so that the cut is actually almost straight.

At the breakdown threshold the current density through a cut should correspond to the critical current density j_c^0 in an infinite medium. The average lowest critical current density, taking account of corrections for the ratio of the length of the cut to the width of the sample, is $j_c = j_c^0(1 - 3 \times 2^{-1/3}(na^2)^{2/3})$. We note that this correction is not analytic in na^2 and is much greater than the correction found on the basis of general considerations.

We have found the typical minimum length of a section. Of course, this result is valid only in the macroscopic limit, $N \gg 1$. In a small sample the fluctuations of this number are decisive, and the minimum length of the cut and therefore the breakdown threshold should fluctuate from sample to sample. This case can also be studied using the optimal-fluctuation method, but we shall confine ourselves to only the macroscopic limit.

8. CYCLOTRON RESONANCE

Let us consider the conductivity of a sample with antidots in a magnetic field at finite frequency. The electrostatic problem remains two-dimensional, if the frequency ω of the electric field is sufficiently low and in consequence the ohmic current is greater than the displacement current. Then the effective conductivity is determined by the expression (19), in which the local conductivity tensor is replaced by the high-frequency response. We shall consider a QHE medium with antidots near cyclotron resonance.

The local conductivity tensor has the form

$$\sigma_{xx} = \beta i\left(\frac{1}{\omega_c - \omega + i\delta} - \frac{1}{\omega_c + \omega - i\delta}\right),$$
$$\sigma_{xy} = \beta\left(\frac{1}{\omega_c - \omega + i\delta} + \frac{1}{\omega_c + \omega - i\delta}\right), \quad (23)$$

where $\beta = e^2 n/2m$, e and m are the electron charge and mass, n is the surface electron density, ω_c is the cyclotron frequency, and δ is the damping. Substitution of expressions (23) into equation (19) gives

$$\sigma_{xx}^{\text{eff}} = \frac{2\beta\omega_c f^2}{-(\omega - i\delta)^2 + f^2\omega_c^2},$$
$$\sigma_{xy}^{\text{eff}} = \frac{2i\beta\omega f}{-(\omega - i\delta)^2 + f^2\omega_c^2}. \quad (24)$$

Absorption of a linearly polarized electromagnetic wave is determined by σ_{xx}^{eff}. It follows from equations (24) that the resonance frequency and the oscillator strengths are renormalized by the factor f. The absorption of a circular wave with cyclotron-active polarization in the plane of the system that is associated with $\sigma_{xx}^{\text{eff}} - i\sigma_{xy}^{\text{eff}}$ has the same resonance frequency shift, and the oscillator strengths are renormalized by the factor f^2. If f is sufficiently small, the renormalization can be very large.

9. APPLICABILITY OF THE MACROSCOPIC APPROXIMATION

Our analysis is based on two important assumptions. The first one is that the relation (1) between the current density and the field is local. The other assumption is that the microscopic properties of the medium are uniform. Both assumptions require that the geomet-

ric dimensions of the system be large compared with the microscopic lengths.

The first assumption requires that the equilibration length be small compared to the geometric lengths. Within uniform regions of a Hall dielectric the equilibration length is, evidently, small, since the displacement of a center of the cyclotron orbit in a strong magnetic field is caused by scattering and is of the order of the cyclotron diameter, which for the lower Landau levels is equal to the magnetic length a_H.

The situation with the edge currents is different. Edge currents are sources of the Hall quantization in quite small regions [10, 11]. In large samples, the electrons injected into a sample from a contact along an edge state move along it until they are scattered into a volume or another edge state. (The phenomenology of the process is described in [9].) These transitions are limited by the necessity of tunneling to a distance much greater than the size of the wave function and therefore the equilibration length is exponentially large.

The equilibration length between the edge and volume states, evidently, depends on the splitting between the Fermi level and the Landau level. If it is small, then the transitions are elastic, otherwise exchange between edge and volume states limits the emission or absorption of phonons with activation energy determined by the energy splitting between delocalized volume states and the Fermi level. It is known that in a Hall insulator state the equilibration length can reach macroscopic values, right up to 100 μm [13].

There are two types of edge currents. Currents along the outer boundary of a sample carry a chemical potential from the source to the sink [10, 11]. A large equilibration length on these edge states can be compensated by large sample dimensions. If this is so, these edge currents can be neglected.

Other edge currents flow around the antidots and are not related directly with the contacts. They play no role in the two cases, when the size of and the distance between antidots are much greater than or, conversely, less than the equilibration length. Indeed, in the first case the edge currents are disconnected from one another and the contacts to the sample, and in the second case they usually mix with volume states.

Mixing also occurs between edge states of different antidots. It is exponentially small in the parameter equal to the ratio of the distance between the edges of antidots and the magnetic length a_H, and the only important fact is that this distance is sufficiently small.

Direct transitions between different edge states or between edge and volume states are determined by the overlapping of their functions. Let us consider elastic transitions. The transition time can be estimated as $(1/\omega_c)\exp(d_1^2/2a_H^2)$, where d_1 is the distance between the corresponding states. The quantity d_1 for transitions between edge states or between edge and volume states is determined primarily by the width of the depletion layer. For transitions between different edge states, d_1 is determined by the distance between the corresponding edges. More accurate estimates require analyzing the profile of the potential and the participation of impurities in the tunneling process [14].

The inner edge currents around antidots are disconnected and do not participate in the overall conductivity, if the distance δ between antidots is greater than the magnetic length a_H. If a sample lies close to the region of a Hall insulator, $\sigma_{xx} \ll \sigma_{xy}$, the criterion for the edge currents to be neglected is that the ratio of the parameter $\exp(-\delta^2/4a_H^2)$, characterizing the overlapping of the edge states around different antidots, to the parameter $H = \sigma_{xx}/\sigma_{xy}$ be small.

The profiles of the potential and electron density confirm also the validity of the assumption that the macroscopic properties of the medium are uniform. The width of the depletion layer around antidots is determined by the screening of the potential as a result of the redistribution of electrons. In the generally accepted approach [14], the screening is determined by the redistribution of electrons, complicated by the structure of the Landau levels. Since in the QHE the chemical potential is a step function of the electron density and, vice versa, the electron density is a step function of the chemical potential, the profile of the potential near the edges is also a step function of the coordinates, and screening occurs with alternation of the Hall conductor and insulator phases.

A simpler situation arises in the presence of a field electrode, screening the lateral potential. In this case the position of the chemical potential on the potential surface is fixed. If, as ordinarily happens, the distance between the two-dimensional layer and the surface is small, ~1000 Å, the width of the depletion layer is determined predominantly by this distance and not by the 2D electrons. This assumption seems to us to correspond better to experiment than the situation described in [14].

We shall distinguish two cases. In the first one, in the outer region of the electronic liquid between antidots, the chemical potential is fixed at a Landau level (Hall conductor). In the second case it lies between the Landau levels (Hall insulator).

In the first case the linear screening is determined by the surface and the redistribution of electrons in the two-dimensional layer, while in the second case it is determined only by the electrons. In both cases the typical spatial scale of the electron density is determined by the geometric factors, i.e., the distance D between the two-dimensional layer and the surface. The behavior of the potential is distinguished from an antidot along the 2D layer by the asymptotic behavior. At large distances from the edge of the capacitor, formed by the edge of an antidot, the electric potential approaches exponentially its value at infinity. The electron density is determined by the normal derivative of this potential

and therefore has the same behavior. In the second case the variation of the potential along the 2D layer is a smooth function of the ratio of the longitudinal coordinate ρ to D. This function can be estimated by treating an antidot as an additional charge Q placed at a distance D near a metallic surface: $\phi(\rho) = Q/\rho - Q/\sqrt{\rho^2 + 4D^2} \sim \phi_S aD^2/\rho^3$, where ϕ_S is the potential of the surface of the semiconductor and ϕ is the equilibrium potential. The corresponding change in the electron density (number of quasiparticles in Landau levels lying next to μ) is proportional to $\delta n = \exp(-(\omega_c/2 - |\mu - e\phi - n\omega_c|)/T)$.

Let us assume that the conductivity $\sigma_{xx} \sim \delta n$. According to this assumption, the change in σ_{xx} should be small. This determines the electrical size of an antidot.

In the first case the size of an antidot increases with D. In the second case the electric size a_e, if it exceeds the geometric size a, should satisfy the inequality $a_e \gg (\phi_S aD^2/T)^{1/3}$. Outside this region σ_{xx} can be assumed to be constant. If the conductivity is determined by hops near the chemical potential, the temperature T must be replaced by a characteristic energy θ, determining the dependence of the conductivity on the chemical potential that is associated with the density of states in the gap: $\sigma_{xx} \sim F(\mu/\theta)$.

The depletion layer around an antidot can consist of alternating Hall phases, which should complicate the problem. However, according to [2] the boundary conditions at the boundary of different Hall phases require that the normal component of the current vanish. If the inner region is closed, it can be taken to be insulating and our entire analysis remains in force. The only quantity requiring redefinition is the boundary of an antidot, which should be understood as the boundary of the last phase of the Hall insulator.

10. CONCLUSIONS

In conclusion we shall underscore the main results.

We have studied the quantum Hall effect in a system containing antidots. The conductivity tensor of a system with antidots in a magnetic field can be expressed in terms of the conductivity of a similar system without a magnetic field and the local conductivity in a magnetic field. We found that the presence of insulating inclusions does not influence the Hall quantization and the vanishing of the diagonal conductivity. A measure of the influence of antidots on the conductivity is the deviation of the local conductivity σ_{xx} from zero and the conductivity σ_{xy} from the quantum values. The finite effective conductivity σ_{xx}^{eff} and the deviation of σ_{xy}^{eff} from the steps in the limit $\sigma_{xx} \longrightarrow 0$ arise near the percolation threshold.

A local electric field in a system containing antidots in a magnetic field differs from the electric field in the absence of a magnetic field by a rotation by the Hall angle. The local current density with prescribed boundary conditions for the current on the outer boundaries does not depend on the magnetic field.

In this paper the distribution of the local field in the particular cases of one and two antidots was found. Knowing the local fields makes it possible to determine the breakdown threshold for QHE in an antidot lattice on the basis of the model of overheating of the electron gas.

The formulas for the effective conductivity, which are applicable to the case of high-frequency electric field, made it possible to find the cyclotron resonance frequency in a system with antidots that is different from the frequency for a uniform medium.

ACKNOWLEDGMENTS

This work was supported in part by the Russian Foundation for Fundamental Research (project 97-02-18397) and the National Program "Physics of Solid-State Nanostructures." One of us (É. M. B.) was also partially supported by INTAS grant no. 960 730.

REFERENCES

1. V. E. Arkhincheev and E. G. Batyev, Solid State Commun. **12**, 1059 (1989).
2. A. M. Dykhne and I. M. Ruzin, Phys. Rev. B **50**, 2369 (1994).
3. I. M. Ruzin and S. Feng, Phys. Rev. Lett. **74**, 154 (1995).
4. G. Nachtwei, G. Lutjering, D. Weiss, *et al.*, Phys. Rev. B **55**, 6731 (1997).
5. E. M. Baskin and M. V. Entin, Phys. Low-Dimens. Semicond. Struct. **1/2**, 17 (1997).
6. B. I. Shklovskii and A. L. Efros, *Electronic Properties of Doped Semiconductors* (Nauka, Moscow, 1979; Springer-Verlag, New York, 1984).
7. S. Feng, B. I. Halperin, and P. N. Sen, Phys. Rev. B **35**, 197 (1987).
8. B. Ya. Balagurov, Zh. Éksp. Teor. Fiz. **82**, 1333 (1982) [Sov. Phys. JETP **55**, 774 (1982)].
9. B. I. Shklovskii and B. Z. Spivak, J. Stat. Phys. **50**, 60 (1984).
10. É. M. Baskin and M. V. Éntin, Zh. Éksp. Teor. Fiz. **98**, 2150 (1990) [Sov. Phys. JETP **71**, 1208 (1990)].
11. M. Butiker, Semicond. Semimet. **35**, 191 (1992).
12. V. T. Dolgopolov, G. V. Kravchenko, and A. A. Shashkin, Solid State Commun. **78**, 999 (1991).
13. M. V. Budantsev, G. M. Gusev, Z. D. Kvon, and A. G. Pogosov, Pis'ma Zh. Éksp. Teor. Fiz. **60**, 834 (1994) [JETP Lett. **60**, 848 (1994)].
14. D. B. Chklovskii, B. I. Shklovskii, and L. I. Glazman, Phys. Rev. B **46**, 4026 (1992).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 4, 2000, pp. 655–661.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 4, 2000, pp. 753–760.

SOLIDS
Electronic Properties

Order Parameter of Layered Superconducting Structures

A. N. Oraevskiĭ
Lebedev Physics Institute, Russian Academy of Sciences, Moscow, 117924 Russia
e-mail: oraevsky@sci.lebedev.ru
Received June 3, 1998

Abstract—It is suggested that the order parameter and the critical temperature in an *S–N* structure be described on the basis of the microscopic Ψ theory, modeling the metal in the normal state by the Ginzburg–Landau equation with a negative coefficient *a*. It is shown that at a contact between layers the order parameter in the layer whose temperature is below the critical temperature induces an order parameter in the other layer whose temperature is higher than the critical temperature. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The physics of contact proximity of a superconductor and a normal metal has been under investigation for many years. Although investigations of this problem peaked in the 1960s [1–4], it continues to attract attention today [5, 6]. The effect of a contact between a normal metal and a superconductor on the critical temperature of the superconductor has been studied theoretically and experimentally [3, 4]. Relations describing the superconducting transition temperature of a composite structure containing a normal metal and a superconductor have been obtained on the basis of a microscopic theory. The computational results are in good agreement with experiment [4]. The overwhelming majority of works on the problem of contact proximity employ a microscopic approach. In the present work an attempt is made to describe the proximity effect and related effects on the basis of the Ψ theory of superconductivity which is based on the nonstationary Ginzburg–Landau (GL) equation [7, 8]

$$\frac{\partial}{\partial t}\Psi(\mathbf{r},t) = (a - b|\Psi|^2)\Psi(\mathbf{r},t) + D\left(\nabla - i\frac{2e}{\hbar c}\mathbf{A}\right)^2\Psi(\mathbf{r},t). \quad (1)$$

The standard notation, similar to that adopted in many works [7–14] based on equation (1), is used in this equation. A description on the basis of the microscopic theory of an effect such as the proximity effect requires a model that allows for the appearance of an order parameter in the normal metal. For this reason, in what follows we shall describe the normal metal by a GL equation with a negative coefficient *a*. The relation [7, 8]

$$\mathbf{n}\left(i\hbar\nabla\Psi + \frac{e}{c}\mathbf{A}\Psi\right) = 0, \quad (2)$$

where **n** is a vector normal to the surface of the superconducting sample, is used as a condition at the boundary with the vacuum. According to [2, 11], at the interface between two different materials (including between a superconductor and a normal metal)

$$\frac{\Psi_1}{N_1V_1} = \frac{\Psi_2}{N_2V_2}, \quad \frac{D_1}{V_1}\frac{d\Psi_1}{dx} = \frac{D_2}{V_2}\frac{d\Psi_2}{dx}, \quad (3)$$

where different indices are used to denote quantities at the interface on the side of the different layers, N_j is the density of states, V_j is the interaction constant for the current carriers, and D_j is the diffusion coefficient of pairs.

2. PERIODIC AND ASYMPTOTIC SOLUTIONS OF THE GL EQUATION

We shall analyze the nonstationary solutions of equation (1) in the simplest one-dimensional case with no magnetic field. We represent the order parameter in the form

$$\Psi(\mathbf{r}) = \left(\frac{\alpha}{b}\right)^{1/2}\psi(x)\exp[i\varphi(x)], \quad a = \alpha\theta, \quad \theta = \frac{T_0 - T}{T_0}, \quad (4)$$

where T_0 is the critical temperature of a bulk uniform superconductor and $\psi(x)$ and $\varphi(x)$ are real functions which did not depend on the coordinates *y* and *z*. According to equation (1), these functions satisfy the equations

$$\frac{d^2\psi}{d\xi^2} = -\psi(\theta - \psi^2) + \frac{G}{\psi^2}, \quad (5)$$

$$\frac{d\varphi}{d\xi} = \frac{G}{\psi^2}, \quad (6)$$

where $\xi = \sqrt{\alpha/D}\,x$ and *G* is an arbitrary constant of integration. In the absence of a superconducting current

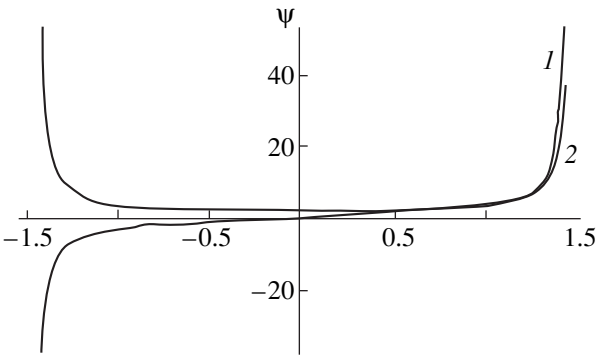

Fig. 1. Asymptotic solutions of equation (7). The order parameter in arbitrary units is plotted along the ordinate. The curve *1* (symmetric) corresponds to the values $\psi(0) = 1.5$ and $\psi'(0) = 0$ and possesses a vertical asymptote $\xi_A = 1.442$. The curve *2* (asymmetric) corresponds to the values $\psi(0) = 0$ and $\psi'(0) = 2.6$ and possesses a vertical asymptote $\xi_A = 1.468$.

$G = 0$, which we shall assume in what follows. Then equation (3) possesses, besides the obvious solutions $\psi = 0$ and $\psi^2 = \theta$, solutions which are determined by the equations

$$\frac{d\psi}{d\xi} = \pm\sqrt{C - \theta\psi^2 + \frac{\psi^4}{2}}, \qquad (7)$$

where C is a constant of integration. For $\theta > 0$ and $0 < C < \theta^2/2$ the solutions of equation (7) are periodic. Indeed, a continuous periodic function necessarily passes through alternating maxima and minima. It follows from equation (7) that

$$\psi_{ex}^2 = \theta \pm \sqrt{\theta^2 - 2C} \qquad (8)$$

at the extremal points. It is obvious that two real extrema (a maximum and a minimum) exist if $\theta > 0$ and $0 < C < \theta^2/2$.

For $C = \theta^2/2$ the order parameter is identically a constant ($\psi^2 = 0$). However, if $\theta < 0$ for any C, or $C < 0$ for any θ, or $C > \theta^2/2$, then the solutions are no longer periodic: They increase with ξ and approach a vertical asymptote (Fig. 1). The numerically computed positions of the asymptotes ξ_A for various values of the order parameter $\psi(0)$at the point chosen as the origin of the coordinates are presented in Table 1. For definiteness, it was assumed in the calculations that $\theta = -0.5$, and the derivative of the order parameter at the initial point was chosen to be 0. It is obvious that as $\psi(0)$ increases, the asymptote shifts toward the origin. Thus, the thinner the superconducting layer is, the larger the admissible value of the order parameter. A characteristic feature of the solutions with a vertical asymptote ("asymptotic" solutions) is that they exist even for negative values of the parameter θ. This suggests that the order parameter in the normal metal can be described by modeling the last GL equation with a negative coefficient a (or θ). In such a material without contact with a superconductor the order parameter is 0, so that the superconducting state does not exist. At the same time, this model potentially admits the existence of a superconducting state under favorable conditions.

In the absence of a field, the boundary condition (2) requires that the derivative vanish at the boundary with a vacuum. This condition determines the relation between the constant C and the value of the order parameter at the boundary. It is obvious that periodic solutions can satisfy the conditions (2) at the boundary with a vacuum on both sides of a finite-size sample. However, solutions with a vertical asymptote cannot satisfy the conditions at the boundary with a vacuum on both sides of the sample. Therefore, in a uniform material adjoining a vacuum they must be dropped as nonphysical solutions. But if the sample consists of two layers of different superconducting materials, then the asymptotic solution on one edge of the sample can made to conform directly with the vacuum, while on the other edge it can be made to conform with the vacuum via a periodic solution for a layer of a different material. The idea of such matching is demonstrated in Fig. 2.

In what follows we shall analyze a two-layer element with layer thicknesses l_1 and l_2. The material in the layers will be superconductors (metals) with different parameters, specifically, different critical temperatures. We shall denote the parameters of the materials by the a_μ, b_μ, D_μ, and T_μ. Then equation (1) for the stationary case with no magnetic field can be written for both materials in the general form

$$D_\mu\frac{d^2\psi}{dx^2} = -\psi(a_\mu - b_\mu\psi^2), \quad \mu = 1, 2. \qquad (9)$$

The asymptotic and periodic solutions of the nonlinear equation (9) can be expressed in terms of elementary functions. We introduce the following notation for them. We denote a periodic function that is symmetric with respect to the origin of the coordinates by $P_s(x; a, b, C)$ and an asymptotic symmetric function by $A_s(x; a, b, C)$. We shall place the origin of the coordinates on the left-hand boundary of the left-hand layer. Then for the left-hand layer

$$\Psi_1(x) = P_s(x; a_1, b_1, C_1), \qquad (10)$$

Table 1

$\psi(0)$	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
ξ_A	1.657	1.170	0.898	0.726	0.609	0.524	0.460	0.409	0.369

and for the right-hand layer

$$\Psi_2(x) = A_s(x - l_1 - l_2;\, a_2, b_2, C_2). \tag{11}$$

The solutions (10) and (11) satisfy the condition (2) at the boundary with a vacuum. The conditions (3) at the boundary of the layers reduce to the relations

$$\begin{aligned} \frac{D_1}{V_1}\frac{d}{dx}[P_s(l_1;\, a_1, b, C_1)] &= \frac{D_2}{V_2}\frac{d}{dx}[A_s(l_2;\, a_2, b_2, C_2)], \\ \frac{1}{N_1 V_1}P_s(l_1;\, a_1, b, C_1) &= \frac{1}{N_2 V_2}A_s(l_2;\, a_2, b_2, C_2), \end{aligned} \tag{12}$$

which determine the values of C_1 and C_2 in terms of the parameters of the materials and the thicknesses of the layers. This gives the solution for a two-layer structure.

Together with the solution described by equations (10) and (11), there also exists a zero solution. Thus, competition between two solutions occurs in the layer structure under study. The stability of the solutions determines which one is realized.

3. CRITICAL TEMPERATURE AND STABILITY OF THE SOLUTIONS

It is simplest to determine the stability of the zero solution. For this, it is necessary to use the nonstationary equation (1). In the absence of a magnetic field, we shall seek this solution in the form

$$\Psi_{1,2}(t, x) = e^{\lambda t}\delta\psi_{1,2}(x), \tag{13}$$

where λ is a Lyapunov exponent determining the stability of the regime. Stability analysis assumes that the quantity $\delta\psi_{1,2}(x)$ is small. We shall use the linearized equation (1) normalized to the parameters of the left-hand layer:

$$D_\mu\frac{d^2\psi}{dx^2} = (\lambda - a_\mu)\psi. \tag{14}$$

The solution of equation (14) for the left-hand layer ($\theta_1 > 0$), satisfying the condition on the left-hand boundary with the vacuum, has the form

$$\delta\psi_1(x) = A\cos\left(\sqrt{\frac{a_1 - \lambda}{D_1}}x\right). \tag{15}$$

For $\lambda = 0$ it is a linear approximation of a periodic symmetric solution. Thus solution of equation (14) for $\delta\psi_2$, which is a linear approximation of a symmetric asymptotic solution and satisfies the condition at the boundary with a vacuum on the right-hand side, is given by the relation

$$\delta\psi_2(x) = B\cosh\left[\sqrt{\frac{\lambda - a_2}{D_2}}(x - l_1 - l_2)\right]. \tag{16}$$

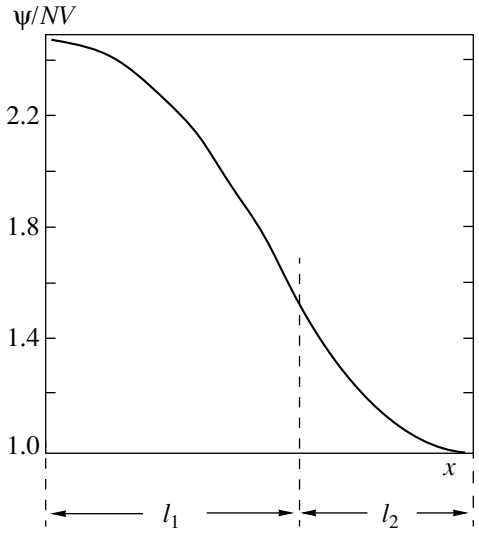

Fig. 2. Qualitative form of the curve of the order parameter in a two-layer superconducting structure. Here l_1 is the thickness of the layer with the higher critical temperature (left-hand layer); l_2 is thickness of the layer with the lower critical temperature (right-hand layer).

Then, according to equations (3),

$$\begin{aligned} &D_1 N_1\sqrt{\frac{a_1 - \lambda}{D_1}}\tan\left(\sqrt{\frac{a_1 - \lambda}{D_1}}l_1\right) \\ &= D_2 N_2\sqrt{\frac{\lambda - a_2}{D_2}}\tanh\left(\left(\sqrt{\frac{\lambda - a_2}{D_2}}l_2\right).\right) \end{aligned} \tag{17}$$

Obvious, $\lambda > 0$ means that the zero solution is unstable. Therefore equation (17) with $\lambda = 0$,

$$\begin{aligned} &D_1 N_1\sqrt{\frac{a_1}{D_1}}\tan\left(\sqrt{\frac{a_1}{D_1}}l_1\right) \\ &= D_2 N_2\sqrt{-\frac{a_2}{D_2}}\tanh\left(\left(\sqrt{-\frac{a_2}{D_2}}l_2\right),\right) \end{aligned} \tag{18}$$

determines the ratio of the layer parameters that corresponds to the boundary of instability of the zero solution. Depending on circumstances, equation (18) can determine the following: (a) the general critical temperature of a two-layer sample; (b) the sample thicknesses for which the block as a whole possesses a fixed critical temperature; and, (c) the critical temperature of the right-hand layer with a fixed temperature of the left-hand layer, if the layers are thermally insulated from one another.

We shall now present numerical examples illustrating these situations.

First example. The temperature of both samples is the same. For fixed parameters of the materials the condition (18) determines the general critical temperature of the sandwich. The critical temperature of a system consisting of a normal metal and a superconductor has

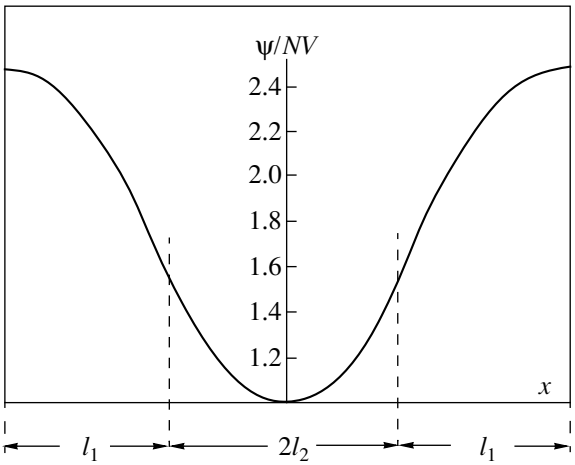

Fig. 3. Periodic structure, symmetric regime. The order parameter in the layer with the lower critical temperature is symmetric with respect to the center of the layer.

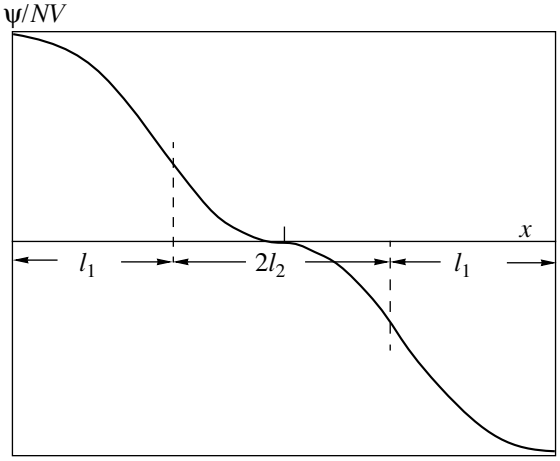

Fig. 4. Periodic structure, antisymmetric regime. The order parameter in the layer with the lower critical temperature is antisymmetric with respect to the center of the layer.

been studied previously both experimentally and theoretically. The condition (18) is the same, to within the notation, as the relation obtained in [3] for the general superconducting transition temperature of a layered structure. In [3] excellent agreement between the calculations and experiment was demonstrated for a sandwich consisting of lead and aluminum.

Second example. The layers are thermally insulated and fabricated from the same material: $T_1 = T_2 = 100$ K. The temperature of the left-hand layer is $T = 50$ K. We set $\sqrt{a_1/D_1}\,l_1 = \sqrt{a_2/D_2}\,l_2 = 1.5$. Then the thermally insulated right-hand layer will remain in the superconducting state right up to temperature $T_c \approx 390$ K.

It is clear that the parameters being determined are not absolutely arbitrary. Thus, the general critical temperature of a sample cannot exceed the critical temperature T_1 ($T_1 > T_2$). When the layers are thermally insulated, the fixed temperature of the left-hand layer must be lower than the critical temperature of the layer. Then, however, the critical temperature of the right-hand layer can be appreciably higher than that of the left-hand layer. In principle, it can exceed room temperature, as the numerical example presented above shows. In the example presented, the temperature of the layers was assumed to be different purely speculatively. In reality, the thermal insulation of the layers will require the insertion of an additional insulating layer. This additional layer can change the results obtained. For this reason, a model of a three-layer element is discussed in Section 4.

A multilayer structure can be periodic. It can be constructed from two-layer elements joined at the surfaces of layers consisting of identical materials. Two cases are possible: The asymptotic functions describing the order parameter in the layers with a lower critical temperature can be symmetric or antisymmetric (Figs. 3 and 4). The order parameter in the layer with the higher critical temperature must always be symmetric. The antisymmetric asymptotic solutions change sign on passing through the center of the corresponding layer. This actually means that the phase of the order parameter changes by π. Since the phase jump occurs at a point where the modulus of the order parameter is 0, this does not lead to an abrupt change in the superconducting current. Therefore competition between two possible regimes arises in a periodic structure.

The analysis of the instability of the zero value of the order parameter with respect to fluctuations with an antisymmetric structure of the asymptotic solution is similar to the analysis for a symmetric structure. It is only necessary to replace the hyperbolic cosine in the solution for the right-hand layer by a hyperbolic sine. As a result, we arrive at the following equation for the Lyapunov exponent λ:

$$D_1 N_1 \sqrt{\frac{a_1 - \lambda}{D_1}} \tan\left(\sqrt{\frac{a_1 - \lambda}{D_1}} l_1\right) = D_2 N_2 \sqrt{\frac{\lambda - a_2}{D_2}} \coth\left(\sqrt{\frac{\lambda - a_2}{D_2}} l_2\right). \tag{19}$$

Introducing the variable $u = l_2\sqrt{(\lambda + |a_2|)/D_2}$ and performing simple transformations, it can be shown that equations (17) and (19), which have the same left-hand sides, can be reduced to an equation of the form

$$a\sqrt{b - cu^2}\tan\left(\sqrt{b - cu^2}\right) = \begin{cases} u\tanh u \\ u\coth u, \end{cases} \tag{20}$$

where a, b, and c are constants. Graphical analysis of this equation shows that the largest Lyapunov exponents for antisymmetric regimes is always less than the analogous exponents for the symmetric solutions. Therefore the symmetric solutions always win in the

mutual competition. Figure 5 illustrates this assertion for the particular case $a = b = c = 1$.

4. THERMALLY INSULATED LAYERS

To prepare a sandwich with different layer temperatures it is necessary to introduce an intermediate insulating layer. To investigate such structures the boundary conditions employed for describing Josephson junctions must be applied [8]:

$$\frac{d\psi_1}{dx} = \nu_2\psi_2, \quad \frac{d\psi_2}{dx} = \nu_1\psi_1. \tag{21}$$

The index 1 denotes the order parameter in the left-hand superconducting layer at the boundary with the insulator, the index 2 denotes the order parameter in the right-hand superconducting layer at the boundary with the insulator, and $\nu_{1,2}$ are real constants, whose values depend on the properties of the materials and the thickness of the insulating layer. Using the corresponding expressions (15) and (16) for the fluctuations of the order parameter and the relations between them (21), due to the presence of the insulator, we obtain the following equation at the boundary of the transition from the normal to the superconducting regime:

$$\frac{1}{\nu_1\nu_2}\sqrt{-\frac{a_1a_2}{D_1D_2}}\tan\left(\sqrt{\frac{a_1}{D_1}}l_1\right) = \coth\left(\sqrt{-\frac{a_2}{D_2}}l_2\right). \tag{22}$$

The right-hand side of this expression is not less than 1. The left-hand side can assume any value as a function of the thickness of the layer. Thus, the conclusion concerning the possibility of initiating a superconducting state in the "warm" layer at a temperature above the critical temperature remains qualitatively in force. True, the smaller the quantity $\nu_1\nu_2$ characterizing the degree of contact between the layers, the more difficult it is to find the required thickness l_1 of the layer. Ultimately, this could turn out to be impossible, if $\nu_1\nu_2$ is so small that increasing or decreasing the sample thickness by one atomic layer will destroy the equality following from equation (22).

5. CYLINDRICAL AND SPHERICAL STRUCTURES

Layer structures can possess cylindrical or spherical geometry: a cylinder (ball) consisting of one material is inserted into a hollow cylinder (sphere) consisting of a different material. Determining the parameters of superconducting structure of this kind is essentially similar to doing so for a planar layers, except that the second derivative in equation (9) must be replaced by a Laplacian in a cylindrical or spherical coordinate system. Thus, we have for the cylindrical isotropic case

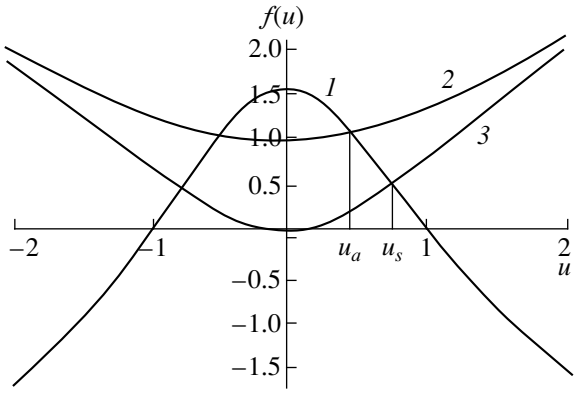

Fig. 5. Example of the graphical solution of equation (20): (*1*) $f(u) = \sqrt{1-u^2}\tan(\sqrt{1-u^2})$; (*2*) $f(u) = u\coth u$; (*3*) $f(u) = u\tanh u$.

$$\frac{d^2\psi}{d\xi^2} + \frac{1}{\xi}\frac{d\psi}{d\xi} + (\theta - \psi^2)\psi = 0. \tag{23}$$

Similarly, for the spherical isotropic case we arrive at the equation

$$\frac{d^2\psi}{d\xi^2} + \frac{2}{\xi}\frac{d\psi}{d\xi} + (\theta - \psi^2)\psi = 0. \tag{24}$$

In equations (23) and (24) ξ must be treated as a normalized radial coordinate: $\xi = \sqrt{a/D}r$.

The positions of the asymptotes for the cylindrical and spherical cases are presented in Table 2. Comparing with the corresponding values for a planar layer shows that for spherical and cylindrical geometries the asymptote lies farther from the origin of the coordinates than in the planar case. Hence, for a radius equal to the linear size of the planar layer the admissible values of the order parameter in spherical and cylindrical geometries are larger than in the planar case.

To determine the boundaries of instability of the zero solution it is necessary to investigate the linearized equation. In the cylindrical geometry it has the form

$$\frac{d^2\delta\psi}{dr^2} + \frac{1}{r}\frac{d\delta\psi}{dr} + \frac{a_\mu}{D_\mu}\delta\psi = 0. \tag{25}$$

Table 2

$\psi(0)$	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
$\xi_{A,\,cyl}$	2.033	1.434	1.100	0.889	0.746	0.642	0.563	0.501	0.452
$\xi_{A,\,sp}$	2.307	1.627	1.248	1.009	0.846	0.728	0.639	0.569	0.513

As is well-known [15, 16], the solutions of equation (25) are zero-order Bessel functions of the first and second kinds: $J_0(\sqrt{a_{1,2}/D_{1,2}}\,r)$ and $Y_0(\sqrt{a_{1,2}/D_{1,2}}\,r)$.

Let the material with the lower critical temperature be located inside a cylinder. Then the linear approximation for the asymptotic solution is a Bessel function of the first kind with imaginary argument $J_0(i\sqrt{|a_2/D_2|}\,r)$, since $a_2 < 0$. A Bessel function of the second kind is unsuitable as a solution, since it possesses a singularity and zero. The linear combination

$$L_0\left(\sqrt{\frac{a_1}{D_1}}r;\,\eta\right) = J_0\left(\sqrt{\frac{a_1}{D_1}}r\right) + \eta Y_0\left(\sqrt{\frac{a_1}{D_1}}r\right), \quad (26)$$

where η is a constant, must be taken as the solution for the outer cylindrical layer. The value of η is found from the condition at the boundary with a vacuum:

$$\begin{aligned}&\frac{d}{dr}L_0\left(\sqrt{\frac{a_1}{D_1}}r;\,\eta\right) \equiv L_1\left(\sqrt{\frac{a_1}{D_1}}r;\,\eta\right)\\ &\equiv \sqrt{\frac{a_1}{D_2}}J_1\left(\sqrt{\frac{a_1}{D_1}}r\right) + \eta\sqrt{\frac{a_1}{D_1}}Y_1\left(\sqrt{\frac{a_1}{D_1}}r\right) = 0,\end{aligned} \quad (27)$$

or

$$\eta = -J_1\left(\sqrt{\frac{a_1}{D_1}}r\right)\Big/ Y_1\left(\sqrt{\frac{a_1}{D_1}}r\right). \quad (28)$$

Then the condition for matching of the solutions at the point r_1 becomes

$$\begin{aligned}&D_2N_2\frac{i\sqrt{|a_2/D_2|}\,J_1(i\sqrt{|a_2/D_2|}\,r_1)}{J_0(i\sqrt{|a_2/D_2|}\,r_1)}\\ &= D_1N_1\frac{(i\sqrt{|a_1/D_1|}\,r_1)L_1(\sqrt{|a_1/D_1|}\,r_1;\,\eta)}{L_0(\sqrt{|a_1/D_1|}\,r_1;\,\eta)}\end{aligned} \quad (29)$$

(r_1 and r_2 are the inner and outer radii of the cylindrical sample). The relation (29) is equivalent to equation (18) for a planar structure. The same consequences follow qualitatively from it has from equation (18). Carbon nanotubes can be tried for producing superconducting layered elements with cylindrical and spherical geometries.

The linearized equation for a spherical geometry is

$$\frac{d^2\delta\psi}{dr^2} + \frac{2}{r}\frac{d\delta\psi}{dr} + \frac{a_\mu}{D_\mu}\delta\psi = 0. \quad (30)$$

The solution of equation (30) can be expressed in terms of a Bessel function of order 1/2: $(1/\sqrt{r})J_{\pm1/2}(\sqrt{a_{1,2}/D_{1,2}}\,r)$. In turn, $J_{-1/2}(z) = \sqrt{2/\pi z}\cos z$, $J_{1/2}(z) = \sqrt{2/\pi z}\sin z$ [15]. Now the matching equation at the boundary of the spherical layers, similar to equations (18) and (29), can be written down. This equation is too complicated to present here.

6. DISCUSSION

The effects discussed above do not depend critically on the form of the pantry conditions at the point of contact. The boundary conditions are only required to match the asymptotic solution through the superconducting layer with the condition at the boundary with a vacuum. This general assertion is supported by the fact that the effect exists with the boundary conditions (18) at the contact point as well as with the condition (21).

The fluctuations of the order parameter can play a large role in the region where the order parameter vanishes. For this reason, strictly speaking, the GL equations extended to the fluctuation region [17] should be used to analyze the stability of the zero solution. The generalized GL equations in the linear approximation differ from equation (1) by the fact that the coefficient θ is replaced by the coefficient $\theta|\theta|^{1/3}$. It can be shown that this does not qualitatively change the results presented above. Thus, in a layered superconducting structure a kind of initiation of an order parameter in the layer whose temperature is higher than the critical temperature of the corresponding material can occur. This is caused by the order parameter in the neighboring layer whose temperature is lower than the critical temperature.

The superconducting regime in a system of thermally insulated layers is interesting. In a three-layer element, where two superconductors with good thermal conductivity are separated by an insulator which conducts heat poorly, a regime with a different temperature should be established if one of the superconducting layers is in contact with a refrigerant while the other is in contact with the surrounding environment at a higher temperature. The insulator must bear the sharp temperature drop between the superconducting layers. The thermal resistance of the insulator can be increased by increasing its thickness, and good tunneling transmittance can be achieved by selecting a material with a small barrier.

It is obvious that the characteristic features of two- and multilayer structures are not exhausted by what we have set above. Certainly, nonstationary regimes in layered structures are of interest. To study them it is necessary to use a system of equations that takes into account not only the dynamics of the order parameter of the superconducting Bose condensate but also the associated dynamics of free quasiparticles [18].

ACKNOWLEDGMENTS

I am grateful to É. M. Belenov (now deceased), with whom I repeatedly discussed the problem of superconductivity in layered structures. I thank A.F. Andreev, L.V. Keldysh, D.A. Kirzhnits (deceased), A.I. Golovashkin, V.F. Elesin, Yu.V. Kopaev, and G.F. Zharkov for discussions concerning the physics of the superconducting state.

REFERENCES

1. E. A. Lynton, *Superconductivity*, 4th ed. (Chapman and Hall, London, 1971, 4th ed.; Mir, Moscow, 1974).
2. P. G. De Gennes, Rev. Mod. Phys. **36**, 225 (1964).
3. N. R. Werthamer, Phys. Rev. **132**, 2440 (1963).
4. J. J. Hauser and H. C. Theuerer, Phys. Lett. **14**, 270 (1965).
5. I. A. Devyatov and M. Yu. Kupriyanov, Zh. Éksp. Teor. Fiz. **104**, 3897 (1993) [JETP **77**, 874 (1993)].
6. I. A. Devyatov and M. Yu. Kupriyanov, Zh. Éksp. Teor. Fiz. **112**, 342 (1997) [JETP **85**, 189 (1997)].
7. V. L. Ginzburg and L. D. Landau, Zh. Éksp. Teor. Fiz. **20**, 1064 (1950).
8. E. M. Lifshitz and L. P. Pitaevskiĭ, *Course of Theoretical Physics. Vol. 5. Statistical Physics* (Nauka, Moscow, 1978; Pergamon, New York, 1980), Part 2.
9. L. P. Gor'kov and G. M. Éliashberg, Zh. Éksp. Teor. Fiz. **55**, 2431 (1968) [Sov. Phys. JETP **28**, 1291 (1969)].
10. A. A. Abrikosov, Zh. Éksp. Teor. Fiz. **32**, 1442 (1957).
11. P. De Gennes, *Superconductivity of Metal and Alloys* (Benjamin, New York, 1966; Mir, Moscow, 1968).
12. M. Tinkham, *Introduction to Superconductivity* (McGraw-Hill, New York, 1975; Atomizdat, Moscow, 1980).
13. V. F. Elesin and Yu. V. Kopaev, Usp. Fiz. Nauk **133**, 259 (1981) [Sov. Phys. Usp. **24**, 116 (1981)].
14. A. M. Gulyan and G. F. Zharkov, *Superconductors in External Fields* (Nauka, Moscow, 1990).
15. I. S. Gradshteĭn and I. M. Ryzhik, *Table of Integrals, Series, and Products* (Nauka, Moscow, 1971, 5th ed.; Academic, New York, 1980).
16. B. G. Korentsov, *Introdution to the Theory of Bessel Functions* (Nauka, Moscow, 1971).
17. L. N. Bulaevskiĭ, V. L. Ginzburg, and A. A. Sobyanin, Zh. Éksp. Teor. Fiz. **94** (7), 355 (1988) [Sov. Phys. JETP **67**, 1441 (1988)].
18. A. N. Oraevskiĭ, Zh. Éksp. Teor. Fiz. **103**, 981 (1993) [JETP **76**, 480 (1993)].

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 4, 2000, pp. 662–670.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 4, 2000, pp. 761–770.

SOLIDS
Electronic Properties

Electron Localization in a Nondegenerate Semiconductor with a Random Potential due to Charged Impurities

N. G. Zhdanova, M. S. Kagan*, and E. G. Landsberg

Institute of Radio Engineering and Electronics, Russian Academy of Sciences, Moscow, 103907 Russia
** e-mail: kagan@landau.ac.ru, kagan@mail.splire.ru*

Received August 11, 1999

Abstract—The saturation of the electron mobility, as determined according to the magnetoresistance, was observed in a semiconductor with a large-scale potential due to charged impurities. It was shown that the saturation is due to the existence of a quantum mobility threshold. A negative magnetoresistance of nondegenerate electrons, which is due to the suppression of quantum interferences corrections to the conductivity by the magnetic field, was found. The magnitude of these effects near the mobility threshold was explained by the absence of short, closed, electronic trajectories in the large-scale potential. A relation was established between the amplitude of the random potential and the saturated values of the mobility and the quantum corrections to the conductivity. *© 2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

A disordered spatial distribution of scattering centers (for example, charged impurities) results in localization of electrons with energy below a certain value ε_c, called the mobility threshold. Localization phenomenon have been studied in detail in systems with a degenerate electron gas. The character of the conductivity of a degenerate electron gas is determined by the position of the Fermi level ε_F with respect to the mobility threshold. For $\varepsilon_F > \varepsilon_c$ the mobility is of a metallic character, while for $\varepsilon_F < \varepsilon_c$ the system behaves like an insulator, and its mobility vanishes at the temperature $T = 0$. The metal–insulator transition, which occurs when the Fermi level crosses the mobility threshold, is known as Anderson's transition. Localization effects can also be observed in nondegenerate systems, such as an electron gas above the surface of solid hydrogen [1] or semiconductors with Fermi energy in the band gap [2, 3]. A system with a nondegenerate electron gas lies on the insulator side of Anderson's transition. However, a finite conductivity exists at zero temperature T as a result of the thermal "tail" of the Boltzmann electron energy distribution function. Since in a nondegenerate electron gas electrons with energy of the order of kT, for low T when $kT < \varepsilon_c$, make the main contribution to transport phenomenon, this makes it possible to study the behavior of electrons in the critical energy range near the mobility threshold. In a well-known sense, a nondegenerate electron gas is more "pure" for studying localization phenomenon than a degenerate gas, since the interelectron interaction can be neglected because the electron density is so low. In addition, for a nondegenerate electron gas the amplitude of the random potential due to the charged impurities and the mobility threshold should not depend on the free-electron density, which can be varied over a wide range, for example, by means of optical excitation of electrons from impurities.

The existence of a mobility threshold should determine the character of transport phenomenon in a nondegenerate electron gas when the wavelength λ of electrons with average thermal energy is of the order of the electron mean free path l (the Ioffe–Regel' condition). This condition for strong localization can be easily satisfied in a compensated semiconductor with deep impurities at not too-low temperatures and moderate doping levels. In the present work, localization effects were studied in compensated Ge doped with deep multiply charged impurities. A mobility threshold should exist in this material because of the strong scattering by the charged impurities. On the other hand, disorder in the spatial arrangement of the same charged impurities should result in the appearance of a random Coulomb potential, and it should cause the conduction to exhibit percolation behavior. This raises the question of the relation between the quantum mobility threshold and the amplitude of the random potential.

The next question concerns the magnitude of the interference corrections to the conductivity. It has been shown in [4] that for a three-dimensional electron gas the quantum corrections to the conductivity as result of the interference of electron waves on closed trajectories in the presence of multiple scattering are of the order of λ^2/l^2 and are small to the extent that the parameter λ/l is small. Near the energy ε_c the interference corrections to the conductivity should increase, and it can be expected that for $\lambda \approx l$ (at $T = 0$) these corrections should be of the order of 1. On the other hand, the interference corrections to the conductivity should be strongly suppressed in systems with a random potential [5].

1063-7761/00/9004-0662$20.00

2. EXPERIMENT

Copper-doped Ge crystals (*n*-Ge), partially compensated with shallow donors (antimony) so that only doubly and triply negatively charged copper ions were present at equilibrium, were investigated. The antimony density varied from 1.5×10^{15} to 1.2×10^{16} cm^{-3}. At thermal equilibrium the Fermi level lies near the upper level of copper (0.26 eV from the conduction band). The conductivity at low temperatures was obtained by optical excitation of electrons from the copper levels. It has been shown in [6] that at low temperatures, when the compensating shallow donors are frozen out, their filling is controlled by direct interimpurity recombination shallow copper–deep acceptor and depends on the intensity I of the optical generation from deep levels of copper. As I varied, the degree of filling of the donors could vary from 10^{-3} up to 0.2. Thus, virtually all impurity centers are charged (Cu^{3-}, Cu^{2-}, and Sb^{+} ions).

A chaotic spatial arrangement of charged impurities leads to the appearance of a random Coulomb potential with amplitude γ. The values of γ can be determined by several methods. In the first place, it can be determined according to the classical broadening of the spectrum of impurities states. Specifically, the measured [6] donor density of states has been found to depend on the energy as $\exp[(\varepsilon_{FD} - \varepsilon)/\gamma]$, where ε_{FD} is the Fermi quasilevel of the donors. In the second place, it can be determined from the dependences of the interimpurity recombination coefficient β on I and T. It has been shown in [6] that the broadening of the energy spectrum of the impurities by the random potential leads to a dependence $\beta \propto I\exp(\gamma/kT)$. The experimental values of γ are presented in the table. The table also presents the values of γ calculated using the formula $\gamma = e^2 N_t^{2/3}/\kappa N_s^{1/3}$ [7], where $N_t = \sum Z N_Z$ is the total density of charged impurities taking account of their charge Z, N_s is the density of the screening charges, which in our case is equal to the total density of filled (neutral) donors and triply charged copper ions, e is the elementary charge, and k is the permittivity. It is evident that the computed values of γ agree well with the experimental values.

The strong impurity scattering in our samples due to the multiply charged impurities makes it possible to satisfy the condition $\tau_\varphi \gg \tau$ (τ is the average mean-flight time, τ_φ is the phase interruption time of the electronic waves due to inelastic scattering), which is necessary for observing weak-localization effects. A negative magnetoresistance was observed in our experimental samples [2]. The magnetoresistance was reconstructed from the measured dependence of the ac $\tilde{i} = \tilde{H}\, di/dH$, caused by the modulation of the magnetic field, $H + \tilde{H}$ and $\tilde{H} \ll H$ on the constant magnetic field H, since $\tilde{i} \propto d\sigma/dH$ (σ is the conductivity). Figure 1 shows the relative magnetoresistance $\Delta\sigma/\sigma$ versus the magnetic field for samples with different impurity concentrations (different values of γ) and close values of the ratio $kT/\gamma \approx 0.6$. It is evident that there exists a range of weak magnetic fields where the magnetoresistance is negative. In stronger fields it becomes positive and increases as H^2. As will be shown below, the negative magnetoresistance arising in weak fields, when the classical magnetoresistance is small, is due to the suppression of quantum interference corrections to the conductivity in a magnetic field [4, 5]. The change in the sign of the magnetoresistance with increasing H is

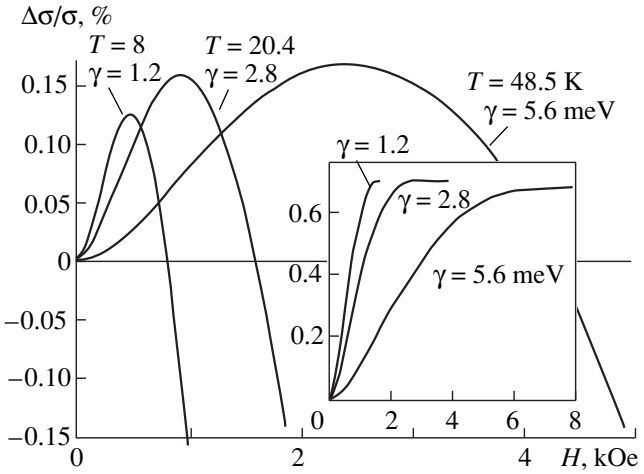

Fig. 1. Magnetic-field dependence of the relative magnetoresistance for samples with different doping levels and close values of the ratio kT/γ. Inset: negative magnetoresistance.

Amplitude γ of the random potential and mobility threshold ε_c (meV)

Sample	$\gamma = \dfrac{e^2 N_t^{2/3}}{k N_s^{1/3}}$ [7]	γ from the interimpurity recombination coefficient [6]	γ from the donor density of states [6]	kT_c	$\dfrac{\hbar}{2\tau_c}$	ε_c from comparison of $\Delta\sigma/\sigma(H)$ with Eq. (4)
1	1.2	–	–	1.1	1.2	–
2	2.8	2.7	2.8	2.3	2.8	3.5
3	5.6	5.5	5.7	5.6	6.0	9.0

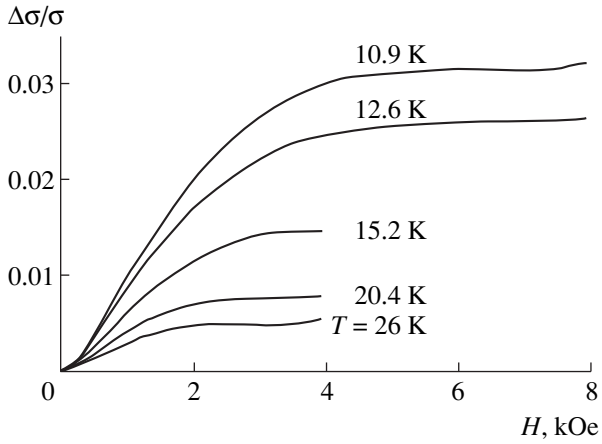

Fig. 2. Magnetic-field dependence of the negative magnetoresistance at different temperatures for a sample with $\gamma = 2.8$ meV.

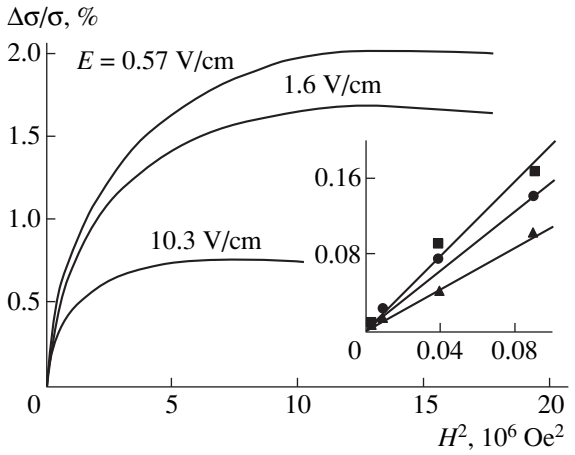

Fig. 3. Magnetic-field dependence of the negative magnetoresistance with different electric fields. Inset: initial sections of the curves.

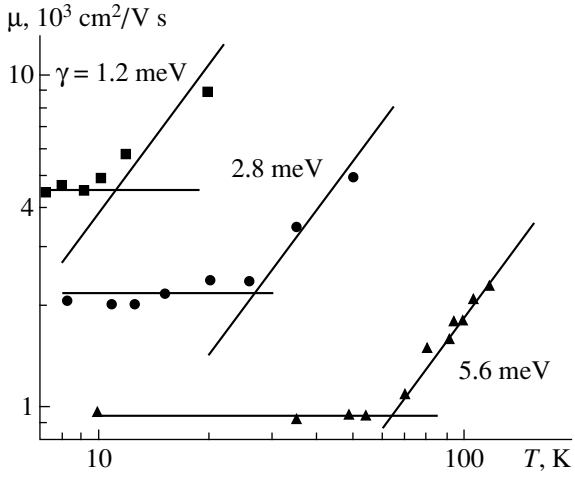

Fig. 4. Temperature dependence of the electron mobility for samples with different γ.

due to the saturation of the negative magnetoresistance and a transition to the classical dependence

$$\frac{\Delta\sigma}{\sigma} \propto -\left(\frac{\mu H}{c}\right)^2$$

(the condition $\mu H/c \ll 1$, where μ is the electron mobility and c is the speed of light, is virtually always satisfied in our experiments). The magnetoresistance did not depend on the free-electron density, which can be varied from 10^7 up to the 10^{11} cm^{-3} by varying the illumination intensity. This shows that the interelectronic interaction indeed plays no role in our case. The inset shows the magnetic-field dependence of the negative magnetoresistance for the same conditions (in constructing these curves, the classical magnetoresistance was subtracted from the total magnetoresistance). In weak magnetic fields the negative magnetoresistance is proportional to the squared magnetic field (see Fig. 7a below). As the magnetic field increases further, the dependence of the negative magnetoresistance on H decreases and saturates. It is evident that in samples with different values of γ but close values of kT/γ, the negative magnetoresistance saturates at different values of H, and this value is virtually the same in the saturation region. Figure 2 shows curves of the negative magnetoresistance as a function of H for different values of T for samples with antimony density 1.5×10^{15} cm^{-3}. As the temperature increases, the negative magnetoresistance decreases and for $kT > \gamma$ it vanishes (for this sample for $T > 26$ K). The larger the value of γ, the higher the temperature at which negative magnetoresistance is observed is. For samples with antimony density 1.2×10^{16} cm^{-3} ($\gamma = 5.6$ meV) it is observed right up to 60 K (see Fig. 7b below).

Heating of the electron gas also suppresses the negative magnetoresistance. However, this occurs differently when the electrons are heated by an electric field and by light. In an electric field the negative magnetoresistance decreases in the saturation region (Fig. 3) and on the quadratic section (inset). Photoheating of electrons becomes substantial at low temperatures, where the rate of thermal ionization of the donors becomes less than the rate of ionization of the donors by the background radiation, and it intensifies as temperature decreases. Photoheating is manifested in an increase of the electron mobility μ at low temperatures, since in the presence of electron scattering by charged impurities the mobility μ should increase with the electron energy. In this case the saturated value of the negative magnetoresistance also decreases (see the experimental points for $\gamma = 5.6$ meV in Fig. 7b), but the slope of the curve of $\Delta\sigma/\sigma$ versus H on the quadratic section no longer changes with decreasing temperature.

Figure 4 shows the temperature dependence of the electron mobility μ, determined from the slope of the quadratic section of the positive magnetoresistance, for samples with different values of γ. The mobility was determined from the formula $\Delta\sigma/\sigma = -r_m(\mu H/c)^2$, where

be magnetoresistance constant $r_m = 4$ for impurity scattering was used. At temperatures below a critical value T_c the mobility saturates ($\mu = \mu_c$). At temperatures $T > T_c$ a power-law growth of the mobility with temperature is observed. The quantities kT_c for different samples were found to be close to the corresponding values of γ (see table). Figure 5 shows the dependence $\mu_c(\gamma)$. The saturated low-temperature value of the mobility μ_c was found to be inversely proportional to γ and not the density of charged centers, as should happen for impurity scattering.

Thus, the saturated value of the mobility and the interference corrections to the conductivity are determined by the random Coulomb potential of the charged impurities.

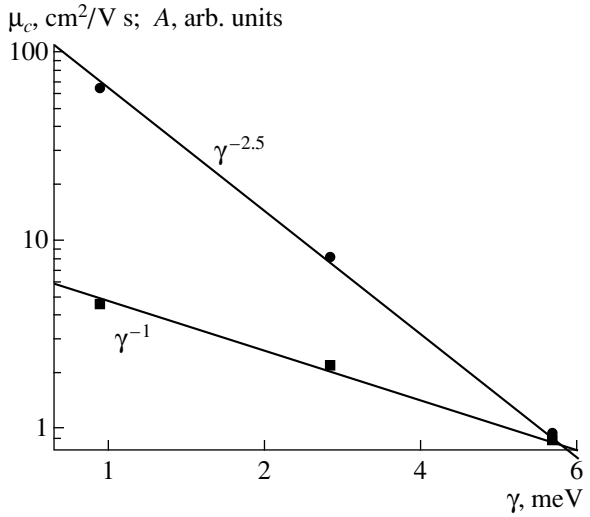

Fig. 5. Saturated mobility μ_c and coefficient A versus γ.

3. COMPARISON WITH THE THEORY

Saturation of the mobility at low temperatures attests to the existence of a quantum mobility threshold. It is known (see the reviews [8, 9] and the initial works cited) that in disordered metals, when the Fermi energy is close to ε_c, electronic diffusion vanishes at $T = 0$ in accordance with the expression

$$D = D_0\left(1 - \frac{\varepsilon_c}{\varepsilon_F}\right)^s,$$

where $D_0 = v^2\tau/3$ is the diffusion coefficient, v is the electron velocity, τ is the electron free-flight time (determined in our case by scattering by charged impurities), and s is a critical exponent. Correspondingly, the conductivity should vanish. For a nondegenerate semiconductor it is natural to study not the conductivity but rather the mobility of the electrons [10]. To determine the critical exponent s we used the expression

$$\mu = -\frac{2e}{3m}\frac{\int_{\varepsilon_c}^{\infty}\varepsilon\tau\nu\frac{df}{d\varepsilon}\left(1 - \frac{\varepsilon_c}{\varepsilon}\right)^s d\varepsilon}{\int_{\varepsilon_c}^{\infty}\nu f\, d\varepsilon}. \quad (1)$$

Here $\nu(\varepsilon)$ is the density of states and f is the distribution function. The lower limit of integration and the factor $(1 - \varepsilon_c/\varepsilon)^s$ take account of the fact that diffusion vanishes at the mobility threshold. For a Boltzmann distribution function, $f = \exp[-(\varepsilon - \varepsilon_F)/kT]$, the limit of the expression (1) as $T \longrightarrow 0$ gives $\mu \propto (kT)^{s-1}$. Mobility saturation can be observed only if $s = 1$, in agreement with the prediction of the theory [9, 11–13] for the quantum mobility threshold. For the classical percolation threshold the critical exponent $s \approx 1.7$–1.9 [7], and mobility saturation should not be observed. We note that the theoretical conclusion that s is exactly 1 is not generally accepted. In its explicit form this hypothesis has been formulated in [14]. The theoretical derivation of this assertion, claiming mathematical strictness, is given in [13]. On the other hand, numerical calculations [15] give $s = 1.3$–1.5. However, our experimental data support the hypothesis $s = 1$.

The condition for strong Ioffe–Regel–Mott localization, $\lambda \approx l$, equivalent to the relation $\varepsilon_c \approx \hbar/2\tau_c$, can be used to determine the threshold energy ε_c. The values of the mobility threshold ε_c determined for different samples according to the low-temperature values of $\tau_c = m\mu_c/e$ (m is the effective mass), turned out to be close to the corresponding values of γ (see table). Thus, the value of the saturated mobility is related with the amplitude of the random potential of charged impurities as $\mu_c \approx e\hbar/2m\gamma$. This relation can be interpreted as the quantum limit of the mobility, which is reached as a result of the modulation of the electron wave function by the random potential. Treating γ as the quantum uncertainty of the energy and taking account of the fact that this uncertainty and the mobility threshold should be of the order of the electron energy, we have for the momentum uncertainty $\Delta p = \sqrt{2m\gamma}/2$. In terms of scattering, this modulation determines the mean free path

$$l = \hbar/2\Delta p = \hbar/\sqrt{2m\gamma} = \lambda(\gamma).$$

This, at the energy γ we have $l \approx \lambda$, but both quantities have a strongly nonclassical meaning. We note that the classical mean free path for scattering by the random potential cannot be less than the radius of the optimal fluctuation [7], which in our samples is approximately 30 times greater than the mean free path determined according to the saturated value τ_c has $l = \tau_c/\sqrt{2\gamma/m}$. Thus, mobility saturation at low temperatures and its dependence on γ show that electron scattering at temperatures $kT < \gamma$ is strongly nonclassical.

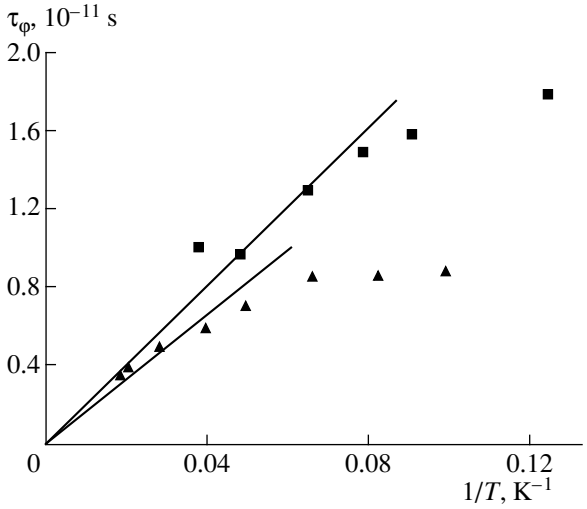

Fig. 6. Temperature dependence of the phase interruption time.

For $kT > \gamma$ the mobility starts to increase. We endeavored to compare $\mu(T)$ with the standard temperature dependence of the impurity mobility $\mu(T) = AT^{3/2}$, shown by the straight lines in Fig. 4. The coefficient A (see Fig. 5) turned out to depend on γ ($A \propto \gamma^{-2.5}$) and not on the effective density of charged centers ($N_{\text{eff}} = \sum Z^2 N_Z$, N_Z is the density of centers with charge Z). This is clearly seen by comparing the curves in Fig. 4 for samples with $\gamma = 1.2$ and 2.7 meV, for which N_{eff} has approximately the same value, while the amplitude of the random potential is different because the degree of compensation is different. The fact that μ is related with the amplitude of the random potential shows that the mobility in this temperature range is determined not only by the energy dependence of the mean-free flight time but also by a transition to a percolation character of the mobility.

We shall now examine the characteristic features of the negative magnetoresistance. The main difficulty for the analysis of our experimental results using the existing theory of weak localization [4, 5,16] is that theory makes use of the condition $\lambda \ll l$, which does not hold near the mobility threshold. For this reason, the quantitative comparison, presented below, of the experimental data with the theory can be viewed only as plausible estimates, since at present a strict theory of interference corrections to the conductivity near the mobility threshold has not been developed. Nevertheless, we endeavored to compare these data with the calculations. The dependence of the negative magnetoresistance on the magnetic field for a three-dimensional noninteracting electron gas is described, according to the theory of weak localization, by the expression

$$\frac{\Delta\sigma}{\sigma} = \frac{3}{2}\frac{\lambda^2}{l l_H} f_3\left(\frac{2l_\varphi}{l_H}\right), \qquad (2)$$

derived for a degenerate electron gas under the conditions $\lambda \ll l \ll l_\varphi, l_H$. Here $l = v\tau$, $l_H = \sqrt{c\hbar/eH}$ is the magnetic length, $l_\varphi^2 = D_0\tau_\varphi$, τ_φ is the phase interruption time (coherence break down time) of the electronic wave because of inelastic scattering (in our case – by acoustic phonons [16]), and f_3 is a function found in [4], and all quantities are referred to the Fermi energy. For weak magnetic fields ($2l_\varphi/l_H \ll 1$) the asymptotic expression $f_3 \approx (1/48)(2l_\varphi/l_H)^3$ is valid [4]. This gives a quadratic dependence of the negative magnetoresistance on H:

$$\frac{\Delta\sigma}{\sigma} = \frac{1}{64}\frac{\lambda^2}{l l_\varphi}\left(\frac{2l_\varphi}{l_H}\right)^4 \equiv \frac{1}{12\sqrt{3}}\left(\frac{\tau_\varphi}{\tau}\right)^{3/2}(\omega_c\tau)^2, \qquad (3)$$

$\omega_c = eH/mc$ is the cyclotron frequency. The asymptotic dependence $\Delta\sigma/\sigma \propto \sqrt{H}$ is expected in the magnetic-field range where $l < l_H \ll l_\varphi$. Finally, in strong fields, where $l_H \approx l$, the interference corrections to the conductivity should be completely suppressed by the magnetic field, and the negative magnetoresistance should saturate.

The ratio λ/l can be estimated using the expression (3). The deviation from a quadratic dependence first appear in magnetic fields where $l_H \approx 2l_\varphi$. Substituting the corresponding experimental values of H and $\Delta\sigma/\sigma$ into equation (3), we obtain once again $l \approx \lambda$, i.e., the negative magnetoresistance data likewise indicate the existence of a mobility threshold.

The time τ_φ can be determined using the expression on the right-hand side in equation (3). It is easy to verify that here the energy dependence of the transport coefficients can be neglected. Even assuming that these dependences are classical ($\tau_\varphi(\varepsilon) \propto \varepsilon^{-1/2}$, $\tau(\varepsilon) \propto \varepsilon^{3/2}$) we obtain $\Delta\sigma/\sigma \propto \tau_\varphi^{3/2}\tau^{1/2} \propto \varepsilon^{-3/4}\varepsilon^{3/4} = \text{const}$. For this reason, the magnetoresistance is determined by the values of the transport coefficients at the mobility threshold. Here the parameter ε_c plays the same role as the Fermi level ε_F for a degenerate gas. This is the well-known basis for using equation (3) for a nondegenerate electron gas.

The temperature dependences of $\tau_\varphi \propto 1/T$ obtained from the quadratic section of the negative magnetoresistance using the expression (3) are presented in Fig. 6. It is evident that at high temperatures the phase interruption time τ_c, and it saturates as T decreases. The values of τ_φ agree well with the theoretical scattering times for electrons scattered by acoustic phonons in germanium. The free-flight times τ_c found from the saturated values of the mobility μ_c are 0.75×10^{-13}, 1.65×10^{-13}, and 3.4×10^{-13} s for samples with $\gamma = 1.2$, 2.8, and 5.6 meV, respectively, and are two orders of magnitude

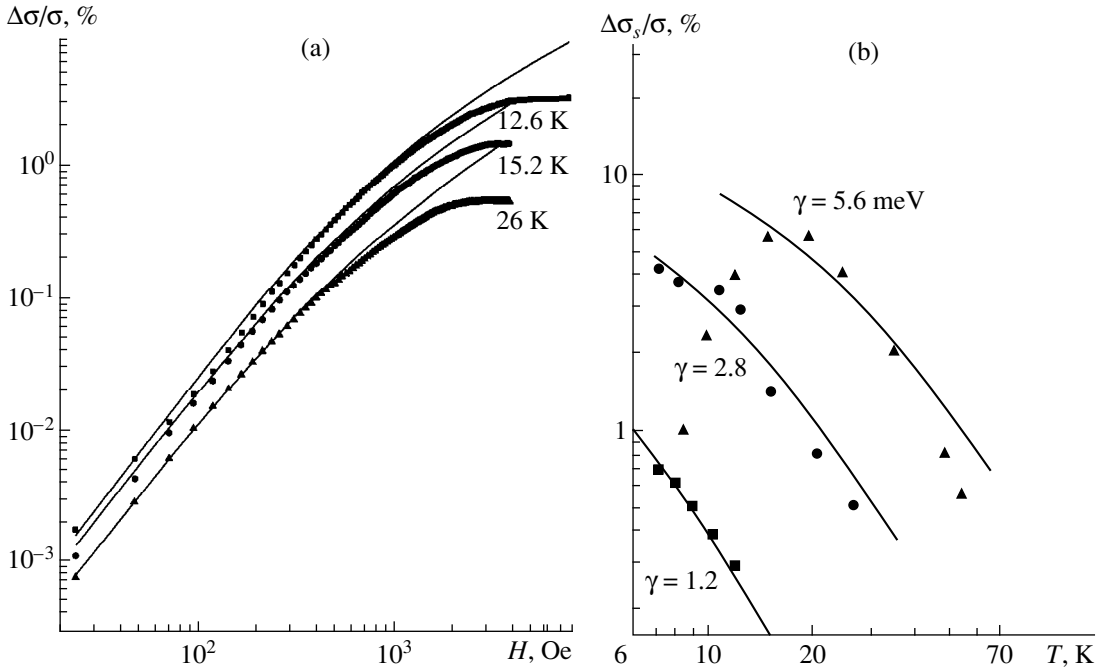

Fig. 7. (a) Negative magnetoresistance at different temperatures. Solid lines represent the calculation using equation (4). (b) Temperature dependence of the quantum corrections to the conductivity. Solid lines represent the calculation using equation (6), averaged over the Boltzmann distribution function.

less than τ_φ, i.e., the conditions $\tau_\varphi \gg \tau$, which is necessary for negative magnetoresistance to arise, was indeed satisfied in our samples. The phase interruption time τ_φ is inversely proportional to the temperature. This is natural, since for scattering by acoustic phonons we have (had energy ε_c) $\tau_\varphi = \tau_{ac} \propto \varepsilon_c^{-1/2} T^{-1}$. The weakening of the temperature dependence of τ_φ at low T is due, in our opinion, to scattering by the zero-point lattice vibrations, which is manifested at such comparatively high temperatures, once again, because of the existence of a mobility threshold. Indeed, scattering by zero-point vibrations should become appreciable when the energy loss, equal to $\sqrt{8mu^2\varepsilon}$, in a collision of an electron with energy ε with a phonon (u is the sound velocity) becomes comparable to kT. For the standard Boltzmann gas, when $\varepsilon \approx kT$, this gives very low temperatures, $kT < mu^2$. Near the mobility threshold, at $\varepsilon \approx \varepsilon_c \approx \gamma$, to reach zero scattering with $\gamma = 2.8$ and 5.6 meV we obtain, respectively, $T < 10$ K and $T < 7$ K, which is close to the observed temperatures (see Fig. 6). We note that the critical temperature $T_c \propto 1/\sqrt{\gamma}$ at which τ_φ saturates, while the saturated value $\tau_\varphi \propto (T_c\sqrt{\gamma})^{-1} \propto 1/\gamma$ also agrees with experiment (Fig. 6). Of course, in these arguments it is essential that in the random potential the minimum kinetic energy of electrons participating in conduction is of the order of γ.

We note that the saturation of τ_φ with decreasing temperature has also been observed in a degenerate electron gas [17, 18]. The theoretical interpretation [19, 20] of this saturation was based on the large role of the interelectron interaction, which is absent in our case.

We used the expression (2) for the single-electron quantum corrections, averaging it over the Boltzmann distribution function, to calculate the negative magnetoresistance in the entire range of magnetic fields:

$$\frac{\Delta\sigma}{\sigma} = \left(\frac{3}{2}\int_{\varepsilon_c}^{\infty} d\varepsilon\varepsilon\tau\nu\frac{\lambda^2}{ll_H}\left(1-\frac{\lambda}{l}\right)e^{-\varepsilon/kT} f_3\left(\frac{2l_\varphi}{l_H}\right)\right) \times\left(\int_{\varepsilon_c}^{\infty} d\varepsilon\varepsilon\tau\nu\left(1-\frac{\lambda}{l}\right)e^{-\varepsilon/kT}\right)^{-1}. \quad (4)$$

Here the factor $1 - \lambda/l$ takes account of the fact that the diffusion vanishes at the mobility threshold. The energy dependences of the transport coefficients entering here where assumed to be classical ($\lambda = \hbar/\sqrt{2m\varepsilon}$, $\tau \propto \varepsilon^{3/2}$, $\tau_\varphi = \tau_{ac} \propto \varepsilon^{-1/2}T^{-1}$, and $\nu(\varepsilon) \propto \sqrt{\varepsilon}$). (We note that the specific energy dependence $\nu(\varepsilon)$ of the density of states is not essential for our purposes. Its value at the mobility threshold is, of course, important.) We also used the condition $\varepsilon_c = \hbar/2\tau_c$, so that the only adjustable parameter is the mobility threshold ε_c. It is easy to show that for small H the expression (4) as $T \longrightarrow 0$ reduces to the formula (3) with energy ε_c. The computational results

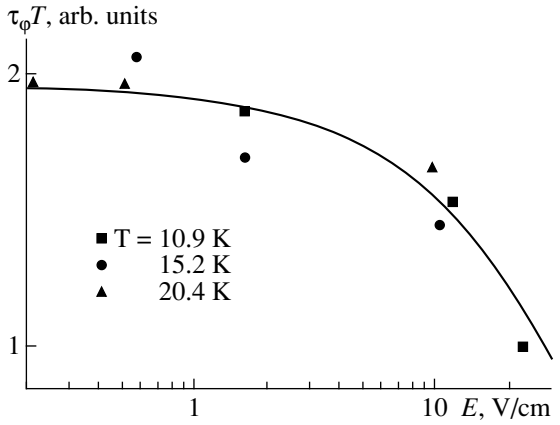

Fig. 8. $\tau_\varphi T$ versus the electric field. Solid line represents calculation using equation (7).

are shown by the solid lines in Fig. 7a. We note that at low temperatures, $kT \ll \varepsilon_c$, the energy dependences of the transport coefficients become negligible, so that the curves calculated using equations (2) and (4) are identical (top curve in Fig. 7a). The values found for ε_c on the basis of the best agreement with experiment are presented in the table. They agree quite well with the value of ε_c determined from the saturated mobility and with the values of γ.

It is evident from Fig. 7a that the experimental magnetic-field dependences of the negative magnetoresistance agree well with the dependences calculated using equation (4) in a quite wide range of magnetic fields, but saturation of the experimental negative magnetoresistance is reached in much weaker magnetic fields then expected (corresponding to the condition $l_H \approx l$). In our case the quantity l obtained near the mobility threshold from the experimental data ($l = v\tau_c$, $v = \sqrt{2\varepsilon_c/m}$) is of the quarter of 10^{-6} cm, and saturation of the negative magnetoresistance could be expected for magnetic fields at least an order of magnitude greater than observed experimentally (see Figs. 1, 2, and 7). The saturation of the negative magnetoresistance was observed starting in fields where $l_H \approx l_\varphi$. For this reason, the square-root dependence $\Delta\sigma/\sigma \propto \sqrt{H}$, which should occur for $l_H \ll l_\varphi$, is not observed in the experiment.

The saturated values of the relative negative magnetoresistance $\Delta\sigma_s/\sigma$ give quantum corrections to the conductivity. The dependence $\Delta\sigma_s/\sigma(T)$ for samples with different values of γ is shown in Fig. 7b. As T increases, the interference corrections strongly decrease. We note that their magnitude does not exceed several percent, though at the mobility threshold the electron wavelength λ is close to the electron mean-free path l, and it was to be expected that $\Delta\sigma_s/\sigma$ will be of the order of 1.

4. DISCUSSION

We shall consider first the characteristic features of the negative magnetoresistance. According to the theory [4, 16], interference of electronic waves on closed trajectories with radii ranging from l to l_φ makes the main contribution to the quantum corrections to the conductivity. Low values of the magnetic fields in which saturation occurs (large magnetic lengths $l_H \gg l$) indicate that the contribution of short trajectories to the interference corrections is small. Short closed trajectories disappear because of the presence of the large-scale random potential of charged impurities, on account of which a random three-dimensional conducting network forms near the percolation threshold [7]. In this case, for interference an electron must go around the "humps" in the large-scale potential relief. The characteristics scale of the random potential is of the order of the radius R_s of the optimal fluctuation. An estimate of R_s using equation (13.8) from [7] shows that in our case $R_s \gg l_\varphi$. The probability of short closed trajectories is low, so that the interference corrections to the conductivity should be strongly suppressed [5]. To illustrate this we estimated the interference corrections to the conductivity for the three-dimensional case using and expression from [5]

$$\frac{\Delta\sigma_s}{\sigma} = \int_\tau^{\tau_\varphi} \frac{\lambda^2 v}{(Dt)^{3/2}} F[x(t)]dt, \qquad (5)$$

multiplying the integrand by the distribution function $F(x)$ of closed electronic trajectories over their lengths x. For $F(x)$ we used the expression $F = \exp(-l_\varphi/vt)$, which, generally speaking, requires substantiation. We used it simply by analogy with equation (39.10) in [7], $F = \exp(-R_s/R)$, derived for the distribution function of fluctuations with the potential γ over the radii R. Here we took account of the fact that the length of a trajectory it is vt and that interference becomes negligible at distances greater than l_φ. Calculation of the integral on the right-hand side of equation (5) gives

$$\frac{\Delta\sigma_s}{\sigma} = \frac{\lambda^2}{ll_\varphi}. \qquad (6)$$

Averaging this expression over the Boltzmann distribution function we obtained the temperature dependences of $\Delta\sigma_s/\sigma$ shown by the solid lines in Fig. 7b. They agree well with the experimental data. (The decrease in $\Delta\sigma_s/\sigma$ at low temperatures, which is shown for a sample with $\gamma = 5.6$ meV, as indicated above, is due to photoheating.) We note that here, once again, the mobility threshold ε_c was the only adjustable parameter. It agrees with the value of ε_c found by other methods. Thus, the small values of the quantum corrections to the conductivity near the mobility threshold are due to the large-scale random potential of the charged impurities, which for a three-dimensional conducting network with a scale of the order of the percolation radius, in

which the probability of closed electronic trajectories with a small radius is exponentially small. Actually, as one can see from the expression (6), interference occurs only on trajectories with a radius of the quarter of l_φ. This explains the low saturation magnetic fields for the negative magnetoresistance. Indeed, in the experiment this occurs for $l_H \approx l_\varphi$.

We shall now consider the effect of heating of the electron gas on the interference corrections to the conductivity. In the presence of photoheating the form of the electron energy distribution function f changes. The Boltzmann exponential is replaced by a much weaker dependence of the form $f \propto 1/\varepsilon^2$ [21]. For this reason, the contribution of high-energy electrons to the quantum corrections (the quantity λ^2/ll_φ averaged over the distribution function) increases and the saturated negative magnetoresistance strongly decreases at low temperatures (points for $\gamma = 5.6$ meV for $T < 20$ K in Fig. 7b). On the quadratic section of the negative magnetoresistance the energy dependence of the transport coefficients is negligible, and they are referred to the energy ε_c. For this reason, here, photoheating has no effect on the negative magnetoresistance. In an electric field it decreases in the saturation region and on the quadratic section of the dependence $\Delta\sigma/\sigma(H)$ (see Fig. 3). The latter means that the quantity τ_φ decreases with increasing electric field (Fig. 8). (To eliminate the dependence $\tau_\varphi(T)$, the field dependence of the product $\tau_\varphi T$ is shown in Fig. 8.) The decrease in τ_φ in an electric field cannot be explained by heating of the electron gas. Indeed, an electron in an electric field E acquires over the mean free path an additional energy eEl. Substituting here the experimental values of l, we obtained at $eEl \ll \gamma, kT$ for $E = 30$ V/cm. On the other hand, the energy acquired by an electron between elastic collisions on different sections of a closed trajectory is a random quantity (of the order of eEl) because of the random character of the scattering events. For this reason, the uncertainty of the electron energy as result of inelastic scattering can be represented in the form

$$\frac{\hbar}{\tau_\varphi^E} = \frac{\hbar}{\tau_\varphi} + eEl. \qquad (7)$$

The quantity τ_φ^E calculated using this expression is shown in Fig. 8 (solid line). We note that the observed dependence $\tau_\varphi(E)$ is at variance with the theoretical representations [22], according to which in the absence of heating the electric field should not affect the phase interruption time, since the resulting change in the electron energy on a closed trajectory is zero.

Let us now see what the idea of a random conducting network for the observed features of the mobility. At the percolation threshold the transverse section of the conducting channels ("wires") of this random network approaches zero, and the characteristics size of the "network" (the correlation radius) approaches infinity. On the other hand, according to the experiments the conductivity (mobility) at low temperatures is determined by the existence of the quantum mobility threshold. In the "wire network" language this means that the minimum electron energy for which conductivity still exists determines the minimum cross section of a wire of the order of λ^2. The existence of such a "quantum percolation" threshold [23] makes it possible to estimate the threshold mobility from other physical considerations. We shall consider the percolation network to be a three-dimensional system of one-dimensional channels in which an electron moves without scattering. The minimum conductivity of each such channel is of the order of $e^2/\hbar$. The specific conductivity of the random network is determined by the conductivity of an individual element of the network and is equal to $e^2/\hbar\xi$ [24], where ξ is the correlation radius. To obtain the mobility we must multiply the specific conductivity σ by the volume occupied by a single electron ($\xi\lambda^2$) and divide by the elementary charge e. This result is $\mu = e\lambda^2/\hbar$. For $\varepsilon_c \approx \gamma$ this expression is equivalent to the condition $\gamma = \hbar/2\tau$, found from experiment, and it is actually the Ioffe–Regel' condition.

5. CONCLUSIONS

The data presented in this paper show that the mobility threshold for conducting electrons in the random Coulomb potential of charged impurities is of a quantum nature. The physical reason for this threshold is the existence of the percolation conducting network and its minimum conductivity, and not coherent backscattering, which leads to the appearance of interference corrections to the conductivity. Due to the presence of the large-scale random potential, these corrections, just as the negative magnetoresistance caused by their suppression by a magnetic field, are small. Because of inelastic scattering, the phase interruption time τ_φ of an electron wave is equal to the free-flight time in the presence of scattering by acoustic phonons, and it saturates at low temperatures. The saturation of τ_φ is due to scattering by zero-point lattice vibrations. The values of τ_φ decrease in comparatively weak electric fields, when appreciable heating of the electron gas still does not occur. The average free-flight time τ saturates at low temperatures, and its saturated values τ_c is the quantum limit of τ, related with the amplitude γ of the random potential by the relation $\tau_c = \hbar/2\gamma$.

ACKNOWLEDGMENTS

We thank A.I. Larkin, A.M. Finkel'shtein, A.Ya. Shul'man, V.A. Volkov, I.M. Suslov, and M.A. Baranov for a discussion of the results. This work was supported by the Russian Foundation for Basic Research (project no. 97-02-16819).

REFERENCES

1. P. W. Adams and M. A. Paalanen, Phys. Rev. Lett. **58**, 2106 (1987).
2. A. M. Bolibok, N. G. Zhdanova, M. S. Kagan, and E. G. Landsberg, Pis'ma Zh. Éksp. Teor. Fiz. **45**, 152 (1987) [JETP Lett. **45**, 188 (1987)]; A. M. Bolibok, M. S. Kagan, and E. G. Landsberg, Zh. Éksp. Teor. Fiz. **94**(2), 267 (1988) [Sov. Phys. JETP **67**, 367 (1988)].
3. E. M. Gershenzon, L. B. Litvak-Gorskaya, G. Ya. Lugovaya, and E. Z. Shapiro, Fiz. Tekh. Poluprovodn. (Leningrad) **20**, 99 (1986) [Sov. Phys. Semicond. **20**, 58 (1986)].
4. A. Kawabata, J. Phys. Soc. Jpn. **49**, 628 (1980).
5. B. L. Altshuler, A. G. Aronov, D. E. Khmelnitskiĭ, and A. I. Larkin, in *Quantum Theory of Solids*, Ed. by I. M. Lifshits (Mir, Moscow, 1982).
6. N. G. Zhdanova, M. S. Kagan, E. G. Landsberg, and V. V. Petrishchev, Pis'ma Zh. Éksp. Teor. Fiz. **58**, 529 (1993) [JETP Lett. **58**, 533 (1993)].
7. B. I. Shklovskiĭ and A. L. Éfros, *Electronic Properties of Doped Semiconductors* (Nauka, Moscow, 1979; Springer-Verlag, New York, 1984).
8. N. Mott, J. Phys. C **20**, 3075 (1987).
9. P. A. Lee and T. V. Ramakrishnan, Rev. Mod. Phys. **57**, 287 (1985).
10. M. H. Cohen, E. N. Economou, and C. M. Soucoulis, Phys. Rev. B **30**, 4493 (1984).
11. A. Kawabata, Solid State Commun. **38**, 823 (1981).
12. B. Shapiro, Phys. Rev. B **25**, 4266 (1982).
13. I. M. Suslov, Zh. Éksp. Teor. Fiz. **108**, 1686 (1995) [JETP **81**, 925 (1995)].
14. H. Kunz and R. Souillard, J. Phys. Lett. **44**, L506 (1983).
15. Isa Kh. Zharekeshev and B. Kramer, Phys. Rev. Lett. **79**, 717 (1997).
16. B. L. Al'tshuler, A. G. Aronov, D. E. Khmel'nitskiĭ, and A. I. Larkin, Zh. Éksp. Teor. Fiz. **81**, 768 (1981) [Sov. Phys. JETP **54**, 411 (1981)].
17. P. Mohanty, E. M. Q. Jarivala, and R. A. Webb, Phys. Rev. Lett. **77**, 3366 (1997); P. Mohanty and R. A. Webb, Phys. Rev. B **55**, 13452 (1997).
18. A. G. Huibers, M. Switkes, C. M. Marcus, *et al.*, Phys. Rev. Lett. **81**, 200 (1998).
19. I. L. Aleiner, B. L. Altshuler, and M. E. Gershenzon, cond-mat/9808053.
20. D. S. Golubev and A. D. Zaikin, Phys. Rev. Lett. **81**, 1074 (1998).
21. Sh. M. Kogan, Fiz. Tekh. Poluprovodn. (Leningrad) **11**, 1979 (1977) [Sov. Phys. Semicond. **11**, 11558 (1977)]; V. N. Abakumov, V. I. Perel, and I. N. Yassievich, *Modern Problems in Condensed Matter Sciences*, Ed. by V. M. Agranovich and A. A. Maradudin (North-Holland, Amsterdam, 1994), Vol. 33.
22. B. L. Altshuler, A. G. Aronov, and D. E. Khmelnitskii, Solid State Commun. **39**, 619 (1981).
23. B. Shapiro, Phys. Rev. Lett. **48**, 823 (1982).
24. W. L. McMillan, Phys. Rev. B **24**, 2739 (1981).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 4, 2000, pp. 671–688.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 4, 2000, pp. 771–789.

SOLIDS
Electronic Properties

Gluing of Josephson Vortices by Cherenkov-Trapped Swihart Waves

A. S. Malishevskiĭ*, V. P. Silin, and S. A. Uryupin
Lebedev Physical Institute, Russian Academy of Sciences, Moscow, 117924 Russia
**e-mail: malish@sci.lebedev.ru*
Received November 1, 1999

Abstract—A theory of Josephson vortices-multikinks moving with a constant velocity is formulated for nondissipative Josephson junctions on the basis of the Aubry–Volkov analytical model. Cherenkov trapping of Swihart waves by moving vortices is predicted. A description of the structure of vortices which is due to the waves trapped by them is given. It is predicted that the characteristic velocities of the moving vortices are discrete. This phenomenon is due to the structure of the Swihart waves trapped by the vortices. The concept of gluing of elementary Josephson vortices by Swihart waves in a coherently-phased, multivortex, structure is advanced. © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Interest in moving Josephson vortices, carrying $n > 1$ magnetic flux quanta, called $2\pi n$-kinks in long Josephson junctions arose a comparatively long time ago [1]. For discrete Josephson junctions, important theoretical results were obtained in [2]. The definite progress made in the description of long Josephson junctions was obtained in a theory that takes account of strong surface losses due to normal electrons [3]. However, in [3], in the first place, the case of very low temperatures, when there are no normal electrons in a superconductor, was neglected and, in the second place, the phenomena predicted by our work (see below) had not been observed. For long Josephson junctions without dissipation, an analytical description of Josephson vortices (4π-kinks) moving with a constant velocity was given in [4–6], which are based on local Josephson electrodynamics applied to the case of a very high critical Josephson current density. The latter was the reason why the results of [4–6] did not attract a great deal of attention. The presence of a moving 4π-kink was established numerically in [7], likewise on the basis of nonlocal Josephson electrodynamics. The important step made in [7] was the analysis of an approximation of nonlocal Josephson electrodynamics that corresponds to weak nonlocality, which takes account of the fourth-order spatial derivative together with the second derivative in the equation for the phase difference. As shown in our work, this approximation is suitable for describing Josephson vortices in Josephson junctions with a low critical current density. However, the work [7] did not attract attention. The next step was made in [8], where a 4π-kink was once again studied on the basis of nonlocal Josephson electrodynamics. Here, it should be said that the numerical analysis in [7, 8] led to a completely definite velocity of 4π-kinks, which corresponded to the analytical results of [4–6]. In [8], an analogous property was established for the 6π-kink, obtained for the first time numerically and corresponding to a Josephson vortex carrying three magnetic flux quanta. It is very important to note that in [8], together with a 4π-kink with a monotonic spatial dependence, a 4π-kink with a nonmonotonic dependence was also found. The idea of the existence of a similar 6π-kink was advanced in [8]. We note that the nonmonotonic and monotonic 4π-kinks moved with different velocities. To some extent this has something in common with the results of [2] for the discrete model of a Josephson junction.

The material which has now accumulated on the theory of moving Josephson vortices in long nondissipative Josephson junctions poses a variety of questions for us. These include: 1) the question of the possibility of the existence of various multikink; 2) the question of the velocities of such multikinks; and 3) the question of why such motion occurs with definite velocities. Even on the basis of the results of [4–8] on nonlocal Josephson electrodynamics, it can be asserted that nonlocal Josephson electrodynamics, in contrast to the standard local Josephson electrodynamics, in many cases was successful in finding multikinks in long Josephson junctions without dissipation. It should be underscored that the standard Josephson electrodynamics of long Josephson junctions does not describe the physical phenomenon of Cherenkov interaction of Josephson vortices with Swihart waves, which is of decisive importance for understanding the effects predicted below, since in the standard Josephson electrodynamics the velocity of a vortex is always less than the phase velocity of the waves. These physical considerations are the basis for using nonlocal Josephson electrodynamics in the present paper. The content of this paper attests to the adequacy of the physical model which we have selected for nonlocal Josephson electrodynamics for describing the phenomena which we predict. In the present paper the physical results pre-

dicting qualitatively new properties of moving Josephson vortices are presented. The Aubry–Volkov model [9–12] (see also [13]) is used. In this model, instead of the standard sinusoidal nonlinearity, the relation

$$j(\varphi) = j_c\{(\varphi/\pi) - 2I[(\varphi/2\pi) + 1/2]\}, \qquad (1.1)$$

where j_c is the critical current density and $I[x]$ is the integer part of the number x, is used for the relation between the superconducting current j and the phase difference φ of Cooper pairs on different sides of a Josephson junction. The nonlinearity (1.1) corresponds to that indicated in [14]. We have published preliminary results for 4π- and 6π-kinks in the Aubry–Volkov model (1.1) in [15, 16].

In ordinary local Josephson electrodynamics the equation

$$\frac{1}{\omega_j^2}\frac{\partial^2\varphi}{\partial t^2} - \lambda_j^2\frac{\partial^2\varphi}{\partial z^2} + \pi\frac{j(\varphi)}{j_c} = 0, \qquad (1.2)$$

where $\pi^{-1}j_c\sin\varphi$ is often chosen for the function $j(\varphi)$, is used to describe free one-dimensional vortices. Here and below, following the Aubry–Volkov model, we shall work with the expression (1.1). In equation (1.2)

$$\omega_j = 4(|e|j_c d/\hbar\varepsilon)^{1/2}, \qquad (1.3)$$

$$\lambda_j = c(\hbar/\lambda|e|j_c)^{1/2}/4 \qquad (1.4)$$

are, respectively, the Josephson frequency and length, which are defined somewhat differently than in [11], which is dictated by the possibility of describing vortex solutions graphically in a form similar to the generally accepted one. In equations (1.3) and (1.4) $2d$ is the width of a nonsuperconducting layer, ε is the permittivity of this layer, e is the electron charge, c is the speed of light, $\hbar$ is Planck's constant, and λ is the London penetration depth.

The Aubry–Volkov model presupposes not only the use of the relations (1.1) but also the application of the Fourier transform in finding the required solutions. This makes it possible, specifically, to study the discontinuous functions $\operatorname{sgn}x$ (sign function), $\theta(x)$ (Heaviside function), and $I[x]$ as functions determined in accordance with the Fourier transform, when the values of the functions at a jump are given by the half-sum of the values to the left and right of the jump. This refinement makes the mathematical apparatus which we used below quite understandable.

A very important advantage of the Aubry–Volkov model over the standard model of a sinusoidal nonlinearity should be underscored. It is expressed in the fact that the Aubry–Volkov model makes it possible to use the Fourier transform, which ordinarily is successful for solving linear problems, for solving nonlinear problems of nonlocal Josephson electrodynamics successfully and productively. At the same time, the problems of finding the nonlinear states are solved using an accurate treatment of discontinuous functions. The corresponding simplicity of the Aubry–Volkov model is its advantage on the one hand and a disadvantage on the other. Specifically, the use of the relation (1.1) of the Aubry–Volkov model together with other possible nonlinearities (compare [14]) gives in many cases only a qualitative description of Josephson vortices. This indicates, specifically, the limits of applicability of the Aubry–Volkov model. We shall indicate a well-known qualitative property of the Aubry–Volkov model. Specifically, in nonlinear Josephson electrodynamics such a model forbids (see below) the existence of freely moving 2π-kinks. However, even this drawback can be viewed as an advantage, because it indicates a way to improve the model.

For comparison with the results obtained below, we indicate that the solution of equations (1.1) and (1.2), describing a Josephson vortex moving with a constant velocity v and carrying an elementary flux (quantum) $\phi_0 \equiv \pi\hbar c/|e|$ of the magnetic field (2π-kink), has the form

$$\varphi_{2\pi}(z - vt) = \pi + \pi\left[1 - \exp\left(-\frac{|z - vt|}{\lambda_j\sqrt{1 - v^2/v_s^2}}\right)\right]\operatorname{sgn}(z - vt), \qquad (1.5)$$

where $v_s = \omega_j\lambda_j$ is the Swihart velocity. This solution corresponds to the following expressions for the intensity of the magnetic field in a nonsuperconducting layer of a Josephson function (compare [11]):

$$H_y(z - vt) = -\frac{\phi_0}{4\pi\lambda\lambda_j}\exp\left(-\frac{|z - vt|}{\lambda_j\sqrt{1 - v^2/v_s^2}}\right) \qquad (1.6)$$

and the energy of the vortex

$$W = W_{2\pi}/\sqrt{1 - v^2/v_s^2}, \qquad (1.7)$$

where $W_{2\pi} \equiv \phi_0^2/32\pi\lambda\lambda_j$ is the energy of a 2π-kink at rest.

In the present paper, a general approach is given for theoretical investigation of $2\pi n$-kinks (Section 2). In Section 3 the general situation concerning the phenomenon of Cherenkov trapping of generalized Swihart waves moving with the Josephson vortices is formulated. Sections 4 and 5 are devoted to the structure of 4π- and 6π-kinks, carrying trapped waves, which determines the spectrum of the characteristic velocities of Josephson vortices. In Section 6, 8π-kinks, which are Josephson vortices carrying four magnetic flux quanta, are examined. Such kinks have not been previously discussed.

Among the results obtained in our paper, besides ascertaining the establishment of new vortex solutions of Josephson electrodynamics, the theoretical prediction of Cherenkov trapping of Swihart waves by Josephson vortices should be underscored. This phenomenon determines the spectrum of the values of the velocity of the free

motion of vortices and predicts the phenomenon of gluing of Josephson vortices by trapped Swihart waves.

2. TRAVELING $2\pi n$-KINK

We shall consider one-dimensional vortices traveling with constant velocity v. For these vortices $\varphi(z, t) = \psi(z - vt) \equiv \psi(\zeta)$, and on the basis of the nonlocal Josephson electrodynamics the following equation is satisfied:

$$F(\psi) + \frac{v^2}{\omega_j^2}\frac{d^2\psi(\zeta)}{d\zeta^2} = \frac{l}{\pi}\frac{d}{d\zeta}\int_{-\infty}^{\infty} d\zeta' K_0\left(\frac{|\zeta - \zeta'|}{\lambda}\right)\frac{d\psi(\zeta')}{d\zeta'}. \quad (2.1)$$

Here $K_0(x)$ is the Macdonald function, $l \equiv \lambda_j^2/\lambda$, and $F(\psi)$ corresponds to equation (1.1):

$$F(\psi) = \psi - 2\pi I[(\psi/2\pi) + 1/2]. \quad (2.2)$$

A traveling $2\pi n$-kink is determined by the boundary conditions at infinity, $\psi(-\infty) = 0$ and $\psi(+\infty) = 2\pi n$, and by the set of conditions

$$\psi(\xi_1) = \pi, \quad \psi(\xi_2) = 3\pi, \ldots, \psi(\xi_n) = (2n - 1)\pi. \quad (2.3)$$

The last set of formulas and equation (1.4) make it possible to write

$$I[(\psi/2\pi) + 1/2] = \sum_{n'=1}^{n} \theta(\zeta - \xi_{n'}). \quad (2.4)$$

On account of equation (2.4), equation (2.1) is linear. For this reason, it can be solved by a Fourier transformation. Using the Fourier transform of the Heaviside function $\theta(\zeta - \xi_{n'})$,

$$\pi\delta(k) - i\exp(-ik\xi_{n'})\text{V.P.}(1/k), \quad (2.5)$$

where V.P.$(1/k)$ denotes the Cauchy principal value, we have for the solution of equation (2.1) that describes a moving $2\pi n$-kink

$$\psi(\zeta) = \sum_{n'=1}^{n} \psi_{2\pi}(\zeta - \xi_{n'}), \quad (2.6)$$

$$\psi_{2\pi}(\zeta) = \pi + \text{V.P.}\int_{-\infty}^{+\infty} dk k^{-1}\sin(k\zeta)\left[1 + K(k) - \frac{k^2v^2}{\omega_j^2}\right]^{-1} + \pi\int_{-\infty}^{+\infty} dk\cos(k\zeta)\delta\left[1 + K(k) - \frac{k^2v^2}{\omega_j^2}\right], \quad (2.7)$$

where

$$K(k) = \frac{lk^2}{\sqrt{1 + k^2\lambda^2}} \equiv \frac{\lambda_j^2k^2}{\lambda\sqrt{1 + k^2\lambda^2}}. \quad (2.8)$$

The vanishing of the argument of the δ function in equation (2.7) corresponds to the condition of Cherenkov interaction (resonance) of a vortex moving with constant velocity v and a generalized Swihart wave with spectrum $\omega(k)$:

$$\omega^2(k) \equiv \omega_j^2[1 + K(k)] = k^2v^2. \quad (2.9)$$

The formula (2.7) can be rewritten in a different, convenient form

$$\psi_{2\pi}(\zeta) = \pi + \pi[1 - f(\zeta)]\,\text{sgn}\,\zeta + 2\pi C\cos(k_0\zeta)\theta(-\zeta), \quad (2.10)$$

$$C = \frac{2}{\lambda_j^2k_0^2}\frac{(1 + \lambda^2k_0^2)^{3/2}}{2(v/v_s)^2(1 + \lambda^2k_0^2)^{3/2} - 2 - \lambda^2k_0^2}, \quad (2.11)$$

$$f(\zeta) = \frac{2}{\lambda_j^2k_0^2}\frac{(1 + \lambda^2k_1^2)^{3/2}\exp(-k_1|\zeta|)}{2 - \lambda^2k_1^2 - 2(v/v_s)^2(1 + \lambda^2k_1^2)^{3/2}} + \frac{2\lambda^2}{\pi\lambda_j^2}\int_1^{\infty} dr r\sqrt{r^2 - 1}\frac{\exp(-r|\zeta|/\lambda)}{r^4 + (r^2 - 1)[(\lambda/\lambda_j)^2 + (vr/v_s)^2]^2}. \quad (2.12)$$

Here $k_1 = k_1(v)$ is the modulus of the purely imaginary root of equation (2.9), describing the monotonic dependences of $2\pi n$-kinks, and $k_0 = k_0(v)$ is the modulus of the real root of equation (2.9), corresponding to generalized Swihart waves, interacting resonantly (by the Cherenkov mechanism) with moving vortices.

In the present paper, we shall be concerned with Josephson junctions with comparatively low Josephson critical current density, when

$$\lambda \ll \lambda_j. \quad (2.13)$$

This enables us to neglect the second term in equation (2.12), which we shall not mention below.

The formula (2.6), according to equations (2.9)–(2.12), satisfies boundary conditions at infinity. We shall show below that the satisfaction of the conditions (2.3) determines the quantities $\xi_{n'}$, the velocity v of a multikink, and therefore also the wave vector $k_0(v)$ of generalized Swihart waves trapped by the vortex.

In concluding this section we shall consider the general relations which determine the magnetic field and energy of a vortex in a long Josephson junction.

Following the general principles of the nonlocal electrodynamics of Josephson junctions [4], in the limit $\lambda \gg 2d$ the magnetic field directed along the y axis and

produced by the distribution of the phase difference $\psi(\zeta)$ can be represented in the form

$$H(x, z - vt) = H(x, \zeta) = -\frac{\phi_0}{4\pi^2\lambda^2} \times \int_{-\infty}^{\infty} d\zeta' K_0\left(\frac{\sqrt{(\zeta - \zeta')^2 + (x \mp d)^2}}{\lambda}\right)\frac{d\psi(\zeta')}{d\zeta'}. \quad (2.14)$$

Here $|x \mp d|$ is the distance from the coordinates $\pm d$ of the boundaries of the tunnel junction to the coordinate x of the point of observation in the superconductor.

When the current density through the junction is related with the phase difference by the relation (1.1), the total energy of the state described by the solution $\psi(\zeta)$ of equation (2.1) is given by

$$W = \frac{\phi_0^2}{64\pi^3\lambda\lambda_j^2} \times \int_{-\infty}^{\infty} d\zeta\left\{\left[\psi(\zeta) - 2\pi I\left[\frac{\psi(\zeta)}{2\pi} + \frac{1}{2}\right]\right]^2 + \frac{v^2}{\omega_j^2}\left(\frac{d\psi}{d\zeta}\right)^2\right\} + \frac{\phi_0^2}{64\pi^4\lambda^2}\int_{-\infty}^{\infty} d\zeta \int_{-\infty}^{\infty} d\zeta' K_0\left(\frac{|\zeta - \zeta'|}{\lambda}\right)\frac{d\psi(\zeta)}{d\zeta}\frac{d\psi(\zeta')}{d\zeta'}, \quad (2.15)$$

where the first term is the energy of the Josephson current, scaled to unit length along the y axis, and the second and third terms are the energies of electromagnetic fields in the tunnel junction and the superconductors [4].

3. GENERAL CONSISTENCY EQUATIONS AND CHERENKOV TRAPPING OF WAVES BY VORTICES

We shall now consider the consequences of the condition (2.3) for the example of several lowest-order kinks. First, we shall consider a very simple 2π-kink, for which, according to [11], the solution

$$\varphi_{2\pi}(z) = \pi + 2\int_0^{\infty} dk k^{-1}[1 + K(k)]^{-1}\sin(kz), \quad (3.1)$$

describing a 2π-kink at rest, holds. For a moving vortex, a 2π-kink, we have according to the expressions (2.6) and (2.10)

$$\psi(\xi_1) = \pi + \pi C \neq \pi, \quad (3.2)$$

when $C \neq 0$, which occurs for $v \neq 0$. This means that in the Aubry–Volkov model the Cherenkov interaction of a 2π-kink with Swihart waves, which leads to $C \neq 0$, forbids free motion of a 2π-kink. This conclusion is characteristic of the Aubry–Volkov model. For higher-order multikinks the Aubry–Volkov model leads to more interesting results. In what follows, using an appropriate choice of the origin of the coordinate system, we take $\xi_1 = -\xi_n$.

We now consider a 4π-kink. Setting $\xi_1 = -\zeta_0$ and $\xi_2 = \zeta_0$, we obtain from the conditions (2.3)

$$f(2\zeta_0) = C, \quad (3.3)$$

$$\cos(k_0\zeta_0) = 0. \quad (3.4)$$

The equation (3.4) possesses a discrete set of solutions

$$k_0\zeta_0 = k_0(v)\zeta_0 = \pi(n + 1/2), \quad n = 0, 1, 2, \ldots. \quad (3.5)$$

Eliminating ζ_0 from equation (3.3) and using equation (3.5), we obtain a single equation for the velocity v of a 4π-kink, determining the discrete set of characteristic velocities of the vortex. The relation (3.4) makes it possible to write down the solution (2.6) for a 4π-kink in the following form:

$$\psi_{4\pi}(\zeta) = \psi_{2\pi, m}(\zeta + \zeta_0) + \psi_{2\pi, m}(\zeta - \zeta_0) + \psi_w(\zeta, \zeta_0), \quad (3.6)$$

where

$$\psi_{2\pi, m}(\zeta) \equiv \pi + \pi[1 - f(\zeta)]\operatorname{sgn}\zeta \quad (3.7)$$

corresponds to a monotonic function, and

$$\psi_w(\zeta, \zeta_s) = 2\pi C\cos[k_0(v)(|\zeta| - \zeta_s)] \times [\theta(-\zeta + \zeta_s) - \theta(-\zeta - \zeta_s)]\operatorname{sgn}\zeta \quad (3.8)$$

describes an oscillatory function, corresponding to Swihart waves trapped by a vortex and localized inside a vortex in the interval $\Xi_s = [-\zeta_s, \zeta_s]$. At the boundaries of the trapping region the derivative of the function (3.8) is zero, which corresponds to zero energy flux of Swihart waves. Denoting the Swihart wavelength by $\lambda_0(v) = 2\pi/k_0(v)$, in accordance with equation (3.5) it can be asserted that the size of the region of trapping of waves by a 4π-kink is given by the relation $2\zeta_0(n) = \lambda_0(n)(n + 1/2)$. The corresponding detailed formulas are given in Section 4.

Similar general consistency equations also arise for a 6π-kink. Setting $\xi_1 = -\zeta_0$, $\xi_2 = 0$, and $\xi_3 = \zeta_0$, we obtain from the conditions (2.3)

$$\cos[k_0(v)\zeta_0] = -1/2, \quad (3.9)$$

$$f(\zeta_0) + f(2\zeta_0) = C. \quad (3.10)$$

These consistency equations determine the discrete "quantized" spectrum of the characteristic velocities $v(n)$ of 6π-kinks, the wave numbers $k_0(n)$ of generalized Swihart waves trapped by a 6π-kink, and the dimensions of the trapping regions $2\zeta_0(n) = \lambda_0(n)[n + 1/2 + (-1)^n/6]$,

$n = 0, 1, 2, \ldots$. Correspondingly, the solution (2.6) of equation (2.1) can be represented in the form

$$\psi_{6\pi}(\zeta) = \psi_{2\pi, m}(\zeta + \zeta_0) + \psi_{2\pi, m}(\zeta) + \psi_{2\pi, m}(\zeta - \zeta_0) + \psi_w(\zeta, \zeta_0). \quad (3.11)$$

The last term of this formula, in accordance with equation (3.8), describes the field of Swihart waves trapped by a 6π-kink. A detailed description of "quantized" properties of a 6π-kink is given in Section 5.

Finally, we shall describe 8π-kinks. Let $\xi_1 = -\zeta_2$, $\xi_2 = -\zeta_1$, $\xi_3 = \zeta_1$, and $\xi_4 = \zeta_2$. Then the conditions (2.3) give the following general consistency relations for 8π-kinks:

$$\cos[k_0(v)\zeta_1] + \cos[k_0(v)\zeta_2] = 0, \quad (3.12)$$

$$f(\zeta_2 - \zeta_1) + f(\zeta_1 + \zeta_2) + f(2\zeta_2) = C, \quad (3.13)$$

$$f(2\zeta_1) + 2f(\zeta_1 + \zeta_2) + f(2\zeta_2) = 4C\cos^2[k_0(v)(\zeta_2 - \zeta_1)/2], \quad (3.14)$$

which determine the discrete ("quantized") spectrum of the characteristic velocities $v(n)$ of 8π-kinks, the wavelengths $\lambda_0(n) = 2\pi/k_0(n)$ of trapped Swihart waves, and the characteristic dimensions ζ_1 and ζ_2 of the trapping regions. Corresponding to the consistency condition (3.12) and equation (2.6), we now have for the solution of equation (2.1)

$$\psi_{8\pi}(\zeta) = \psi_{2\pi, m}(\zeta + \zeta_2) + \psi_{2\pi, m}(\zeta + \zeta_1) + \psi_{2\pi, m}(\zeta - \zeta_1) + \psi_{2\pi, m}(\zeta - \zeta_2) + \psi_w(\zeta, \zeta_1) + \psi_w(\zeta, \zeta_2). \quad (3.15)$$

In contrast to 4π and 6π-kinks, two oscillatory terms, describing generalized Swihart waves trapped by an 8π-kink and localized, respectively, in the regions Ξ_1 and Ξ_2, where the region Ξ_1 lies inside the region Ξ_2, have now appeared. The latter indicates the possibility of interference phenomena. According to equation (3.12), $(2n + 1)$ Swihart wavelengths fit into the sum of the intervals Ξ_1 and Ξ_2:

$$2n + 1 = (2\zeta_1 + 2\zeta_2)[k_0(v)/2\pi], \quad n = 0, 1, 2, \ldots. \quad (3.16)$$

We shall determine the second integer p, having in mind an interference pattern in the region Ξ_1, associated with overlapping of the regions of trapped Swihart waves. We take

$$p + \delta p = 2\zeta_1[k_0(v)/\pi] = 2\zeta_1[2/\lambda_0(v)], \quad (3.17)$$

where δp varies in the range $0 < \delta p < 1$. This value of p corresponds to the number of whole half-wavelengths $\lambda_0(v)/2$ which fit into the inner trapping region Ξ_1. Using the relations (3.16) and (3.17), the total field of trapped Swihart waves in the region Ξ_1 is given by the expression

$$4\pi C\sin[\pi(p + \delta p/2)]\sin[k_0(v)\zeta]. \quad (3.18)$$

This simple expression describes the result of interference of two trains of trapped Swihart waves moving together with an 8π-kink. For even values of p and δp close to zero or for odd p and δp close to 1, the fields of interfering Swihart waves are almost in antiphase and the quantity (3.18) is much less than $4\pi C$ in absolute magnitude. Conversely, for even p and δp close to 1 or for odd p and δp close to zero, the interfering waves are almost in-phase and the amplitude (3.18) is close to $4\pi C$. A detailed description of 8π-kinks is given in Section 6.

The results of this section show that in the general case of a traveling $2\pi n$-kink the phase difference $\psi(\zeta)$ (2.6) consists of the following sum of n monotonic functions of the type (3.7) and $s(n) = [2n - 1 + (-1)^n]/4$ oscillatory functions of the type (3.8), describing generalized Swihart waves Cherenkov-trapped by a moving vortex:

$$\psi(\zeta) = \phi_m(\zeta) + \phi_w(\zeta), \quad (3.19)$$

where

$$\phi_m(\zeta) \equiv \sum_{n' = 1}^{n} \{\pi + \pi[1 - f(\zeta - \xi_{n'})]\operatorname{sgn}(\zeta - \xi_{n'})\} \quad (3.20)$$

is the sum of the contributions of the components of the vortex structure of separate vortices, and

$$\phi_w(\zeta) \equiv \sum_{s = 1}^{s(n)} \psi_w(\zeta, \zeta_s) \quad (3.21)$$

corresponds to the collection of Swihart waves trapped by the vortex structure.

4. GLUING OF 4π-KINKS

We begin our detailed exposition of the results of the theory of the quantized motion of 4π-kinks with the case where the velocity of the kinks is close to the Swihart velocity:

$$\gamma^2 \equiv 1 - (v/v_s)^2 \ll 1. \quad (4.1)$$

Keeping in mind the condition (2.13), we obtain for the roots of equation (2.9) the expressions

$$k_0 = \frac{1}{\lambda}\left[\gamma^2 + \sqrt{\gamma^4 + \frac{2\lambda^2}{\lambda_j^2}}\right]^{1/2}, \quad k_1 = \frac{1}{\lambda}\left[-\gamma^2 + \sqrt{\gamma^4 + \frac{2\lambda^2}{\lambda_j^2}}\right]^{1/2}. \quad (4.2)$$

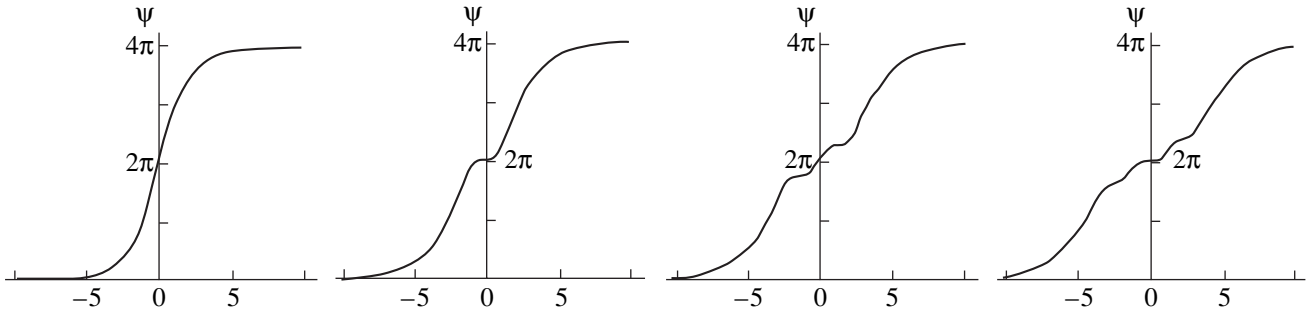

Fig. 1. Phase difference ψ of a fast 4π-kink as a function of the dimensionless coordinate $2^{1/4}\zeta/\sqrt{\lambda\lambda_j}$. The plots (from left to right) correspond to $n = 0, 1, 2, 3$.

Setting $\delta \equiv k_1/k_0$, we write down the relations

$$k_0 = \sqrt{\sqrt{2}/\delta\lambda\lambda_j}, \quad k_1 = \sqrt{\sqrt{2}\delta/\lambda\lambda_j}, \tag{4.3}$$

$$\frac{v^2}{v_s^2} = 1 - \frac{1-\delta^2}{\sqrt{2}\delta}\frac{\lambda}{\lambda_j}. \tag{4.4}$$

In accordance with equations (2.11) and (2.12), we also have

$$C = \delta^2/(1+\delta^2), \tag{4.5}$$

$$f(\zeta) = (1+\delta^2)^{-1}\exp(-\sqrt{\sqrt{2}\delta/\lambda\lambda_j}|\zeta|). \tag{4.6}$$

These relations and the consistency conditions for 4π-kinks (3.3) and (3.5) give the following equation for determining the parameter $\delta = \delta(n)$:

$$\delta(n) = \exp[-\pi(n+1/2)\delta(n)]. \tag{4.7}$$

Besides the formulas (4.3) and (4.4), which determine a discrete set of wave numbers and velocities of 4π-kinks, the parameter $\delta(n)$ also determines the discrete dimensions of the region of trapping of Swihart waves

$$2\zeta_0(n) = \frac{2\pi}{k_0}\left(n+\frac{1}{2}\right) = \frac{\pi}{2^{1/4}}\sqrt{\lambda\lambda_j\delta(n)}(2n+1). \tag{4.8}$$

For this reason, the solution of equation (4.7) in the case (4.1) describes all properties of 4π-kinks. For small numbers n of the vortex modes, from equation (4.7) we have $\delta(0) = 0.474$, $\delta(1) = 0.274$, $\delta(2) = 0.203$, and $\delta(3) = 0.164$. For high-order modes, for which $n \gg 1$, from equation (4.7) follows

$$\delta(n) \approx \pi^{-1}\left(n+\frac{1}{2}\right)^{-1}\ln\left[\pi\left(n+\frac{1}{2}\right)\right] \ll 1. \tag{4.9}$$

In accordance with the formula (4.4), the maximum value of the characteristic velocity of a 4π-kink is

$$v_{\max} = [1 - 0.58(\lambda/\lambda_j)]v_s, \tag{4.10}$$

which corresponds to the zeroth mode $n = 0$. For the high-order modes, $n \gg 1$, we have

$$v(n) = \left\{1 - \frac{\pi}{2^{5/2}}\frac{2n+1}{\ln[(2n+1)\pi/2]}\frac{\lambda}{\lambda_j}\right\}v_s. \tag{4.11}$$

It is obvious that the condition (4.1) imposes an upper limit on the number of high-order modes that corresponds to the smallness of the second term in the braces in equation (4.11) compared to 1. Figure 1 displays the spatial dependences of $\psi_{4\pi}(\zeta)$ for the first four 4π-kinks. As the mode number n of a vortex increases, the width of the region of nonmonotonic spatial dependence of a kink, due to the Cherenkov-trapped Swihart wave, also increases. At the same time, the degree of modulation decreases. On this basis, equation (3.6) can be written in the form

$$\begin{aligned}\psi_{4\pi}(\zeta) &= 2\pi + \pi\,\mathrm{sgn}[\zeta+\zeta_0(n)]\{1 - f(\zeta+\zeta_0(n), n)\} \\ &+ \pi\,\mathrm{sgn}[\zeta-\zeta_0(n)]\{1 - f(\zeta-\zeta_0(n), n)\} \\ &+ (-1)^n 2\pi\delta^2(n)[1+\delta^2(n)]^{-1}\sin\left[\frac{2^{1/4}\zeta}{\sqrt{\delta(n)\lambda\lambda_j}}\right] \\ &\times[\theta(\zeta+\zeta_0(n)) - \theta(\zeta-\zeta(n))],\end{aligned} \tag{4.12}$$

where in accordance with equations (4.1), (4.4), and (4.6)

$$f(\zeta, n) = [1+\delta^2(n)]^{-1} \times \exp\left\{-|\zeta|\Big/\left(\lambda_j\sqrt{\frac{\gamma^2(n)}{2} + \left[\frac{\gamma^4(n)}{4} + \frac{\lambda^2}{2\lambda_j^2}\right]^{1/2}}\right)\right\}. \tag{4.13}$$

In equation (36) of [15] the first factor in the last expression is dropped as an approximation. This corresponds to the approximation (4.9).

In complete analogy to 2π-kinks of the sine-Gordon model [17], the spatial size of our vortices is much smaller than the Josephson length:

$$\lambda_j\sqrt{\frac{\gamma^2(n)}{2}+\left[\frac{\gamma^4(n)}{4}+\frac{\lambda^2}{2\lambda_j^2}\right]^{1/2}} \ll \lambda_j. \qquad (4.14)$$

This similarity is literal for sufficiently high modes, when $\lambda/\lambda_j \ll \gamma^2(n) \ll 1$ and the left-hand side of the inequality (4.14) approximately assumes the standard for Josephson electrodynamics form

$$\lambda_j\gamma(n) = \lambda_j\sqrt{1-v^2(n)/v_s^2} \equiv \tilde{\lambda}_j$$

of the Lorentz-contracted Josephson length. The size of the region of Cherenkov trapping of waves is $\ln(\pi^2 n^2) \gg 1$ times greater than $\tilde{\lambda}_j$.

We now turn to the asymptotic limit of low velocities,

$$v \ll v_s, \qquad (4.15)$$

which obtains for very high-order vortex modes. Then we obtain from equation (2.9)

$$k_0 = \lambda^{-1}(v_s/v)^2 \gg \lambda^{-1}, \quad k_1 = \lambda_j^{-1} \qquad (4.16)$$

and for the expressions (2.11) and (2.12) we have

$$C = 2(\lambda/\lambda_j)^2(v/v_s)^2, \qquad (4.17)$$

$$f(\zeta) = \exp(-|\zeta|/\lambda_j). \qquad (4.18)$$

As a result, on the basis of the consistency conditions (3.3) and (3.5), we obtain the following expressions for the size of the region of Cherenkov trapping of a Swihart wave:

$$2\zeta_0(n) = \lambda_j A(n) \equiv \lambda_j \ln\frac{\pi n\lambda_j}{\lambda} \gg \lambda_j, \qquad (4.19)$$

for the quantum velocity of a 4π-kink:

$$v(n) = v_s\sqrt{\lambda_j/\lambda}\sqrt{A(n)/2\pi n} \ll v_s, \qquad (4.20)$$

and for the wave number of a trapped wave:

$$k_0(n) = 2\pi n/\lambda_j A(n) \gg \lambda^{-1}. \qquad (4.21)$$

Corresponding to these formulas

$$\psi_{4\pi}(\zeta) = \varphi_{2\pi}(\zeta+\zeta_0(n)) + \varphi_{2\pi}(\zeta-\zeta_0(n)) + (-1)^n\frac{4\lambda A(n)}{(2n+1)\lambda_j}\sin\left(\frac{2\pi n\zeta}{\lambda_j A(n)}\right) \times[\theta(\zeta+\zeta_0(n)) - \theta(\zeta-\zeta_0(n))], \qquad (4.22)$$

where the function $\varphi_{2\pi}$ is determined by the expression (1.5) with $v = 0$. The characteristic scale of the spatial variation of a slow 4π-kink corresponds to the ordinary

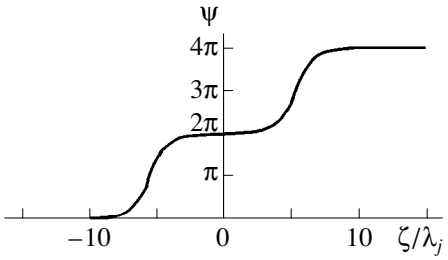

Fig. 2. Gluing of a slow 4π-kink from two 2π-kinks; $\lambda_j/\lambda = 10$, $n = 1600$. The velocity of the vortex is $v \simeq 0.1v_s$, and the distance between the 2π-kinks is $2\zeta_0 \simeq 10.8\lambda_j$.

Josephson length. The scale of oscillations that corresponds to a trapped Swihart wave is small compared to the London length. The amplitude of the trapped wave is also very small. However, only its existence allows the existence of the countable set, which we have obtained, of moving 4π-kinks with a condensation point as $v(n) \longrightarrow 0$. In the formula (4.22) the functions $\varphi_{2\pi}$ describe 2π-kinks separated by a distance $2\zeta_0(n) \gg \lambda_j$. The small-amplitude Cherenkov oscillations, which are described by the last term in equation (4.22), glue these 2π-kinks into a single moving 4π-kink. The cooling phenomenon is illustrated in Fig. 2.

We note that the phase difference (4.12), corresponding to a fast 4π-kink, can also be represented as a sum of two monotonic 2π-kinks,

$$\pi + \pi\,\mathrm{sgn}(\zeta \pm \zeta_0(n))\{1 - f(\zeta \pm \zeta_0(n), n)\}, \qquad (4.23)$$

which are combined into a single 4π-kink and are Cherenkov trapped Swihart waves, whose field is described by the last term in equation (4.12). In this case there is no sense in talking about gluing of 2π-kinks for $n \gg 1$, when the amplitude of the gluing Swihart waves is relatively small.

Using the relations (4.3)–(4.5) and (4.8), we find from equations (A.1), (A.4), and (A.5) the magnetic field of a fast 4π-kink moving with velocity close to the Swihart velocity,

$$H_{4\pi}(x,\zeta) = -\frac{\phi_0}{2^{7/4}\lambda\sqrt{\lambda\lambda_j}}\frac{\sqrt{\delta}}{1+\delta^2}\exp\left(-\frac{|x\mp d|}{\lambda}\right) \times\left\{\exp\left(-2^{1/4}\sqrt{\delta}\frac{|\zeta-\zeta_0|}{\sqrt{\lambda\lambda_j}}\right) + \exp\left(-2^{1/4}\sqrt{\delta}\frac{|\zeta+\zeta_0|}{\sqrt{\lambda\lambda_j}}\right) + 2\delta(-1)^n[\theta(\zeta+\zeta_0) - \theta(\zeta-\zeta_0)]\cos\left(\frac{2^{1/4}|\zeta|}{\sqrt{\delta\lambda\lambda_j}}\right)\right\}. \qquad (4.24)$$

According to equation (4.24) the magnetic field of a fast 4π-kink is concentrated near the plane of the Josephson junction at distances of the order of the London length. The region of localization of the magnetic field is much wider along the propagation axis of a vortex. The monotonic component of the field is localized

on the sections $\pm\zeta_0 - 1/k_0\delta < \zeta < \pm\zeta_0 + 1/k_0\delta$, where $\zeta_0 = 2^{-1/4}\sqrt{\delta\lambda\lambda_j}\,\pi(n + 1/2)$. On the interval $-\zeta_0 < \zeta < \zeta_0$ the magnetic field oscillates with period $2\pi/k_0 = 2^{3/4}\pi\sqrt{\delta\lambda\lambda_j}$, and the amplitude of the oscillations is 2δ times smaller than the maximum value of the monotonic component. We also note that as the mode numbers increase, the maximum value of the magnetic field decreases as $\propto\sqrt{\delta} \approx [(\ln\pi n)/\pi n]^{1/2}$.

The limit $v \ll v_s$, using the relations (4.16), (4.17), and (4.20) we obtain for the magnetic field

$$\begin{aligned} H_{4\pi}(x,\zeta) &= -\frac{\phi_0}{4\lambda\lambda_j}\exp\left(-\frac{|x\mp d|}{\lambda}\right) \\ &\times\left\{\exp\left(-\frac{|\zeta-\zeta_0|}{\lambda_j}\right)+\exp\left(-\frac{|\zeta+\zeta_0|}{\lambda_j}\right)\right\} \\ &-(-1)^n\left(\frac{\phi_0}{k_0\lambda\lambda_j^2}\right)\exp(-k_0|x\mp d|) \\ &\times[\theta(\zeta+\zeta_0)-\theta(\zeta-\zeta_0)]\cos(k_0|\zeta|), \end{aligned} \tag{4.25}$$

where $k_0 = \lambda_j^{-1}[2\pi n/\ln(\pi n\lambda_j/\lambda)] \gg 1/\lambda$. The maximum values of the field (4.25) reached at $\zeta \approx \pm\zeta_0$ are less than for the fast vortices and do not depend on the mode number. In contrast to equation (4.24), the oscillating part of the field of the slow vortices is confined within a distance $\approx 1/k_0$ close to the plane of the Josephson junction which is much less than λ. The amplitude of the oscillations is $k_0\lambda_j/4 \gg 1$ times less than the maximum value of the monotonic component.

Let us now compare the magnetic field (4.24) of a slow 4π-kink inside the junction with the field of the ordinary ($\lambda \ll \lambda_j$) 2π-kink (1.6) with $v = 0$. The expression (4.24) shows that the monotonic component of the magnetic field of a 4π-kink is the field of two ordinary 2π-kinks, which, according to equation (4.19), are separated by a distance $2\zeta_0$ much greater than the scale λ_j of the spatial variation of the 2π-kinks themselves. In the region from $\zeta = -\zeta_0$ to $\zeta = \zeta_0$ the magnetic fields of 2π-kinks are "joined" by a function oscillating with period $2\pi/k_0 \ll 2\zeta_0$ and a low amplitude, describing the magnetic field produced by the trapped Swihart waves, which glue two 2π-kinks into a single 4π-kink.

We shall now discuss the energy of 4π-kinks. We obtain from equations (A.7)–(A.9)

$$\begin{aligned} W_m &= 2(1-C)^2 k_1\lambda_j \\ &\times[1+(1+2k_1\zeta_0)\exp(-2k_1\zeta_0)]W_{2\pi}, \end{aligned} \tag{4.26}$$

$$W_{mw} = 4C(1-C)(1+v^2/v_s^2)k_1\lambda_j W_{2\pi}, \tag{4.27}$$

$$W_w = 2C^2\pi(2n+1)(v/v_s)^2 k_0\lambda_j W_{2\pi}, \tag{4.28}$$

where $W_{2\pi}$ is the energy of a 2π-kink at rest with $\lambda \ll \lambda_j$, and C is given by the formulas (4.5) or (4.17). Specifically, for the main mode with $n = 0$ we find from equations (4.26)–(4.28)

$$W = 3.05\sqrt{\lambda_j/\lambda}\,W_{2\pi},$$

$$W_m = 1.7\sqrt{\lambda_j/\lambda}\,W_{2\pi}, \quad W_{mw} = 0.98\sqrt{\lambda_j/\lambda}\,W_{2\pi},$$

$$W_w = 0.37\sqrt{\lambda_j/\lambda}\,W_{2\pi}.$$

For $n \gg 1$, when the velocity of a vortex is described by the expression (4.11), the energy of the vortex is determined primarily by W_m:

$$\begin{aligned} W \simeq W_m &= 2\left(\sqrt{2}\frac{\ln(\pi n)}{\pi n}\right)^{1/2}\sqrt{\frac{\lambda_j}{\lambda}}W_{2\pi} \\ &= 2W_{2\pi}\Big/\sqrt{1-\left(\frac{v}{v_s}\right)^2} \gg 2W_{2\pi} \end{aligned} \tag{4.29}$$

and we have for the energy of the gluing waves from equation (4.28)

$$W_w \simeq \frac{2\ln^3(\pi n)}{(\pi n)^2}2W_{2\pi}\Big/\sqrt{1-\left(\frac{v}{v_s}\right)^2} \ll W_m. \tag{4.30}$$

The smallness of the energy W_w compared with W_m corresponds to the assertion that the kinks are glued together by Swihart waves. The gluing energy of the kinks is even smaller:

$$\begin{aligned} W_{mw} &\simeq \left(\frac{2\ln(\pi n)}{\pi n}\right)^2 2W_{2\pi}\Big/\sqrt{1-\left(\frac{v}{v_s}\right)^2} \\ &\ll W_w \ll W_m. \end{aligned} \tag{4.31}$$

The expressions (4.29)–(4.31) are valid for vortices with velocities close to the Swihart velocity. This means that the large mode number n cannot exceed the value $(\sqrt{2}\,\lambda_j/\pi\lambda)\ln(\sqrt{2}\,\lambda_j/\lambda) \gg 1$. For modes with even larger numbers, when n satisfies the inequality

$$n \gg \frac{\sqrt{2}\lambda_j}{\pi\lambda}\ln\left(\frac{\sqrt{2}\lambda_j}{\lambda}\right) \gg 1,$$

the total energy W of a vortex is also determined by the sum of the energies of the components of its 2π-kinks:

$$W \simeq W_m \simeq 2W_{2\pi}, \tag{4.32}$$

and the energies of the waves and the gluing energy are, once again, small:

$$\begin{aligned} W_w &\simeq \left[2\lambda\ln^2\frac{\pi n\lambda_j}{\lambda}/(\pi n\lambda_j)\right]2W_{2\pi} \ll W_m, \\ W_{mw} &\simeq \left[2\lambda\ln\frac{\pi n\lambda_j}{\lambda}/(\pi n\lambda_j)\right]2W_{2\pi} \ll W_w \ll W_m. \end{aligned} \tag{4.33}$$

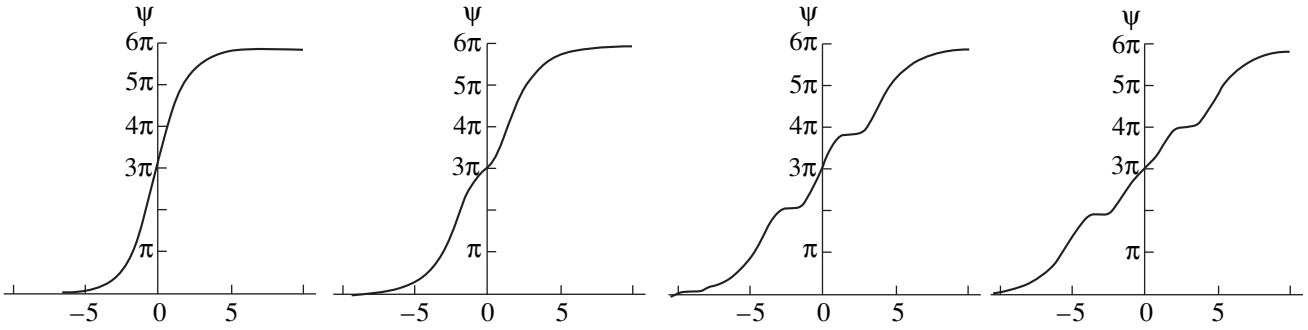

Fig. 3. The phase difference ψ of a fast 6π-kink as a function of the dimensionless coordinate $2^{1/4}\zeta/\sqrt{\lambda\lambda_j}$. The plots (from left to right) correspond to $n = 0, 1, 2, 3$.

The formula (4.32) corresponds to the formula (4.29) in the limit (4.15). Even though the energies (4.33) are small compared to (4.32), only the presence of waves leads to the cooling phenomenon, and therefore to the existence of a 4π-kink. This also pertains to the relations (4.29)–(4.31).

5. GLUING OF 6π-KINKS

We shall make use of the definite similarity between 6π- and 4π-kinks. The relations (4.1)–(4.6) hold for 6π-kinks moving with velocity close to the Swihart velocity. The following new equation for the parameter $\delta(n)$ arises from the consistency equations:

$$\delta^2(n) = \exp\left\{-\pi\delta(n)\left[n + 0.5 + \frac{(-1)^n}{6}\right]\right\} + \exp\left\{-2\pi\delta(n)\left[n + 0.5 + \frac{(-1)^n}{6}\right]\right\}. \quad (5.1)$$

We have for the quantized size of the region of Cherenkov trapping of waves

$$2\zeta_0(n) = 2^{3/4}\pi[n + 0.5 + (-1)^n/6]\sqrt{\delta(n)\lambda\lambda_j}.$$

For the first few modes of a 6π-kink we find from equation (5.1) $\delta(0) = 0.603$, $\delta(1) = 0.434$, $\delta(2) = 0.298$, $\delta(3) = 0.262$, and $\delta(4) = 0.213$. Correspondingly, the highest velocity of a 6π-kink corresponds to $n = 0$ and is

$$v_{\max}(0) = \left[1 - 0.37\frac{\lambda}{\lambda_j}\right]v_s, \quad (5.2)$$

which is somewhat higher than the highest velocity of a 4π-kink (4.10). For $n \gg 1$, but under the condition (4.1), we have

$$\delta(n) = \frac{2}{\pi}\left[n + 0.5 + \frac{(-1)^n}{6}\right]^{-1} \times \ln\left\{\frac{\pi}{2}\left[n + 0.5 - \frac{(-1)^n}{6}\right]\right\} \ll 1.$$

The explicit form of the solution of equation (2.1) is

$$\psi_{6\pi}(\zeta) = 3\pi + \pi\,\mathrm{sgn}[\zeta + \zeta_0(n)] \times \{1 - f(\zeta + \zeta_0(n), n)\} + \pi\,\mathrm{sgn}\,\zeta\{1 - f(\zeta, n)\} + \pi\,\mathrm{sgn}[\zeta - \zeta_0(n)]\{1 - f(\zeta - \zeta_0(n), n)\} + [\theta(\zeta + \zeta_0(n)) - \theta(\zeta - \zeta_0(n))]\pi\delta^2(n) \times (1 + \delta^2(n))^{-1}\{(-1)^n\sqrt{3}\sin[2^{1/4}\zeta/\sqrt{\delta(n)\lambda\lambda_j}] - \cos[2^{1/4}\zeta/\sqrt{\delta(n)\lambda\lambda_j}]\,\mathrm{sgn}\,\zeta\}, \quad (5.3)$$

where $f(\zeta, n)$ is given by the expression (4.13). Figure 3 illustrates the spatial dependence of a moving 6π-kink for the first four modes, for which the roots of equation (5.1) can be found numerically.

Slow 6π-kinks, moving with velocities satisfying the condition (4.15), correspond to high-order modes of vortices, for which

$$n \gg \frac{2^{3/2}\lambda_j}{\pi\lambda}\ln\frac{\sqrt{2}\lambda_j}{\lambda} \gg 1. \quad (5.4)$$

We obtain for the width of the region of Cherenkov trapping of waves, the velocity of a slow 6π-kink, and the wave number of the trapped generalized Swihart wave, respectively,

$$2\zeta_0(n) = 2\lambda_j\ln\frac{\pi n\lambda_j}{2\lambda} \gg 2\lambda_j, \quad (5.5)$$

$$v(n) = v_s\sqrt{\frac{\lambda_j}{\pi n\lambda}\ln\frac{\pi n\lambda_j}{2\lambda}} \ll v_s, \quad (5.6)$$

$$k_0(n) = \frac{\pi n}{\lambda_j}\ln^{-1}\frac{\pi n\lambda_j}{2\lambda} \gg \frac{1}{\lambda}. \quad (5.7)$$

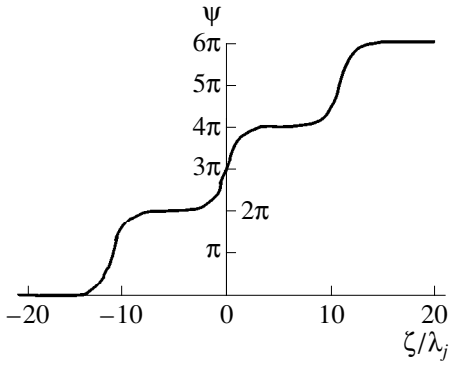

Fig. 4. Gluing of a slow 6π-kink from three 2π-kinks; $\lambda_j/\lambda = 10$, $n = 3200$. The velocity of the vortex is $v \simeq 0.1 v_s$, and the distance between the 2π-kinks is $\zeta_0 \simeq 10.8\lambda_j$.

The solutions of equation (2.1) for slow 6π-kinks can be written in the explicit form

$$\psi_{6\pi}(\zeta) = \varphi_{2\pi}(\zeta + \zeta_0(n)) + \varphi_{2\pi}(\zeta) + \varphi_{2\pi}(\zeta - \zeta_0(n)) + \frac{2\lambda}{n\lambda_j}\ln\frac{\pi n\lambda_j}{2\lambda}\{(-1)^n\sqrt{3}\sin[k_0(n)\zeta] - \cos[k_0(n)\zeta]\operatorname{sgn}\zeta\}[\theta(\zeta + \zeta_0(n)) - \theta(\zeta - \zeta_0(n))], \quad (5.8)$$

where $\varphi_{2\pi}$ is determined by the formula (1.5) with $v = 0$. The last oscillating term of the formula (5.8) is comparatively very small in magnitude. However, the field of the Cherenkov-trapped Swihart waves which is described by this term makes possible the existence of a moving 6π-kink. Just as in the case of a 4π-kink, quantized discrete states of 6π-kinks possess a condensation point as $n \longrightarrow \infty$. According to this formula and equation (5.5) with ζ varying from left to right, $\psi_{6\pi}$ increases by 2π starting approximately at a distance λ_j near $\zeta = -\zeta_0$. Then, at a distance approximately equal to ζ_0 and much greater than the Josephson length λ_j, the phase difference changes very little. Further, near $\zeta = 0$, once again, $\psi_{6\pi}$ increases by 2π at a distance of a approximately λ_j. Once again, equation (5.8) at a distance of about ζ_0 corresponds to a small change in $\psi_{6\pi}$. Finally, a last increase by 2π occurs near $\zeta = \zeta_0$ approximately at a distance λ_j. Figure 4 illustrates the dependence described. On this basis an interesting observation can be made, similar to that made in Section 4 for a 4π-kink trapping a large number of Swihart waves. Specifically, a 6π-kink in the case of slow motion of a multivortex is clearly separated into three 2π-kinks, moving synchronously with a definite velocity (5.6). At the same time, Swihart waves with comparatively low intensities, which are trapped by a 6π-kink and move together with it, make possible the existence of this kink. Being trapped by the vortices, the Swihart waves glue together three 2π-kinks into a single 6π-kink. Thus, once again, we can talk about Swihart waves as a kind of glue for Josephson vortices.

We shall now consider the magnetic field and the energy of a 6π-kink. For a vortex moving with a velocity close to the Swihart velocity, we obtain from equations (A.1), (A.4), (A.5), and (5.3)

$$H_{6\pi}(x, \zeta) = -\frac{\phi_0}{2^{7/4}\lambda\sqrt{\lambda\lambda_j}}\frac{\sqrt{\delta}}{1+\delta^2}\exp\left(-\frac{|x \mp d|}{\lambda}\right) \times\left\{\exp\left(-2^{1/4}\sqrt{\delta}\frac{|\zeta - \zeta_0|}{\sqrt{\lambda\lambda_j}}\right) + \exp\left(-2^{1/4}\sqrt{\delta}\frac{|\zeta|}{\sqrt{\lambda\lambda_j}}\right) + \exp\left(-2^{1/4}\sqrt{\delta}\frac{|\zeta + \zeta_0|}{\sqrt{\lambda\lambda_j}}\right) + 2\delta(-1)^n[\theta(\zeta + \zeta_0) - \theta(\zeta - \zeta_0)] \times\cos\left(\frac{2^{1/4}|\zeta|}{\sqrt{\delta\lambda\lambda_j}} - \frac{(-1)^n\pi}{6}\right)\right\}. \quad (5.9)$$

Just as in the case of a fast 4π-kink, the field (5.9) is localized near the plane of the Josephson junction on a scale of the order of the London length, while the characteristic scale of variation of the monotonic component is approximately $2^{-1/4}\sqrt{\lambda\lambda_j/\delta} \gg \lambda$. The field of Cherenkov-trapped waves is different from zero on an interval Ξ_0 and oscillates with period $2^{3/4}\pi\sqrt{\delta\lambda\lambda_j} \gg \lambda$. As the mode number increases (i.e., the vortex velocity increases), the maximum value of the magnetic field decreases.

For slow 6π-kinks, when the inequality (5.4) is satisfied, we obtain from equations (A.1), (A.4), (A.5), and (5.7)

$$H_{6\pi}(x, \zeta) = -\frac{\phi_0}{4\lambda\lambda_j}\exp\left(-\frac{|x \mp d|}{\lambda}\right) \times\left\{\exp\left(-\frac{|\zeta - \zeta_0|}{\lambda_j}\right) + \exp\left(-\frac{|\zeta|}{\lambda_j}\right) + \exp\left(-\frac{|\zeta + \zeta_0|}{\lambda_j}\right)\right\} - (-1)^n\frac{\phi_0}{k_0\lambda\lambda_j^2}\exp(-k_0|x \mp d|) + [\theta(\zeta + \zeta_0) - \theta(\zeta - \zeta_0)]\cos\left(k_0|\zeta| - (-1)^n\frac{\pi}{6}\right), \quad (5.10)$$

where $k_0 = (\pi n/\lambda_j)\ln^{-1}(\pi n\lambda_j/2\lambda) \gg 1/\lambda$. Similarly to a 4π-kink, the monotonic component of the field of a slow 6π-kink penetrates into superconductors to the London length, and the contribution of the Cherenkov-trapped Swihart waves to the magnetic field is localized near the plane of the junction on a smaller scale, $1/k_0 \ll \lambda$.

The expression (5.10) demonstrates the assertion, made after equation (5.8), about the gluing of vortices. Indeed, the monotonic part of the field (5.10) in the

junction can be represented as a sum $H_{2\pi}(\zeta + \zeta_0) + H_{2\pi}(\zeta) + H_{2\pi}(\zeta - \zeta_0)$, where $H_{2\pi}$ is described by the expression (1.6) with $v = 0$. This means that the field of a 6π-kink can be represented as a superposition of the fields of three comparatively widely separated 2π-kinks. The trapped Swihart waves, having low intensities, glue these three fields into a single field, correspondingly to a slowly moving 6π-kink.

We shall now demonstrate the effect of the gluing on the energy of a 6π-kink in the limit of large n, when a vortex has trapped a large number of wavelengths. Just as for a 4π-kink, in this limit the gluing energy W_{mw} is small compared to W_m and compared to the energy W_w of the waves. For $n \gg 1$ we find from equation (5.3) and the definitions (3.19)–(3.21) and (A.7)–(A.9)

$$W_m \simeq 3k_1\lambda_j W_{2\pi} \simeq 3W_{2\pi}/\sqrt{1-(v/v_s)^2},$$

$$W_w \simeq 4\pi n C^2\left(\frac{v}{v_s}\right)^2 k_0\lambda_j W_{2\pi}$$

$$\simeq \left[\frac{8}{3}\left(\frac{2}{\pi n}\right)^2 \ln^3\frac{\pi n}{2}\right] 2W_{2\pi}/\sqrt{1-\left(\frac{v}{v_s}\right)^2} \ll W_m.$$

We can see that the energy of a vortex is determined primarily by the energy of three 2π-kinks (1.7) and, correspondingly, the energy W_w of the gluing waves is low.

For a vortex with velocity $v \ll v_s$, when the inequality (5.4) holds, we have

$$W_m \simeq 3W_{2\pi},$$

$$W_w \simeq \left[\frac{8}{3}\frac{2\lambda}{\pi n\lambda_j}\ln^2\frac{\pi n\lambda_j}{2\lambda}\right] 3W_{2\pi} \ll W_m.$$

The last expressions show that, to a high degree of accuracy, the energy of a 6π-kink consists of the energy of three 2π-kinks. The energy of the Swihart waves that glue these 2π-kinks together is much less than W_m, which corresponds to the idea of gluing of Josephson vortices by the trapped waves.

The properties established here and in Section 4 are manifested just as strikingly for Josephson vortices corresponding to 8π-kinks. This will be examined in the next section.

6. GLUING OF 8π-KINKS

In Section 3 it was indicated that the structure of the field of Swihart waves Cherenkov-trapped by an 8π-kink is more complicated than that of the field of trapped waves in the case of 4π and 6π-kinks. Indeed, now, besides the number n characterizing the number of wavelengths of the trapped waves in the sum of the intervals Ξ_1 and Ξ_2 and corresponding to equation (3.16), another integer p, determined by the relation (3.17) and corresponding to the interference of waves in the interval Ξ_1, plays an essential role. Using the parameter δ, which was introduced in Section 4 and which for 8π-kinks is a function of the numbers n and $p(\delta(n, p))$, we can write the consistency conditions (3.13) and (3.14) in the form

$$\begin{aligned} &\exp[-k_0(\zeta_2-\zeta_1)\delta] + \exp[-k_0(\zeta_1+\zeta_2)\delta] \\ &\quad + \exp[-2k_0\zeta_2\delta] = \delta^2, \\ &\exp[-2k_0\zeta_1\delta] + 2\exp[-k_0(\zeta_1+\zeta_2)\delta] \\ &\quad + \exp[-2k_0\zeta_2\delta] = 4\delta^2\cos^2(k_0(\zeta_2-\zeta_1)/2). \end{aligned} \tag{6.1}$$

Introducing the notation $a \equiv \pi(2n+1)\delta$, we can write, using equations (3.16) and (6.1), the following equation for the parameter δ:

$$\frac{e^a(1+\delta^2)^2}{(1+e^a)(\delta^2e^a-1)} = 4\delta^2\cos^2\left[\frac{1}{2\delta}\ln\left(\frac{\delta^2e^a-1}{1+e^a}\right)\right]. \tag{6.2}$$

For the intervals Ξ_1 and Ξ_2 we have

$$2\zeta_1 = \sqrt{\frac{\lambda\lambda_j}{\delta\sqrt{2}}}\ln\left(\frac{\delta^2e^a-1}{e^{-a}+1}\right),$$

$$2\zeta_2 = \sqrt{\frac{\lambda\lambda_j}{\delta\sqrt{2}}}\ln\left(\frac{e^a+1}{\delta^2-e^{-a}}\right).$$

Just as in the two preceding sections, for small values of n corresponding to a very small difference of the characteristic velocity of an 8π-kink from the Swihart velocity, equation (6.2) together with the determination of a can be solved numerically. The values obtained for $\delta(n, p)$ and for the quantity $\delta p(n, p)$, characterizing the interference pattern, determined according to equation (3.17) are presented in the table. According to equation (4.4), $\delta(n, p)$ determines the characteristic velocity of an 8π-kink. The 8π-kink, which depends monotonically on the argument ζ and does not oscillate, corresponds to $n = p = 0$. The maximum characteristic velocity of uniform motion of such an 8π-kink is

$$v_{\max} = v_s(1 - 0.29\lambda/\lambda_j),$$

which is higher than the maximum velocities of a 4π-kink (4.10) and a 6π-kink (5.2), and for the regions of Cherenkov trapping of waves we have

$$2\zeta_1(0, 0) = 0.91\sqrt{\lambda\lambda_j}, \quad 2\zeta_2(0, 0) = 3.42\sqrt{\lambda\lambda_j}.$$

Only one wavelength $\lambda(0, 0) = 2\pi/k_0(0, 0) = 4.33\sqrt{\lambda\lambda_j}$ fits into the segment $2\zeta_1 + 2\zeta_2$. This wavelength is greater than the wavelength of the trapped waves in 8π-kinks with $n > 0$. Figure 5 shows plots of 8π-kinks for the smallest numbers $n = 0$ ($p = 0$) and $n = 1$ ($p = 0, 1, 2$).

We shall now consider high-order modes of vortices

$$1 \ll n \ll \frac{2^{3/2}\lambda_j}{\pi\lambda}\ln\frac{\sqrt{2}\lambda_j}{\lambda},$$

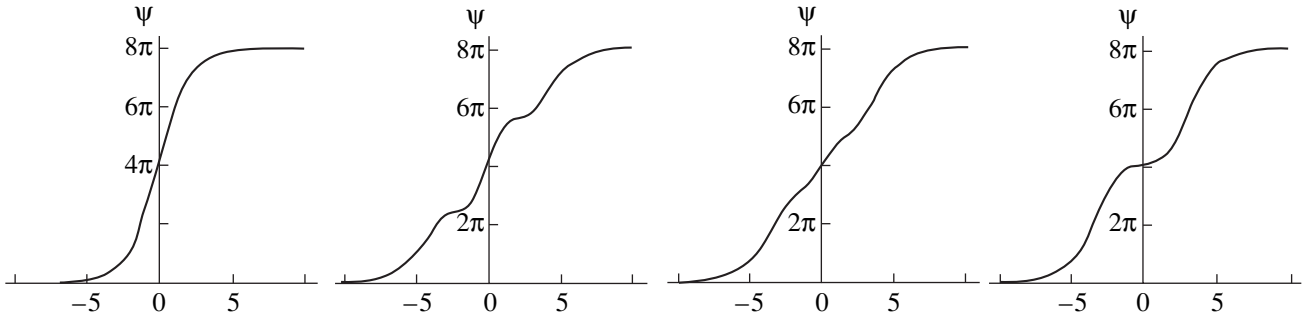

Fig. 5. Phase difference ψ of a fast 8π-kink as a function of the dimensionless coordinate $2^{1/4}\zeta/\sqrt{\lambda\lambda_j}$. The plots (from left to right) correspond to $n = 0$, $p = 0$; $n = 1$, $p = 0$; $n = 1$, $p = 1$; $n = 1$, $p = 2$.

where on account of the right-hand inequality the characteristic velocity of an 8π-kink is still essentially the same as the Swihart velocity. Then equations (6.1) can be reduced to the following:

$$\exp[-k_0(v)(\zeta_2 - \zeta_1)\delta] = \delta^2,$$

$$\exp[-2k_0(v)\zeta_1\delta] = 4\delta^2\cos^2[k_0(v)(\zeta_2 - \zeta_1)/2].$$

Hence, using the relations (3.16), (3.17), and (4.3) we obtain

$$2\zeta_1 = 2^{-1/4}\pi\sqrt{\lambda\lambda_j}\delta(p + \delta p), \tag{6.3}$$

$$2\zeta_2 = 2^{-1/4}\pi\sqrt{\lambda\lambda_j}\delta(4n + 2 - p - \delta p), \tag{6.4}$$

$$\delta = 2A/\pi \tag{6.5}$$

and the following equation for $\delta p(n, p)$:

$$\exp[-(2n + 1 - p - \delta p)A] = 2A/\pi, \tag{6.6}$$

where $A \equiv (2n + 1 - 2p - 2\delta p)^{-1}\ln|2\sin(\pi(p + \delta p)/2)|$.

A particular case of such an interference of trapped waves on the section Ξ_1, when the amplitude of the oscillatory dependence (3.18) is much less than $4\pi C$, corresponds to the solution of equation (6.6) with

$$n + \frac{n}{2}\ln^{-1}\frac{\pi n}{2} < p < 2n - 3,$$

when

$$\delta p = \frac{1 - (-1)^p}{2} + (-1)^p\pi^{-1}\exp\left\{-N\left[p - n - \frac{(-1)^p}{2}\right]\right\}, \tag{6.7}$$

where

$$N = N(n, p) \equiv 4[4n - 2p + 1 + (-1)^p]^{-1} \times \ln\left\{\pi\left(n - \frac{p}{2}\right) + \frac{\pi}{4}[1 + (-1)^p]\right\}.$$

The formulas (4.3), (4.4), (6.3)–(6.5), and (6.7) make it possible to obtain the following discrete spectrum characterizing the quantities for fast 8π-kinks:

$$\frac{v(n, p)}{v_s} = 1 - 2^{-3/2}\pi N^{-1}(n, p)\frac{\lambda}{\lambda_j},$$

$$2\zeta_1(n, p) = 2^{-1/4}p\sqrt{\pi N\lambda\lambda_j},$$

Table

p	n							
	0		1		2		3	
	$\delta(n, p)$	$\delta p(n, p)$	$\delta(n, p)$	$\delta p(n, p)$	$\delta(n, p)$	$\delta p(n, p)$	$\delta(n, p)$	$\delta p(n, p)$
0	0.672	0.421	0.355	0.818				
1			0.439	0.735				
2			0.511	0.139	0.303	0.379	0.228	0.672
3					0.466	0.954	0.263	0.708
4					0.483	0.039	0.286	0.186
5							0.472	0.987
6							0.476	0.007

$$2\zeta_2(n, p) = 2^{-1/4}(4n + 2 - p)\sqrt{\pi N\lambda\lambda_j},$$

$$\lambda_0(n, p) = \frac{2\pi}{k_0(n, p)} = 2^{3/4}\sqrt{\pi N\lambda\lambda_j}.$$

As the number p increases, the size of the region Ξ_1 increases and the size of the region Ξ_2 decreases. In other words, the size of the region of such interference of trapped waves, where the waves compensate one another, becomes larger. At the same time, for a fixed value of n, the characteristic velocity of an 8π-kink, which is essentially the same as the Swihart velocity, increases.

The equation (6.6) holds in the limit of large values of n and not too large values of p, when only the first few terms need be retained, approximately, on the left-hand sides of equations (6.1). Conversely, for p close to $2n$ the size $2\zeta_1$ of the inner trapping region is close to $2\zeta_2$ and the approximation (6.6) of equations (6.1) is not accurate enough. For this reason, for such values of p we shall find the values of $\delta(n, p)$ and $\delta p(n, p)$ directly from equation (6.1), after which we shall determine the parameters characterizing an 8π-kink. Thus, for $p = 2n$, $\delta p \simeq 0$ and $p = 2n - 1$, $\delta p \simeq 1$ we find: $\delta \simeq 0.474$, $2\zeta_1 \simeq 3.64n\sqrt{\lambda\lambda_j}$, $2\zeta_2 \simeq 3.64(n+1)\sqrt{\lambda\lambda_j}$, $v/v_s \simeq 1 - 0.58(\lambda/\lambda_j)$, and $\lambda_0 = 2\pi/k_0 \simeq 3.64\sqrt{\lambda\lambda_j}$. Correspondingly, for $p = 2n - 2$, $\delta p \simeq 0$ and $p = 2n - 3$, $\delta p \simeq 1$ we find $\delta \simeq 0.274$, $2\zeta_1 \simeq 2.77(n-1)\sqrt{\lambda\lambda_j}$, $2\zeta_2 \simeq 3.64(n+2)\sqrt{\lambda\lambda_j}$, $v/v_s \simeq 1 - 1.19(\lambda/\lambda_j)$, and $\lambda_0 = 2\pi/k_0 \simeq 2.77\sqrt{\lambda\lambda_j}$.

As p decreases further, the asymptotic equation (6.6) is applicable to a high degree of accuracy. For example, even for $p = 2n - 4$ the accuracy of the solution (6.7) of this equation is 3%.

Finally, we shall consider slow 8π-kinks, whose mode numbers satisfy the condition (5.4). Then the relations (4.16)–(4.18) are satisfied, and the consistency equations (3.13) and (3.14) reduce to the following:

$$\exp\left(\frac{\zeta_1 - \zeta_2}{\lambda_j}\right) = \frac{2\lambda}{k_0\lambda_j^2},$$
$$\exp\left(-\frac{2\zeta_1}{\lambda_j}\right) = \frac{8\lambda}{k_0\lambda_j^2}\cos^2\left[\frac{k_0(\zeta_2 - \zeta_1)}{2}\right]. \quad (6.8)$$

Using equation (3.17), according to equation (6.8), the characteristic parameters of a vortex can be represented in the form

$$2\zeta_1 = \pi(p + \delta p)/k_0, \quad 2\zeta_2 = \pi(4n + 2 - p - \delta p)/k_0, \quad (6.9)$$

$$v = v_s/\sqrt{k_0\lambda} \ll v_s,$$

$$\lambda_0 = 2\pi/k_0 = 4\lambda_j A \ll \lambda. \quad (6.10)$$

The unknown $\delta p(n, p)$ is determined by the equation

$$\exp\{-2(2n + 1 - p - \delta p)A\} = \frac{4\lambda A}{\pi\lambda_j}. \quad (6.11)$$

The presence of the small parameter (λ/λ_j) on the right-hand side and the fact that this parameter is positive immediately indicate the inequality $2n + 1 - p - \delta p \gg 1$, which follows from equations (6.10) and (6.11). Hence, in accordance with the formulas (6.9), immediately follows the condition $2\zeta_2 > 2\zeta_1$ for the size of the inner region of Cherenkov trapping of waves to be small compared with the size of the outer region Ξ_2.

Since the right-hand side of equation (6.11) is small, this equation can be written in the following approximate form:

$$2A \simeq \frac{\ln[\pi(2n - p)\lambda_j/2\lambda]}{2n - p} \equiv N_\Lambda(n, p). \quad (6.12)$$

It is obvious that for $p \ll n$ this equation has no solutions. For interference of trapped waves in the region Ξ_1 close to complete compensation, we have from equation (6.12):

$$\delta p(n, p) = [1 - (-1)^p]/2 + (-1)^p\pi^{-1}\exp\{-(p - n)N_\Lambda(n, p)\},$$

and

$$n + n/\ln\frac{\pi n\lambda_j}{2\lambda} < p < 2n - \frac{2\lambda_j}{\pi\lambda}\ln\frac{\lambda_j}{\sqrt{2}\lambda}.$$

Interference leads to doubling of the amplitude of the oscillations if

$$|\sin[\pi(p + \delta p)/2]| \simeq 1. \quad (6.13)$$

The corresponding solution obtains for $1 \ll n - p \ll n$, when $n - p \simeq n\ln 2/\ln[\pi n\lambda_j/2\lambda]$. In accordance with equation (6.13), δp is close to 1 for even p and 0 for odd p.

When the inequality (5.4) holds, equation (6.11) and hence the system (6.8) are asymptotically exact for $p \ll 2n$. We shall show that the consistency equations (3.13) and (3.14) have no solutions for $p \leqslant 2n$, i.e., we shall show that no solutions are lost on switching from equations (3.13)–(3.14) to equation (6.8).

Indeed, we set $p = 2n - p_1$ ($p_1 \ll n$), $x \equiv \pi/k_0\lambda_j = \pi\lambda(v/v_s)^2/\lambda_j \ll \pi\lambda/\lambda_j \ll 1$. Then equation (3.13) assumes the form

$$\exp[-(p_1 + 1 - \delta p)x] + \exp[-(2n + 1)x] + \exp[-(2n + 2 + p_1 - \delta p)x] = 2\lambda x/\pi\lambda_j.$$

This equation possesses solutions that satisfy the condition $x \ll \pi\lambda/\lambda_j$ only for sufficiently large values of p_1: $p_1 > (2\lambda_j/\pi\lambda)\ln(\lambda_j/\sqrt{2}\,\lambda)$. Thus, the exact consistency equations do not admit solutions for values of p close to $2n$. The solutions with smaller p values ($p < 2n - (2\lambda_j/\pi\lambda)\ln(\lambda_j/\sqrt{2}\,\lambda)$), as can be easily check, can be described by the approximate equation (6.8). For this reason, equation (6.8) is suitable for describing all 8π-kinks moving with velocities much less than the Swihart velocity.

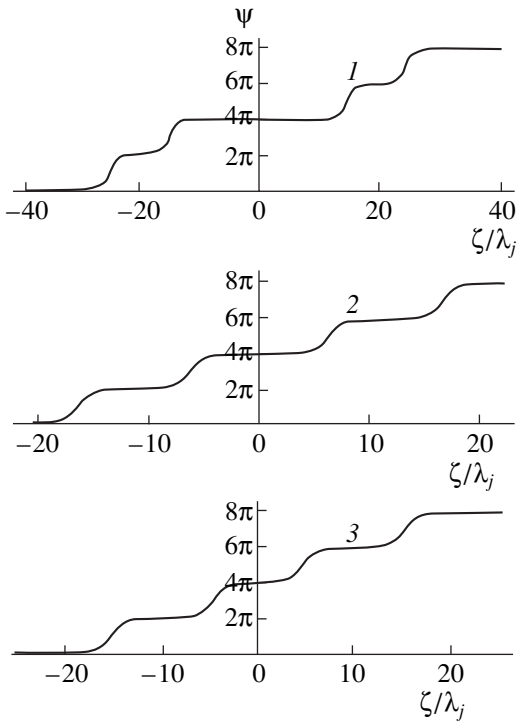

Fig. 6. Typical gluing of a slow 8π-kink; $\lambda_j/\lambda = 10$, $n = 2000$; curve *1* corresponds to $p = 3000$, $v \simeq 0.18 v_s$; curve *2*, $p = 2200$, $v \simeq 0.13 v_s$; curve *3*, $p = 1886$, $v \simeq 0.12 v_s$.

To demonstrate gluing of vortices clearly, we shall write equation (3.15) in the limit $v \ll v_s$, when equations (4.16)–(4.18) hold:

$$\begin{aligned}\psi_{8\pi}(\zeta) &= 4\pi + \pi[1 - \exp(-|\zeta + \zeta_2|/\lambda_j)]\,\mathrm{sgn}(\zeta + \zeta_2)\\ &+ \pi[1 - \exp(-|\zeta + \zeta_1|/\lambda_j)]\,\mathrm{sgn}(\zeta + \zeta_1)\\ &+ \pi[1 - \exp(-|\zeta - \zeta_2|/\lambda_j)]\,\mathrm{sgn}(\zeta - \zeta_2)\\ &+ \pi[1 - \exp(-|\zeta - \zeta_1|/\lambda_j)]\,\mathrm{sgn}(\zeta - \zeta_1)\\ &+ 4\pi(\lambda/\lambda_j)^2(v/v_s)^2\{\cos[k_0(v)(|\zeta| - \zeta_1)]\\ &\times \mathrm{sgn}\,\zeta[\theta(-\zeta + \zeta_1) - \theta(-\zeta - \zeta_1)]\\ &+ \cos[k_0(v)(|\zeta| - \zeta_2)]\\ &\times \mathrm{sgn}\,\zeta[\theta(-\zeta + \zeta_2) - \theta(-\zeta - \zeta_2)]\ \}.\end{aligned} \tag{6.14}$$

The smallness of the oscillating term due to the trapped waves is obvious here, i.e., the smallness of the field of waves, which glues individual 2π-kinks, is obvious. In order to see more easily the spatial separation of the 2π-kinks from which the 8π-kink is glued, we shall rewrite equation (6.10) using the relation (6.12):

$$\lambda_0 = \frac{2\pi}{k_0} = 2\lambda_j \ln\left[\frac{(2n-p)\pi\lambda_j}{2\lambda}\right]/(2n-p).$$

This permits writing immediately

$$2\zeta_1 \simeq p\lambda_j \frac{\ln[(2n-p)\pi\lambda_j/2\lambda]}{2n-p} \gg \lambda_j,$$

$$2\zeta_2 \simeq (4n-p)\lambda_j \frac{\ln[(2n-p)\pi\lambda_j/2\lambda]}{2n-p} \gg \lambda_j.$$

The fact that the inner region of Cherenkov trapping of waves by a slow vortex is large compared with the size λ_j of a 2π-kink should be obvious because equation (6.12) has no solutions for $p \ll n$. Figure 6 illustrates four 2π-kinks, which are strongly separated in space and are glued by Cherenkov-trapped Swihart waves into a single Josephson vortex consisting of four phased elementary vortices.

We shall demonstrate how the interference of trapped waves in the section Ξ_1 is manifested and how the phenomenon of gluing is manifested in the form of the magnetic field and the energy of 8π-kinks. For a vortex with $v \approx v_s$ we have

$$\begin{aligned}H_{8\pi}(x, \zeta) &= -\frac{\phi_0}{2^{7/4}\lambda\sqrt{\lambda\lambda_j}}\frac{\sqrt{\delta}}{1+\delta^2}\exp\left(-\frac{|x \mp d|}{\lambda}\right)\\ &\times\Big\{\exp(-k_1|\zeta + \zeta_2|) + \exp(-k_1|\zeta - \zeta_2|)\\ &+ \exp(-k_1|\zeta + \zeta_1|) + \exp(-k_1|\zeta - \zeta_1|)\\ &+ 2\delta[\theta(\zeta + \zeta_2) - \theta(\zeta - \zeta_2) - \theta(\zeta + \zeta_1) + \theta(\zeta - \zeta_1)]\\ &\times \sin[k_0(\zeta_2 - |\zeta|)] + 4\delta[\theta(\zeta + \zeta_1) - \theta(\zeta - \zeta_1)]\\ &\times \sin\left[\frac{\pi}{2}(p + \delta p)\right]\cos(k_0\zeta)\Big\},\end{aligned}$$

where k_0 and k_1 are described by the expressions (4.3). Just as in the case of 4π- and 6π-kinks, the magnetic field of a fast 8π-kink penetrates into the superconductors to a distance of the order of the London length, and the characteristic distance of variation of the monotonic component is $1/k_1 \gg \lambda$. The field of the Cherenkov-trapped Swihart waves oscillates with period $2\pi/k_0 \gg \lambda$. As the velocity of a vortex decreases, the maximum value of the monotonic component of the field and the amplitude of the field engendered by the trapped characteristic Swihart waves decrease.

The most striking difference of the magnetic field of an 8π-kink from that of 4π- and 6π-kinks is due to the presence of two overlapping regions of trapping of Swihart waves, Ξ_1 and Ξ_2. As a result of the interference of waves from different trapping regions on the section from $-\zeta_1$ to ζ_1, the term with the cosine contains the additional factor $2\sin[(p + \delta p)\pi/2]$. The appearance of such a factor results in the suppression of the amplitude of oscillations of the magnetic field, if p is even and $\delta p \ll 1$or if p is odd and $1 - \delta p \ll 1$.

The magnetic field of an 8π-kink moving with velocity $v \ll v_s$ has the form

$$H_{8\pi}(x, \zeta) = -\frac{\phi_0}{4\lambda\lambda_j}\left\{\exp\left(-\frac{|x \mp d|}{\lambda}\right)\right.$$

$$\times\left[\exp\left(-\frac{|\zeta + \zeta_2|}{\lambda_j}\right) + \exp\left(-\frac{|\zeta + \zeta_1|}{\lambda_j}\right)\right.$$

$$\left.\left. + \exp\left(-\frac{|\zeta - \zeta_1|}{\lambda_j}\right) + \exp\left(-\frac{|\zeta - \zeta_2|}{\lambda_j}\right)\right]\right\}$$

$$-\frac{\phi_0}{k_0\lambda\lambda_j^2}\exp(-k_0|x \mp d|) \qquad (6.15)$$

$$\times\left\{[\theta(\zeta + \zeta_2) + \theta(\zeta - \zeta_1) - \theta(\zeta + \zeta_1) - \theta(\zeta - \zeta_2)]\right.$$

$$\times \sin(k_0(\zeta_2 - |\zeta|)) + 2[(\zeta + \zeta_1) - \theta(\zeta - \zeta_1)]$$

$$\left.\times \sin\left(\frac{\pi}{2}(p + \delta p)\right)\cos(k_0\zeta)\right\},$$

where $k_0 = (v_s/v)^2/\lambda \gg 1/\lambda$. Similarly to the magnetic fields of slow 4π- and 6π-kinks, the oscillating part of the magnetic field is confined within a small distance from the plane of the junction, $1/k_0 \ll \lambda$.

The expression (6.15) shows that when the inequalities $2\zeta_1 \gg \lambda_j$ and $\zeta_2 - \zeta_1 \gg \lambda_j$ are satisfied, the magnetic field produced in the junction by a slow 8π-kink is a sum of the fields of 2π-kinks, which are separated by a large distance, are localized near the points $\zeta = -\zeta_2$, $\zeta = -\zeta_1$, $\zeta = \zeta_1$, and $\zeta = \zeta_2$, and are glued into a single field by Swihart waves of low intensity. Only the presence of such Swihart waves which have a low amplitude and interfere on the interval Ξ_1 makes possible the existence of the solution (6.14)—a slow 8π-kink.

Finally, we shall consider the manifestation of the gluing of vortices for the example of the expressions for the energy of an 8π-kink, trapping a field with a large number of wavelengths. Keeping in mind the inequalities $W_{mw} \ll W_w \ll W_m$, we confine our attention to the two largest contributions to the energy. For a vortex moving with velocity $v \approx v_s$, we have

$$W_m \simeq 4k_1\lambda_j W_{2\pi} = \frac{4W_{2\pi}}{\sqrt{1 - (v/v_s)^2}},$$

$$W_w \simeq 4\pi k_0\lambda_j C^2 (v/v_s)^2$$

$$\times [2n + 1 - (-1)^p(p + \delta p)\cos(\pi\delta p)]W_{2\pi} \ll W_m.$$

The uniqueness of the latter expression as compared with the wave contributions to the energy of 4π and 6π-kinks lies in the appearance of a term containing a cosine and arising because of the interference of trapped waves. We can see that the energy of the vortex consists primarily of the energies (1.7) of four 2π-kinks, and the energy of the gluing waves is small.

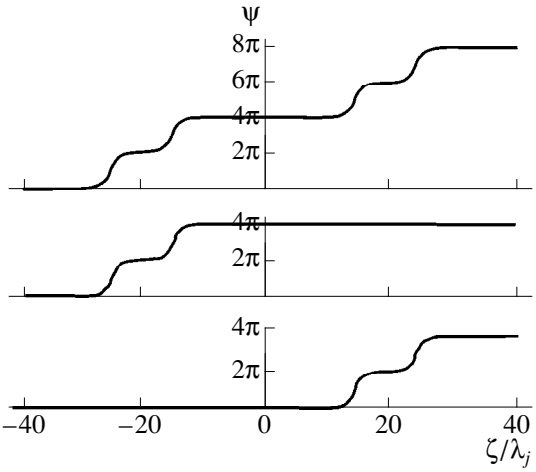

Fig. 7. Gluing of an 8π-kink from two 4π-kink; $\lambda_j/\lambda = 10$. The velocities of the vortices are $\simeq 0.18 v_s$. For an 8π-kink $n = 2000$ and $p = 3000$. For the 4π-kinks $n = 500$. The same number of wavelengths, 500.5, fits on the sections $[-\zeta_2, -\zeta_1]$ and $[\zeta_1, \zeta_2]$ (for an 8π-kink) and on the segments where Cherenkov oscillations are present (for 4π-kinks).

The energy of an 8π-kink in the limit $v \ll v_s$ also consists primarily of the energy of four 2π-kinks which are almost at rest:

$$W_{8\pi} \simeq 4W_{2\pi},$$

and the energy of the trapped Swihart waves, which give rise to the gluing of the 2π-kinks, once again is small.

In concluding this section, we shall discuss the results presented in Fig. 7. The top curve in this figure corresponds to an 8π-kink with $n = 2000$, $p = 3000$, when $2\zeta_1 \simeq 29\lambda_j \gg \zeta_2 - \zeta_1 \simeq 9.66\lambda_j$ and $v/v_s \simeq 0.18$. Each of the two bottom curves, differing from one another only by a shift along the ζ axis, corresponds to a 4π-kink (naturally, glued from 2π-kinks) with $2\zeta_0 \simeq 9.66\lambda_j$ and $v/v_s \simeq 0.18$. Comparing the three curves presented in Fig. 7 shows clearly the meaning of gluing of an 8π-kink from two 4π-kinks, each of which is glued from two 2π-kinks. We underscore that this assertion is based only on the quantitative possibility of a substantial spatial separation of the two pairs of 2π-kinks forming the 8π-kink.

One can talk about a substantial separation of pairs of 2π-kinks forming a single 8π-kink when the size of each pair $(\zeta_2 - \zeta_1)$ is small compared with the distance between the pairs $(2\zeta_1)$. In accordance with equation (6.9), such a possibility occurs for $p \gg n$ and corresponds to the case of strong suppression of Cherenkov oscillations on the segment Ξ_1. Since each pair of

2π-kinks is a 4π-kink, one can talk about gluing in the limit $p \gg n$ as gluing of two 4π-kinks.

As p decreases, the amplitude of the interference pattern in the region Ξ_1 increases and the distance between the pairs of 2π-kinks decreases. The smallest value of p (corresponding to the condition (6.13)) corresponds to the following sizes of the trapping regions: $2\zeta_1 \simeq \lambda_j \ln(\pi n \lambda_j/2\lambda)$, $2\zeta_2 \simeq 3\lambda_j \ln(\pi n \lambda_j/2\lambda)$, i.e., $2\zeta_1 \simeq \zeta_2 - \zeta_1 \gg \lambda_j$. This means that the size of each pair of 2π-kinks and the distance between these pairs are close. In this case, naturally, it is natural to talk about the fact that an 8π-kink is glued from four 2π-kinks separated by approximately the same distances.

7. CONCLUSIONS

Summarizing the material presented above, it should be underscored, first and foremost, that application of the Aubry–Volkov model was very productive. This made it possible to formulate in a comparatively simple analytical form a theory of Josephson vortices which carry a finite number (greater than one) of elementary magnetic fluxes. The analytic theory of our work not only made it possible to formulate a closed theory describing such vortices (multikinks) but it also established a new role for the Cherenkov interaction of Josephson vortices with Swihart waves. It was shown that moving multivortices are coherent structures, carrying within themselves the field of trapped Swihart waves. The frequencies and wave numbers of such waves are related with the velocity of the moving vortices by the Cherenkov resonance condition (2.9). Thus, the phenomenon of Cherenkov trapping of Swihart waves by Josephson multivortices was predicted. On the other hand, the properties which were established for the coherence of multivortices with the waves trapped by them made it possible to predict discretization (nonlinear quantization) of the characteristic velocities of freely moving Josephson multivortices. Finally, we associate the existence of multivortices itself with the phenomenon of gluing of individual vortices by the glue consisting of the trapped Swihart waves. The latter phenomenon, which we predict, is seen especially strikingly for high modes of the coherent structures which we studied, where the size of the region of gluing is much larger than the size of individual glued vortices. We note in closing that the phenomena which we predict could have a bearing not only on the dynamics of Josephson vortices but in a number of other cases on the nonlinear dynamics of vortices. It can be asserted that the theory expounded above satisfies, to a definite extent, the inquiry of the experiment formulated in [1] back in 1981. Indeed, the authors of [1] were confident that their experimental results on microwave radiation from a long Josephson junction were consistent with the fluxon picture, where according to [1] a collective motion of coherent multifluxon groups corresponds to a stable stationary state. The same conclusion was drawn in [1] in connection with an analysis of the current-voltage characteristic (IVC) presented there. Aside from this, it can be inferred that the theoretical discovery in our work of a discrete spectrum of characteristic velocities of Josephson vortices could find a unique reflection in the IVC of long Josephson junctions carrying multivortex structures. A theoretical investigation of the IVCs in the Aubry–Volkov model is a subject of a special investigation.

ACKNOWLEDGMENTS

This work was supported by the interdisciplinary project "Investigations of the Characteristic Features of High-Temperature and Other Superconductors with High Critical Parameters" of the section "Investigation of the Nature and Basic Properties of Superconductors" of the Scientific Council on Superconductivity of the Subprogram "Topical Directions in the Physics of Condensed Matter" and by State support of leading scientific schools. We are grateful to the referee for critical remarks, which improved the exposition of the material.

APPENDIX

We shall derive the contribution of trapped waves to the magnetic field of a $2\pi n$-kink and to the expression for the energy of a Josephson junction. In the calculation of the magnetic field (2.14), we note that the discontinuous functions $\operatorname{sgn}\zeta$ and $\theta(\zeta)$ appearing in the expressions (3.7) and (3.8) need not be differentiated, since the terms arising in so doing cancel one another. The last property follows from the fact that there are no jumps in the phase difference at the matching points $\xi_{n'}$ and it is guaranteed by the identity $1 = f(0) + C$. Taking this remark into account, the magnetic field (2.14) assumes the form

$$H(x, \zeta) = \sum_{n'=1}^{n} H_m(x, \zeta - \xi_{n'}) + \sum_{s=1}^{s(n)} H_w(x, \zeta, \zeta_s), \tag{A.1}$$

where

$$H_m(x, \zeta) = \frac{\phi_0}{4\pi\lambda^2} \times \int_{-\infty}^{\infty} d\zeta' K_0\left(\frac{\sqrt{(\zeta - \zeta')^2 + (x \mp d)^2}}{\lambda}\right)\frac{df(\zeta')}{d|\zeta'|}, \tag{A.2}$$

$$H_w(x, \zeta, \zeta_s) = \frac{\phi_0 C k_0}{2\pi\lambda^2} \times \int_{-\zeta_s}^{\zeta_s} d\zeta' K_0\left(\frac{\sqrt{(\zeta - \zeta')^2 + (x \mp d)^2}}{\lambda}\right)\sin[k_0(|\zeta'| - \zeta_s)]. \tag{A.3}$$

The last expression describes the magnetic field generated by Swihart waves trapped in the interval $\Xi_s = [-\zeta_s, \zeta_s]$, which are emitted by a moving vortex and are trapped by it in the process of motion.

To simplify the expression (A.2) we note the following. Under the conditions discussed $\lambda \ll \lambda_j$, free motion of multikinks is allowed only with velocities $v < v_s$. The characteristic scale of variation of the function $f(\zeta)$ is $\approx 1/k_1$ and it is always much greater than λ—the size of the effective region of integration over ζ' in equation (A.2) (this fact follows from a simple analysis of equation (2.9)). For $k_1\lambda \ll 1$, we find from equation (A.2)

$$H_m(x, \zeta) = -\frac{\phi_0}{2k_1\lambda\lambda_j^2} \times \frac{\exp[-k_1|\zeta| - |x \mp d|/\lambda]}{2 - k_1^2\lambda^2 - 2(v/v_s)^2(1 - k_1^2\lambda^2)^{3/2}}. \quad \text{(A.4)}$$

Next, using the integral representation of the Macdonald function

$$K_0\left(\frac{\sqrt{\zeta^2 + x^2}}{\lambda}\right) = \frac{1}{2}\int_{-\infty}^{\infty} \frac{dk}{\sqrt{k^2 + \lambda^{-2}}} \times \exp[ik\zeta - |x|\sqrt{k^2 + \lambda^{-2}}]$$

and integrating over ζ' and k in the complex plane, we obtain from equation (A.3)

$$H_w(x, \zeta, \zeta_s) = -\frac{\phi_0 C k_0}{4\lambda\sqrt{1 + k_0^2\lambda^2}} \times \exp(-|x \mp d|\sqrt{k_0^2 + \lambda^{-2}})[\sin(k_0|\zeta - \zeta_s|) + \sin(k_0|\zeta + \zeta_s|) - 2\cos(k_0\zeta_s)\sin(k_0|\zeta|)] \quad \text{(A.5)}$$
$$- I(x, \zeta - \zeta_s) - I(x, \zeta + \zeta_s) + 2\cos(k_0\zeta_s)I(x, \zeta),$$

where the function $I(x, \zeta)$ has the form

$$I(x, \zeta) = \frac{\phi_0 C k_0^2}{2\pi}\int_1^{\infty} \frac{dr}{\sqrt{r^2 - 1}} \times \frac{\exp(-r|\zeta|\lambda)}{r^2 + k_0^2\lambda^2}\cos\left(\frac{|x \mp d|}{\lambda}\sqrt{r^2 - 1}\right).$$

The contribution of this integral to the magnetic field (A.1) is small compared with the contribution of H_m, and we shall neglect it in what follows. The relations (A.1), (A.4), and (A.5) form the basis of the description of the magnetic field of moving vortices which have trapped generalized Swihart waves.

In conclusion, we shall present the energy of the vortex structure of a Josephson junction (2.15) as the sum

$$W = W_m + W_w + W_{mw} \quad \text{(A.6)}$$

of the energy W_m of individual vortices, the energy W_w of the Swihart waves, and the interaction energy W_{mw} of vortices with the Swihart waves which are trapped by the vortices and which glue the vortices into a single multikink vortex structure. For this, we employ the notation

$$\phi'_m(\zeta) \equiv -\pi\sum_{n'=1}^{n} \frac{df(\zeta - \xi_{n'})}{d|\zeta - \xi_{n'}|},$$

$$\phi'_m(\zeta) \equiv -2\pi k_0 C\sum_{s=1}^{s(n)} \sin[k_0(|\zeta| - \zeta_s)] \times [\theta(\zeta_s - \zeta) - \theta(-\zeta_s - \zeta)].$$

In addition, we note that from equations (2.4) and (3.19)–(3.21) follows

$$\psi(\zeta) - 2\pi I\left[\frac{\psi(\zeta)}{2\pi} + \frac{1}{2}\right] = F_m(\zeta) + \phi_w(\zeta),$$

where

$$F_m(\zeta) = -\pi\sum_{n'=1}^{n} f(\zeta - \xi_{n'})\,\mathrm{sgn}(\zeta - \xi_{n'}).$$

Then, we can write the following expression for the energy of a set of individual vortices:

$$W_m = \frac{\phi_0^2}{64\pi^3\lambda\lambda_j^2}\left\{\int_{-\infty}^{\infty} d\zeta\left\{\left(\frac{v}{\omega_j}\right)^2[\phi'_m(\zeta)]^2 + F_m^2(\zeta)\right\} + \frac{\lambda_j^2}{\pi\lambda}\int_{-\infty}^{\infty} d\zeta\int_{-\infty}^{\infty} d\zeta' K_0\left(\frac{|\zeta - \zeta'|}{\lambda}\right)\phi'_m(\zeta)\phi'_m(\zeta')\right\}. \quad \text{(A.7)}$$

Correspondingly, we have for the energy of trapped Swihart waves

$$W_w = \frac{\phi_0^2}{64\pi^3\lambda\lambda_j^2}\left\{\int_{-\infty}^{\infty} d\zeta\left\{\left(\frac{v}{\omega_j}\right)^2[\phi'_w(\zeta)]^2 + \phi_w^2(\zeta)\right\} + \frac{\lambda_j^2}{\pi\lambda}\int_{-\infty}^{\infty} d\zeta\int_{-\infty}^{\infty} d\zeta' K_0\left(\frac{|\zeta - \zeta'|}{\lambda}\right)\phi'_w(\zeta)\phi'_w(\zeta')\right\}. \quad \text{(A.8)}$$

Finally,

$$W_{mw} = \frac{\phi_0^2}{32\pi^3\lambda\lambda_j^2} \times \left\{\int_{-\infty}^{\infty} d\zeta\left\{\left(\frac{v}{\omega_j}\right)^2\phi'_m(\zeta)\phi'_w(\zeta) + F_m(\zeta)\phi_w(\zeta)\right\} + \frac{\lambda_j^2}{\pi\lambda}\int_{-\infty}^{\infty} d\zeta\int_{-\infty}^{\infty} d\zeta' K_0\left(\frac{|\zeta - \zeta'|}{\lambda}\right)\phi'_m(\zeta)\phi'_w(\zeta')\right\} \quad \text{(A.9)}$$

describes the action of a set of individual vortices in a vortex structure with trapped Swihart waves which glue individual vortices together.

REFERENCES

1. B. Dueholm, O. A. Levring, J. Mygind, *et al.*, Phys. Rev. Lett. **46**, 1299 (1981).
2. M. Peyrard and M. D. Kruskal, Physica D **14**, 88 (1984).
3. B. A. Malomed, Phys. Rev. B **47**, 1111 (1993).
4. Yu. M. Aliev and V. P. Silin, Phys. Lett. A **177**, 259 (1993).
5. Yu. M. Aliev and V. P. Silin, Zh. Éksp. Teor. Fiz. **104**, 2526 (1993) [JETP **77**(1), 142 (1993)].
6. V. P. Silin and S. A. Uryupin, Zh. Éksp. Teor. Fiz. **108**, 2163 (1995) [JETP **81**, 1179 (1995)].
7. G. L. Alfimov, V. M. Eleonsky, N. E. Kulagin, *et al.*, Chaos **3**, 405 (1993).
8. G. L. Alfimov, V. M. Eleonsky, and L. M. Lerman, Chaos **8**, 257 (1998).
9. S. Aubry and P. J. Le Daeron, Physica D **7**, 240 (1983).
10. S. Aubry, J. Phys. C **16**, 2497 (1983).
11. A. F. Volkov, Physica C **183**, 177 (1991).
12. A. F. Volkov, Physica C **192**, 306 (1992).
13. S. N. Artemenko and S. V. Remizov, Pis'ma Zh. Éksp. Teor. Fiz. **66**, 811 (1997) [JETP Lett. **66**, 853 (1997)].
14. K. K. Likharev, Rev. Mod. Phys. **51**, 101 (1979).
15. A. S. Malishevskii, V. P. Silin, and S. A. Uryupin, Phys. Lett. A **253**, 333 (1999).
16. A. S. Malishevskiĭ, V. P. Silin, and S. A. Uryupin, Pis'ma Zh. Éksp. Teor. Fiz. **69**, 318 (1999) [JETP Lett. **69**, 348 (1999)].
17. A. Barone and G. Paterno, *Physics and Applications of the Josephson Effect* (Wiley, New York, 1982; Mir, Moscow, 1984).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 4, 2000, pp. 689–694.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 4, 2000, pp. 790–796.

SOLIDS
Electron Properties

Antiferromagnetic Hall Effect Nonlinear in the Electric Current

E. A. Turov

Institute of the Physics of Metals, Ural Division, Russian Academy of Sciences, Yekaterinburg, 620219 Russia
e-mail: turov@imp.uran.ru

Received November 25, 1999

Abstract—A new galvanoantiferromagnetic effect is predicted—the Hall effect nonlinear in the electric current, i.e., a potential difference that is transverse to the current and is associated with the antisymmetric contribution to the resistivity tensor, which violates Ohm's law, for example, of the form $\Delta\rho_{ij} = -\Delta\rho_{ji} \propto LE$, where L is the antiferromagnetism vector and **E** is the electric field. The indicated contribution is characteristic for centrally antisymmetric antiferromagnets and can lead to a Hall field that is quadratic in the current: $\Delta E^{\perp} \propto LJ^2$. Similarly, the effect $\Delta E^{\perp} \propto LJ^3$ can exist in centrally symmetric antiferromagnets even in a state where the ordinary Hall effect, which is linear in **J**, is absent (for example, in hematite below the Morin point). © *2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The antiferromagnetoelectric Faraday effect (circular birefringence of light), caused by an electric field **E**, as a result of an antisymmetric contribution of the form $\Delta\varepsilon_{ij} = -\Delta\varepsilon_{ji} \propto LE$, where **L** is the antiferromagnetism vector, to the permittivity, is well-known in optics. This effect is characteristic for centrally antisymmetric exchange magnetic structures (these are the structures with which we shall be concerned). This optical effect was discovered experimentally in [1] for the crystal Cr_2O_3, which possesses a centrally antisymmetric exchange magnetic structure $\bar{1}(-)3_z(+)2_x(-)$ [2]. The corresponding invariant in $\Delta\hat{\varepsilon}$ for Cr_2O_3 has the form L_zE_z.

The analogue of the Faraday effect in kinetics is the Hall effect. There arises the question: What will be the galvanomagnetic analogue of the antiferromagnetoelectric contribution to the Faraday effect? To answer this question it is necessary to take into account a term of the form L_iE_j in the resistivity tensor ρ_{ij}. Such terms violate Ohm's law and they are antisymmetric: $\Delta\rho_{ij} = -\Delta\rho_{ji} \propto LE$, which follows from Onsager relations [2],

$$\rho_{ij}(\mathbf{L}, \mathbf{E}) = \rho_{ji}(-\mathbf{L}, \mathbf{E}).$$

The antisymmetric components of the tensor $\hat{\rho}$ are responsible for the appearance of an electric field perpendicular to the current **J**: $\mathbf{E}^{\perp} \perp \mathbf{J}$. This effect is analogous to the Hall effect, but there are very big differences. In the first place, the effect also arises in the absence of a magnetic field ($B = 0$). In the second place, it is not linear but quadratic in the current **J**: $E^{\perp} \propto LJ^2$. The specific expressions for $\mathbf{E}^{\perp}$ depend on the crystal-chemical structure (system), the exchange magnetic structure, the orientational state (the direction of the vector **L**), and the geometry of the experiments.

Let us assume that the sample has been prepared and mounted in a manner so that the current can flow in only one direction (let this direction be x_1). Then, if symmetry permits the effect under study, a field

$$E_2^{\perp} = \rho_{[21]}J_1 = -rL\rho_1J_1^2 \quad (1)$$

should arise in the transverse direction x_2. Here r is a constant and ρ_1 is the resistivity along x_1. If the directions of the current and transverse field excited by it are interchanged, then we have instead of equation (1)

$$E_1^{\perp} = \rho_{[12]}J_2 = rL\rho_2J_2^2. \quad (2)$$

The indices enclosed in brackets in $\hat{\rho}$ signify the antisymmetry of the corresponding components, $\rho_{[12]} = -\rho_{[21]}$, whence follows, specifically, $\rho_1J_1 = \rho_2J_2$.

In what follows we shall present the results for tetragonal and trigonal antiferromagnets with various exchange magnetic structures in the easy-axis (EA) and easy-plane (EP) states.

2. EASY-AXIS ANTIFERROMAGNETS WITH AN EVEN PRINCIPAL SYMMETRY AXIS

We shall consider first the exchange magnetic structure $\bar{1}(-)4_z(+)2_d(-)$, which is characteristic for certain tririutiles (Fe_2TeO_6, and others, Fedorov space group $G_F \equiv P4_2/mnm$), in the EA state $\mathbf{L} \parallel \mathbf{z} \parallel \mathbf{4}$ (**4** is a fourfold symmetry axis). If $\mathbf{B} \parallel \mathbf{z}$ also, then the latter is stable only in fields $B_z < H_{sf}$, where H_{sf} is the "spin-flop" field. This case is the most unfavorable one in the sense that the effect of interest to us is absent in the EA state.

The same thing can also be said about the exchange magnetic structure $\bar{1}(-)3_z(+)2_x(-)$ and the EA state **L** || **z** || **3** (likewise, for $B_z < H_{sf}$), which are characteristic for the best known centrally antisymmetric antiferromagnet: chromium oxide Cr_2O_3 ($G_F \equiv R\bar{3}c$).

3. TETRAGONAL ANTIFERROMAGNETS WITH AN ODD $4_z(-)$-AXIS IN THE EASY-AXIS STATE

We are talking about to exchange magnetic structures with the classification numbers $\bar{1}(-)4_z(-)2_d(\pm)$, which correspond to certain trirutiles as well as rare-earth phosphates ($DyPO_4$ and others) and vanadates ($GdVO_4$ and others) ($G_F \equiv I4_1/amd$; see, for example, [3]).

Let us examine once again the EA state, which occurs in the indicated phosphates and vanadates. The parity of the 2_d-axis depends on how we orient the *x*-axis (and, correspondingly, the *y*-axis): If we orient it along the edge of the basal square, then $2_d \equiv 2_d(-)$, and if we orient it along the diagonal of the square, then $2_d \equiv 2_d(+)$. To switch from one case to the other, the coordinate system must be rotated by 45° around the *z*-axis.

The compounds $DyPO_4$ (T_N = 3.4 K) and $GdVO_4$ (T_N = 2.4 K) possess the exchange magnetic structure $\bar{1}(-)4_z(-)2_d(-)$ and the EA state **L** || **z**. In contrast to the preceding case of a structure with $4_z(+)$ (Section 2), here the quadratic Hall effect exists.

The invariant, relative to the symmetry elements appearing in the above-indicated classification indices of exchange magnetic structures, expressions for the antisymmetric components of the tensor $\hat{\rho}$ in this case are as follows [2]:

$$\begin{aligned} \rho_{[zx]} &= R_2 B_y + r_2 L_z E_x, \\ \rho_{[yz]} &= R_2 B_x + r_2 L_z E_y, \\ \rho_{[xy]} &= R_3 B_z, \end{aligned} \tag{3}$$

where *R* and *r* are constants. Further, let **J** || **x** and $J_y = J_z = 0$. Then, using equations (3), we have

$$\begin{aligned} E_x^{\parallel} &= \rho_1 J_x, \quad E_y^{\perp} = \rho_{yx} J_x = -R_3 B_z J_x, \\ E_z^{\perp} &= \rho_{zx} J_x = R_2 B_y J_x + r_2 \rho_1 L_z J_x^2. \end{aligned} \tag{4}$$

Thus, indeed, besides the standard Hall effect linear in **J** (vanishing for **B** = 0), there appears a quadratic effect which, in addition, is proportional to the *z* component of the vector **L.** Therefore, the samples require special treatment [4], which removes the domain structure along L_z, in order to observe it.

The formula corresponding to the case **B** || **x** and **J** || **y** can be easily derived from the expression (4). The symmetry element $4_z(-)$ must be applied to it. The result is

$$E_z^{\perp} = -R_2 B_x J_y - r_2 \rho_2 L_z J_y^2. \tag{5}$$

(We note only that $\rho_2 = \rho_1 \equiv \rho_\perp$ for the case under study.) The general expression for an arbitrary direction of **J** in the basal plane (let **J** make an angle φ_J with the *x* axis) for **B** = 0 assumes the form

$$E_z^{\perp} = r_2 \rho_\perp L_z (J_x^2 - J_y^2) \equiv r_2 \rho_\perp L_z J_\perp^2 \cos(2\varphi_J). \tag{6}$$

It is easy to show that this expression is invariant with respect to the elements $\bar{1}(-)$, $4_z(-)$, and $2_d(-)$, appearing in the classification indices of the exchange magnetic structure.

We shall now examine the case **J** || **z** with the same exchange magnetic structure and orientational EP state. Here, we can talk about a second effect, which is associated with terms of the form $L_i E_j$ in the tensor $\hat{\rho}$. This effect occurs only in the presence of a field $\mathbf{B} \neq 0$. Using equations (3), in this case we obtain for the field components transverse to the current

$$\begin{aligned} E_x^{\perp} &= -\frac{R_2 B_y J_z}{1 + r_2 L_z J_z} \simeq -R_2 B_y J_z (1 - r_2 L_z J_z), \\ E_y^{\perp} &= \frac{R_2 B_x J_z}{1 - r_2 L_z J_z} \simeq R_2 B_x J_z (1 + r_2 L_z J_z). \end{aligned} \tag{7}$$

Here there is also a contribution, which quadratic in **J,** to the Hall field (in the approximate expressions (7)), but it is also proportional to the field **B**. Since this effect is proportional to the product of the cofactors R_2 and r_2, which are responsible for the standard and quadratic Hall effects, respectively, it can be inferred that it makes a small contribution to $E^{\perp}$ compared with the second term in equations (4), (5), or (6).

4. EASY-PLANE STATE FOR THE EXCHANGE MAGNETIC STRUCTURE $\bar{1}(-)4_z(-)2_d(-)$

The antiferromagnets (trirutiles) Cr_2WO_6 (T_N = 69 K) and V_2WO_6 (T_N = 370 K) have the magnetic structure indicated in the heading (i.e., an exchange magnetic structure and orientational state). Since in this case **L** $\perp$ **z**, instead of equation (3), the components of the tensor $\rho_{[ij]}$ which correspond to this state must be rewritten as

$$\begin{aligned} \rho_{[xy]} &= R_3 B_z + r_3 (L_x E_y + L_y E_x), \\ \rho_{[zx]} &= R_2 B_y + r_2 L_x E_z, \\ \rho_{[yz]} &= R_2 B_x + r_2 L_y E_z. \end{aligned} \tag{8}$$

Here and everywhere below, we shall assume that the projection $\mathbf{B}_\perp \perp \mathbf{z}$ of the field is large enough so that the vector $\mathbf{L}_\perp \perp \mathbf{z}$ is perpendicular to $\mathbf{B}_\perp$ in the basal plane:

$\mathbf{L}_\perp \perp \mathbf{B}_\perp$. We note that this assumption will also pertain to the case where the state $\mathbf{L}_\perp \perp \mathbf{z}$ appeared in the "spin-flop" process in a field $B_z > H_{sf}$ from the EA state with $\mathbf{L} \equiv \mathbf{L}_s \parallel \mathbf{z}$. Introducing the azimuthal angle φ_B for the vector $\mathbf{B}_\perp$ (measured from the binary symmetry axis $2_x \parallel x$), we have in these cases

$$B_x = B_\perp \cos\varphi_B, \quad B_y = B_\perp \sin\varphi_B,$$
$$L_x = -L_\perp \sin\varphi_B, \quad L_y = L_\perp \cos\varphi_B.$$

We shall examine first what is perhaps the most interesting case of a sample with current orientation $\mathbf{J} \parallel \mathbf{z}$. Then, using equations (8), we find for the field transverse to $\mathbf{J}$ (the Hall field)

$$\begin{aligned} E_x^\perp &= \rho_{[xz]} J_z = -\sin\varphi_B (R_2 B_\perp - r_2 \rho_3 L_\perp J_z) J_z, \\ E_y^\perp &= \rho_{[yz]} J_z = \cos\varphi_B (R_2 B_\perp + r_2 \rho_3 L_\perp J_z) J_z. \end{aligned} \tag{9}$$

Hence we obtain for the field part of the effect ($\propto B_\perp$) that $\mathbf{E}_B^\perp \perp \mathbf{B}_\perp$, and its magnitude is isotropic: $E_B^\perp = |R_2 B_\perp J_z|$ (does not depend on the angle φ_B). The sign changes when J_z changes sign. At the same time, we have for the antiferromagnetic (quadratic) part of the effect

$$\mathbf{E}_L^\perp \cdot \mathbf{L}_\perp = r_2 \rho_3 L_\perp^2 J_z^2 \cos(2\varphi_B). \tag{10}$$

The projection $\mathbf{E}_L^\perp$ on the vector $\mathbf{L}_\perp$ is anisotropic: It depends on the direction of the field $\mathbf{B}_\perp$ (and, therefore, the vector $\mathbf{L}_\perp$) in the basal plane. Specifically,

$$\mathbf{E}_L^\perp \perp \mathbf{L}_\perp \quad \text{for} \quad \varphi_B = \pm\frac{\pi}{4},$$

$$\mathbf{E}_L^\perp \parallel \mathbf{L}_\perp \quad \text{for} \quad \varphi_B = 0, \pm\frac{\pi}{2}.$$

However, the absolute magnitude of the effect LJ^2 remains isotropic (independent of the angle φ_B):

$$E_L^\perp = \left| r_2 \rho_3 L_\perp J_z^2 \right|. \tag{11}$$

We shall now present the results for various cases where $\mathbf{J} \perp \mathbf{z}$. We shall confine our attention to examples that take into account two very simple situations: The direction of the current is along the magnetic field $\mathbf{B}_\perp$ (cases (a) and (b)) or along the vector $\mathbf{L}_\perp$ (the cases (c) and (d)).

(a) $\mathbf{L} \parallel \mathbf{y}$, $\mathbf{J} \parallel \mathbf{B}_\perp \parallel \mathbf{x}$:

$$E_y^\perp = -(R_3 B_z + r_3 \rho_\perp L_y J_x) J_x, \quad E_z^\perp = 0. \tag{12}$$

(b) $\mathbf{L} \parallel \mathbf{x}$, $\mathbf{J} \parallel \mathbf{B}_\perp \parallel \mathbf{y}$:

$$E_x^\perp = (R_3 B_z + r_3 \rho_\perp L_x J_y) J_y. \tag{13}$$

It is easy to verify that the formulas for the cases (a) and (b) are related with one another in accordance with the symmetry of the transformations $4_z(-)$ and $2_d(-)$.

(c) $\mathbf{J} \parallel \mathbf{L} \parallel \mathbf{y}$, $\mathbf{B}_\perp \parallel \mathbf{x}$:

$$E_x^\perp = \frac{R_3 B_z J_y}{1 - r_3 L_y J_y}, \quad E_z^\perp = -\frac{R_2 B_x J_y}{1 + r_2 L_y J_y}. \tag{14}$$

(d) $\mathbf{J} \parallel \mathbf{L} \parallel \mathbf{x}$, $\mathbf{B}_\perp \parallel \mathbf{y}$:

$$E_y^\perp = -\frac{R_3 B_z J_x}{1 + r_3 L_x J_x}, \quad E_z^\perp = \frac{R_2 B_y J_x}{1 - r_2 L_x J_x}. \tag{15}$$

The relations (14) and (15) are also related with the indicated transformations.

Thus, here, the effect LJ^2 in which we are interested exists for $\mathbf{J} \perp \mathbf{L}$ only in the geometry (a) and (b), when $\mathbf{J} \parallel \mathbf{B}_\perp$, while for (c) and (d) only the second effect, which is analogous to one described by equations (1) in the state $\mathbf{J} \parallel \mathbf{L} \parallel \mathbf{z}$, occurs.

5. EASY-PLANE STATE FOR THE EXCHANGE MAGNETIC STRUCTURE $\bar{1}(-)4_z(+)2_d(-)$

According to [3], the trirutile Cr_2TeO_6 ($T_N = 105$ K) possesses such a magnetic structure. This material is of interest if only because the classification indices of this structure differ from those in the preceding case only by the parity of the 4_z symmetry axis: $4_z(-) \longrightarrow 4_z(+)$. Moreover, the other trirutile Fe_2TeO_6 ($T_N = 219$ K) mentioned above, which possesses an even axis $4_z(+)$, can transform as a result of a "spin-flop" transition into an EP state in fields $B_z > H_{sf}$. We shall hereby clarify the role of the parity of the element 4_z for the phenomena of interest to us.

The form of the tensor $\rho_{[ij]}$ also changes when the parity of 4_z changes. Instead of equation (8), we now have [2]

$$\begin{aligned} \rho_{[xy]} &= R_3 B_z + r_3 (L_x E_x + L_y E_y), \\ \rho_{[zx]} &= R_2 B_y + r_2 L_y E_z, \\ \rho_{[yz]} &= R_2 B_x + r_2 L_x E_z. \end{aligned} \tag{16}$$

(The changes do not concern the field terms, which are identical for all uniaxial crystals.)

We shall write out the results for $\mathbf{E}^\perp$ for the same geometric situations as in Section 4.

For $\mathbf{J} \parallel \mathbf{z}$

$$\begin{aligned} E_x^\perp &= -(R_2 B_\perp \sin\varphi_B + r_2 \rho_3 L_\perp J_z \cos\varphi_B) J_z, \\ E_y^\perp &= (R_2 B_\perp \cos\varphi_B - r_2 \rho_3 L_\perp J_z \sin\varphi_B) J_z. \end{aligned} \tag{17}$$

In accordance with what we have said above, the field terms of the Hall field (linear in J_z) are identical to those for the preceding case (Section 4): $\mathbf{E}_B^\perp \perp \mathbf{B}_\perp$ and $E_B^\perp = |R_2 B_\perp J_z|$ for any value of φ_B. For the quadratic antiferromagnet effect

$$\mathbf{E}_L^\perp \cdot \mathbf{L}_\perp = 0, \tag{18}$$

so that in contrast to equation (10), here, $\mathbf{E}_L^\perp \perp \mathbf{L}_\perp$ is independent of the angle φ_B. However, the relation (11) for the absolute magnitude of the effect holds in the present case also.

Next, for $\mathbf{J} \perp \mathbf{z}$ we have under the same conditions as for equations (12)–(15)

(a) $\mathbf{L} \parallel \mathbf{y}$, $\mathbf{J} \parallel \mathbf{B}_\perp \parallel \mathbf{x}$:

$$E_y^\perp = -\frac{R_3 B_z J_x}{1 + r_3 L_x J_x}, \quad E_z^\perp = 0; \tag{19}$$

(b) $\mathbf{L} \parallel \mathbf{x}$, $\mathbf{J} \parallel \mathbf{B}_\perp \parallel \mathbf{y}$:

$$E_x^\perp = \frac{R_3 B_z J_y}{1 - r_3 L_x J_y}, \quad E_z^\perp = 0; \tag{20}$$

(c) $\mathbf{J} \parallel \mathbf{L} \parallel \mathbf{y}$, $\mathbf{B}_\perp \parallel \mathbf{x}$:

$$E_x^\perp = (R_3 B_z + r_3 \rho_\perp L_y J_y) J_y, \quad E_z^\perp = -R_2 B_x J_y; \tag{21}$$

(d) $\mathbf{J} \parallel \mathbf{L} \parallel \mathbf{x}$, $\mathbf{B}_\perp \parallel \mathbf{y}$:

$$E_y^\perp = -(R_3 B_z + r_3 \rho_\perp L_x J_x) J_x, \quad E_z^\perp = R_2 B_y J_x. \tag{22}$$

Of course, in this case, as should happen, the formulas for the situations (a) and (b) or (c) and (d) transform into one another under the transformations $4_z(+)$ and $2_d(-)$, appearing in the classification indices of the exchange magnetic structure under study.

Finally, we note that in contrast to the structure $\bar{1}(-)4_z(-)2_d(-)$, the quadratic effect $E_L^\perp \propto LJ^2$ for the structure $\bar{1}(-)4_z(+)2_d(-)$ exists in the configuration $\mathbf{J}_\perp \parallel \mathbf{L}_\perp$ but not $\mathbf{J}_\perp \parallel \mathbf{B}_\perp$ (Section 4). Once again, the parity of the 4_z symmetry axis comes into play.

6. EASY-PLANE STATE FOR THE EXCHANGE MAGNETIC STRUCTURE $\bar{1}(-)3_z(+)2_x(-)$

We shall now consider the second of the systems which we intended to study: trigonal (rhombohedral) antiferromagnets. We shall confine our attention to the exchange magnetic structure $\bar{1}(-)3_z(+)2_x(-)$ and EP state ($\mathbf{L} \perp \mathbf{3} \parallel \mathbf{z}$). True, the well-known antiferromagnet with such an exchange magnetic structure, Cr_2O_3 (T_N = 318 K), is found in a natural EA state with $\mathbf{L} \parallel \mathbf{z}$, but, in a magnetic field $B_z > H_{sf} \approx 100$ kOe it can transform into the "spin-flop" phase with $\mathbf{L} \perp \mathbf{z}$, which we shall assume have occurred. Moreover, once again, it is assumed that a field $\mathbf{B}_\perp \perp \mathbf{z}$ is present and that this field, overcoming the basal anisotropy, aligns the vector's $\mathbf{L}_\perp \perp \mathbf{B}_\perp$ in the basal plane *xy*.

The tensor $\rho_{[ij]}$ for the state of interest to us has the form

$$\begin{aligned} \rho_{[xy]} &= R_3 B_z + r_3 (L_x E_x + L_y E_y), \\ \rho_{[zx]} &= R_2 B_y + r_2 (L_x E_x - L_y E_y) + r_1 L_y E_z, \\ \rho_{[yz]} &= R_2 B_x + r_2 (L_x E_y + L_y E_x) + r_1 L_x E_z. \end{aligned} \tag{23}$$

In many cases, the formulas for $\mathbf{E}^\perp$ in this state will be somewhat more complicated than the formulas presented above for a tetragonal antiferromagnet, but we need only present them in an approximate form, neglecting terms which contain products of the form $r_i r_j$ and $R_i r_j$ and are small compared to 1.

Once again we begin with the case $\mathbf{J} \parallel \mathbf{z}$. The approximate results for $E_x^\perp$ and $E_y^\perp$ with the indicated quantities neglected are exactly described by equations (17), obtained for the exchange magnetic structure $\bar{1}(-)4_z(+)2_d(-)$, and it is pointless to write them out once again. Of course, relations of the form (18) and (11), determining the perpendicularity of $\mathbf{E}_L^\perp$ to $\mathbf{L}_\perp$ and the isotropic absolute magnitude of the effect LJ^2, are also preserved.

In concluding our analysis of centrally antisymmetric antiferromagnets, before proceeding to the centrally symmetric case, we shall present several formulas for the exchange magnetic structure discussed (with $3_z(+)$) in the situation $\mathbf{J} \perp \mathbf{z}$. Specifically, we shall confine our attention to the cases for which the quadratic Hall effect is different from zero. This will occur for two geometric configurations (c, d) from the four configurations considered (a, b, c, d) for the corresponding tetragonal structures:

c) $\mathbf{J} \parallel \mathbf{L} \parallel \mathbf{x}$, $\mathbf{B}_\perp \parallel \mathbf{y}$:

$$E_y^\perp = -(R_3 B_z + r_3 \rho_\perp L_x J_x) J_x, \quad E_z^\perp = (R_2 B_y + r_2 \rho_\perp L_x J_x) J_x; \tag{24}$$

d) $\mathbf{J} \parallel \mathbf{L} \parallel \mathbf{y}$, $\mathbf{B}_\perp \parallel \mathbf{x}$:

$$E_x^\perp = R_3 B_z J_y + r_3 \rho_\perp L_y J_y^2, \quad E_z^\perp = -R_2 B_x J_y. \tag{25}$$

Comparing these formulas with equations (19)–(22) demonstrates at least one case where the crystal system influences the results. Even for the configurations (c) and (d), where the LJ^2 effect occurs for both systems, in the rhombohedral case for (d) it does not occur according to equation (25) for the component $E_z^\perp$.

7. CENTRALLY SYMMETRIC EXCHANGE MAGNETIC STRUCTURE OF THE TYPE $\bar{1}(+)3_z(+)2_x(-)$

So far we have been discussing centrally antisymmetric antiferromagnets. It is in these materials that a Hall effect quadratic in the current $E^\perp \propto LJ^2$ arises. To

complete our analysis we must show very briefly, at least at the fundamental level, what occurs instead of this effect in centrally symmetric antiferromagnets. We shall consider the exchange magnetic structure $\bar{1}(+)3_z(+)2_x(-)$, which occurs in hematite α-Fe_2O_3. It is obvious that here the antisymmetric components of the tensor ρ_{ij}, which violate Ohm's law, will be quadratic in the field **E**:

$$\Delta\rho_{ij} = -\Delta\rho_{ji} = LE^2. \quad (26)$$

In the temperature range $T > T_M = 260$ K, which is above the Morin point, where hematite, being in an EP state, possesses weak ferromagnetism and, in a field $\mathbf{E}^\perp$, a spontaneous term $\propto LJ$, an additional term that is cubic in the current **J** appears because of equation (26). We shall write the results only schematically:

$$E^\perp = r_1 LJ + r_3 LJ^3 \quad (27)$$

(once again, r_i are constants). The first term has been observed experimentally in [5]. The second term seems to be small compared to the first term; no special attempts were made to observe it.

A more favorable situation is the situation in the EA state, in which hematite is found at temperatures $T < T_M$. Here we shall write out one formula for the specific case where a current $\mathbf{J} \equiv \mathbf{J}_\perp \perp \mathbf{z}$ is studied with $\mathbf{L} \parallel \mathbf{z}$. In this case a Hall field

$$E_z^\perp = r_2 L_z \rho_\perp^2 J_\perp^3 \sin(3\varphi_J), \quad (28)$$

where φ_j is the angle between the current **J** and the symmetry axis 2(–), parallel to the *x* axis, should arise. The main point is that this effect can be observed against a zero background. Of course, once again, care must be taken to make sure that the sample consists of a single domain. A very suitable object for such an investigation is hematite at temperatures below the Morin point.

8. DISCUSSION

The theory presented above is a phenomenological symmetry theory. Therefore without additional microscopic calculations it is difficult to say anything about the magnitude of the predicted effects. Still, in my opinion, this is not the main difficulty of the experimental observation of these effects. Ultimately, the optical analog, the antiferromagnetoelectric Faraday effect, turned out to be measurable experimentally [1]. The fact that the spontaneous antiferromagnetic Hall effect ($E^\perp \propto LJ$), was discovered in hematite [5], and turned out to be orders of magnitude stronger than the ordinary (normal) Hall effect ($E^\perp \propto BJ$), is also encouraging.

The difficulty lies in finding and, possibly, preparing appropriate conducting centrally antisymmetric antiferromagnets (when the LJ^2 effect is being studied). Unfortunately, I have no information about the electrically conducting properties of antiferromagnets for whose symmetry these effects were studied in the present paper. This pertains trirutiles as well as rare-earth phosphates and candidates. The Cr_2O_3 crystals on which optical experiments have been performed (see the review [1] and the corresponding citations) are evidently optically transparent (i.e., virtually nonconducting) insulators. It can only be inferred that when these oxides are specially treated with the addition of certain dopants, they can become semiconductors. The existence of other conducting oxides (and, specifically, α-Fe_2O_3) of this type is hopeful.

Of course, other centrally antisymmetric antiferromagnets with adequate conductivity can also be found among crystals with orthorhombic symmetry, for example, rare-earth orthoferrites, orthochromites, and orthoaluminates ($G_F \equiv Pbnm$), in which the rare-earth magnetic ions can become ordered in a centrally antisymmetric manner (see, for example, Table 14.1 in [6]). However, for them it is necessary to perform a special analysis, similar to the one performed above, using the above-indicated symmetry and the exchange magnetic structures and orientational states corresponding to them. As an example, we shall nonetheless present here a formula for the structure $\bar{1}(-)2_x(-)2_y(+)$. Specifically, the orthoaluminates $TbAlO_3$, $DyAlO_3$, and $GdAlO_3$ possess such an exchange magnetic structure [2,7]. For the first two $\mathbf{L} \parallel \mathbf{x}$ and for the third $\mathbf{L} \parallel \mathbf{y}$. For the $\mathbf{L} \parallel \mathbf{x}$ state, it is not difficult to prove that relations with the following form are invariant with respect to the elements appearing in the indicated classification indices of the exchange magnetic structure:

$$E_z^\perp = rL_x J_y^2.$$

This attests to the possibility of the existence of a quadratic Hall effect for $TbAlO_3$ and $DyAlO_3$. At the same time, the effect is absent in the state $\mathbf{L} \parallel \mathbf{y}$, characteristic for $GdAlO_3$. This once again demonstrates the strong dependence of the effect $E^\perp \propto LJ^2$ on the orientational state (direction of vector **L**). It even appears in the EP state of a tetragonal antiferromagnet, if the 4_z symmetry axis is odd (see equations (9) or (10)). Such an anisotropy does not occur for the analogous configuration in the case of an even axis ($4_z(+)$ and $3_z(+)$).

In closing, I wish to make several important remarks.

The reader who has not delved deeply into the heart of the matter could have the impression that the quadratic Hall effect studied here is a trivial effect due to the magnetoelectric contribution to the magnetization $M_E = \hat{\alpha} E$, which, via terms of the form M_i in the tensor $\rho_{[ij]}$ [2] (having the same symmetry as the terms with B_i) also leads to terms of the form $L_i E_j$ in this tensor, since the magnetoelectric susceptibility tensor depends on *L*, $\hat{\alpha} \sim L$. However, this is not the case. The coefficients r_i in the expressions (3), (8), and so are, generally

speaking, independent of $\hat{\alpha}$, so that the coefficients r_i can be different from 0 even if $\hat{\alpha} = 0$. A similar situation has already been encountered repeatedly for effects which are linear in the vector **L**. In the first place, the spontaneous Hall and Faraday effects in the centrally symmetric antiferromagnets, which are proportional to **L**, by no means are determined by the weakly ferromagnetic contribution to the magnetization $\mathbf{M}_s = \hat{d}\mathbf{L}$. Experiments (see [3]) have shown that the indicated effects remain virtually unchanged when the magnetization produced by a field **H** varies severalfold. Furthermore, the study of the effect of a field **E** on the NMR frequency in centrally antisymmetric antiferromagnets has shown that the magnetoelectric channel of this influence (via $\mathbf{M}_E$) by no means gives a complete picture of the splitting of the NMR spectrum by the field **E**. This splitting is determined by independent terms of the form L_iE_j in the hyperfine field [8]. Finally, in the so-called nonreciprocal linear birefringence effect, which is due to terms of the type L_ik_j in the permittivity $\hat{\varepsilon}$ (**k** is the wave vector), the $\mathbf{M}_E$ channel likewise is not the determining channel. The tensor $\hat{\varepsilon}$ contains independent terms which are due to the linear spatial dispersion of different nature (see [1, 4] and the citations).

Now I shall consider the question of how the Hall fields which are linear and quadratic in the current can be distinguished experimentally. This can be done very simply: Measurements must be performed for two opposite directions of the current **J** and the half-sum of the results taken. It is important that in so doing another extraneous transverse field effect (transverse galvanomagnetic effect) due to the anisotropy of asymmetric part of the tensor $\hat{\rho}$, which under conditions where Ohm's law holds for this part of $\hat{\rho}$ is also linear in **J**, is eliminated. This pertains especially to the EP state, in which the indicated anisotropy in $\hat{\rho}$ appears as result of terms of the type *LL* and *LB* [2,5]. There is also a well-known standard method for distinguishing the Hall effect from the transverse galvanomagnetic effect: Measure the transverse fields $\mathbf{E}^{\perp}$ corresponding to the ρ_{ij} and ρ_{ji}. Then the half-difference of the results will give the Hall effect and the half-sum will give the transverse galvanomagnetic effect.

Off all of the effects predicted above, at present time the effect $E^{\perp} \propto LJ^3$, represented by equations (27) and (28) seems to be most accessible experimentally (because samples (hematite) are very accessible). In the first case, it is necessary to check the existence of a deviation of $E^{\perp}$ from linearity in the current. In the second case (below the Morin point, where $\mathbf{L} \parallel \mathbf{z}$) an attempt can be made to observe the Hall field, which here is proportional to $L_z J_{\perp}^3 \sin(3\varphi_j)$. But, still, the quadratic Hall effect in centrally antisymmetric antiferromagnets is most attractive for an experimental search precisely because of its unusual nature ($\propto J^2$). Moreover, it should be expected that it will be greater in magnitude.

ACKNOWLEDGMENTS

This work was partially supported by the Russian Foundation for Basic Research (project no. 99-02-16268).

I thank M.I. Kurkin and V.V. Nikolaev for a helpful discussion.

REFERENCES

1. R. V. Pisarev, Ferroelectrics **162**, 191 (1994); B. B. Krichevtsov, V. V. Pavlov, and R. V. Pisarev, Zh. Éksp. Teor. Fiz. **94** (2), 284 (1988) [Sov. Phys. JETP **67**, 378 (1988)].
2. E. A. Turov, *Kinetic, Optical, and Acoustic Properties of Antiferromagnets* (Izd. Ross. Akad. Nauk, Sverdlovsk, 1990).
3. D. E. Cox, in *Magnetoelectric Interaction Phenomena in Crystals*, Ed. by A. J. Freeman and H. Schmid (Gordon and Breach, London, 1975).
4. S. Shtrikman and D. Treves, Phys. Rev. **130**, 986 (1963).
5. K. B. Vlasov, B. A. Rozenberg, E. A. Turov, and V. G. Shavrov, in *Dynamic and Kinetic Properties of Antiferromagnets*, Ed. by S. V. Vonsovskiĭ and E. A. Turov (Nauka, Moscow, 1986), Chap. 2; K. B. Vlasov, E. A. Rozenberg, A. G. Titova, *et al.*, Fiz. Tverd. Tela (Leningrad) **22**, 1656 (1980) [Sov. Phys. Solid State **22**, 967 (1980)].
6. A. K. Zvezdin, V. M. Matveev, A. A. Mukhin, and A. I. Popov, *Rare-Earth Ions in Magnetically Ordered Crystals* (Nauka, Moscow, 1985).
7. R. M. Hornreich, Solid State Commun. **7**, 1081 (1969).
8. V. V. Leskovets and E. A. Turov, Fiz. Tverd. Tela (Leningrad) **42** (5), (2000) [Phys. Solid State **42** (2000) (in press)].

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 4, 2000, pp. 695–713.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 4, 2000, pp. 797–817.

NONLINEAR PHYSICS

Statistical Limit in a Completely Integrable System with Deterministic Initial Conditions

A. V. Gurevich*, [1], N. G. Mazur, and K. P. Zybin***

** Division of Theoretical Physics, Lebedev Physical Institute, Russian Academy of Sciences, Moscow, 117924 Russia*

*** Schmidt Institute of Physics of the Earth, Russian Academy of Sciences, Moscow, 123810 Russia*

[1] e-mail: alex@td.lpi.ac.ru

Received October 6, 1999

Abstract—The asymptotic behavior of the solutions of the KdV equation in the classical limit with an oscillating nonperiodic initial function $u_0(x)$ prescribed on the entire x axis is investigated. For such an initial condition, nonlinear oscillations, which become stochastic in the asymptotic limit $t \longrightarrow \infty$, develop in the system. The complete system of conservation laws is formulated in the integral form, and it is demonstrated that this system is equivalent to the spectral density of the discrete levels of the initial problem. The scattering problem is studied for the Schrödinger equation with the initial potential $-u_0(x)$, and it is shown that the scattering phase is a uniformly distributed random quantity. A modified method is developed for solving the inverse scattering problem by constructing the maximizer for an N-soliton solution with random initial phases. A one-to-one relation is established between the spectrum of the discrete levels of the initial state of the system and the spectrum established in phase space. It is shown that when the system passes into the stochastic state, all KdV integral conservation laws are satisfied. The first three laws are satisfied exactly, while the remaining laws are satisfied in the WKB approximation, i.e., to within the square of a small dispersion parameter. The concept of a quasisoliton, playing in the stochastic state of the system the role of a standard soliton in the dynamical limit, is introduced. A method is developed for determining the probability density $f(u)$, which is calculated for a specific initial problem. Physically, the problem studied describes a developed one-dimensional turbulent state in dispersion hydrodynamics. *© 2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The asymptotic behavior of the solutions of the Korteveg–de Vries (KdV) equation

$$u_t + 6uu_x + \varepsilon^2 u_{xxx} = 0 \qquad (1)$$

was studied in [1] in the semiclassical (Witham) limit

$$\varepsilon \ll 1. \qquad (2)$$

The initial functions were assumed to be hydrodynamic, i.e., smooth, oscillatory, nonperiodic, and defined on the entire x axis. In this formulation the problem is fundamentally different from the classical KdV solutions studied previously [2–5] in that, as will be shown in the present work, the asymptotic behavior of the solution as $t \longrightarrow \infty$ acquires a statistical character as a result of phase mixing. The infinite system of interacting solitons with uniformly distributed random phases, which develops in the process, leads to stochastic oscillations. Such a system can be described in terms of a continuous random process. Specifically, in the asymptotic limit $t \longrightarrow \infty$ the exact dynamical value of $u(x, t)$ becomes meaningless and one can talk only about $f(u, x, t)$—the density of the probability of finding at the point (x, t) a given value of u or about $f(u, u'; x, t, x', t')$—a two-point distribution function, i.e., the joint probability density of a value u occurring at the point (x, t) and a value u' occurring at the point (x', t'), and so on. It is remarkable that it becomes possible to simplify substantially the description of the asymptotic behavior of the solution. Specifically, for a spatially uniform initial function $u_0(x)$ the probability density $f(u)$ depends only on the velocity u, and the spatial correlation function is $K = K(s)$, where $s = |x - x'|$.

We note that random processes in integrable systems have been studied previously in a number of works, but these works were concerned with taking account of the influence of a small stochastic perturbation on the soliton solutions (see, for example, [6, 7]) or with the evolution of random initial data (see, for example, [8, 9]). The question of the appearance of a stochastic regime as a result of purely dynamical evolution of an integrable system with a uniquely determined deterministic initial potential $-u_0(x)$ is posed in the present paper (as well as in our preceding work [1]) for the first time.

It should be underscored that the analysis performed in [1] showed that with time a developing system of interacting solitons arrives in a state where the semiclassical multimodal Witham structure breaks down as a result of phase mixing. This still does not prove completely the appearance of a stochastic state of solitons: Additional analysis based on the solution of the scatter-

ing problem for the Schrödinger equation is required. This solution will be presented below. No less important is the fact that the specific construction of a solution in the statistical limit is a very complicated problem. In [1] the construction is obtained only in the degenerate case, when the initial function $u_0(x)$is a sequence of extremely widely spaced semiclassical pulses. The density of solitons formed as a result of the decay of the initial pulses approaches zero. For this reason, to a first approximation, solitons can be treated as free and noninteracting, which makes it possible to investigate the problem as the statistics of free solitons. Degeneracy is removed when the density of solitons is finite. This results in fundamental changes in the structure of the phase space. The interaction between solitons now plays a determining role. The present paper is devoted to the specific construction of the asymptotic statistical limit, developing from a smooth semiclassical initial function $u_0(x)$ with an arbitrary density of solitons.

The paper is organized as follows. In Section 2 a general formulation of the problem is given and the process leading to the development of a smooth deterministic initial state in time is described. It is shown that small-scale nonlinear oscillations in the form of an infinite system of solitons arise and a gradual stochastization of these oscillations occurs. In Section 3 the complete system of integral conservation laws, which determine the properties of the asymptotic solution, is formulated. In Section 4 the solution of the scattering problem for the Schrödinger equation is analyzed. It is shown that for the initial potential under study the scattering phases are distributed randomly and uniformly, which rules out the possibility of a dynamical description of the system in the asymptotic limit $t \longrightarrow \infty$. In Section 5 a method of constructing the maximizer for a system of interacting solitons, defined on a certain scale L and possessing uniformly distributed random phases, is developed on the basis of the known N-soliton solution. The limit with a fixed dimensionless soliton density $C = \varepsilon N/L$ is studied in the limits $N \longrightarrow \infty$ and $L \longrightarrow \infty$. It is shown that the condition for the existence of a maximizer with a fixed density C leads to a linear integral equation. In Section 6 a solution of this equation is found and the spectrum of solitons in a stochastic state is determined. The conservation laws play the main role in determining the final form of the spectrum. The new concept of a quasisoliton is introduced in Section 7. This object is of fundamental significance in the stochastic state of the system: It plays a role similar to that of an ordinary soliton in the dynamical state. In Section 8 the probability density $f(u)$ is determined on the basis of the soliton spectrum obtained. A general computational procedure is indicated. Specific calculations are performed using a maximizer, which makes it possible to develop an algorithm for finding the N-soliton solution rapidly for large values of N. It is shown that the distribution $f(u)$ obtained after averaging over the initial phases does not depend on N, but rather it is determined only by the properties of the initial semiclassical function $u_0(x)$.

The theory constructed actually describes the appearance and development of a turbulent state in one-dimensional dispersion hydrodynamics. The statistical properties of developed turbulence are determined completely by the initial large-scale hydrodynamic flow. A method making it possible to determine the characteristics of turbulence on the basis of a prescribed initial state is developed. Here the finiteness of the dispersion parameter ε plays a fundamental role, though the value of ε itself can be as small as desired.

2. FORMULATION OF THE PROBLEM

We shall study the solution of the KdV equation in the semiclassical limit (1) and (2). The initial function $u_0(x)$ is a smooth, bounded, nonperiodic, oscillating function on the entire x axis. The detailed conditions imposed on the class of ergodic functions $u_0(x)$ are presented in [1]. As an example, we shall consider an initial deterministic function $u_0(x)$ in the form of an infinite sequence of nonoverlapping pulses $F_i(x)$:

$$u(x, 0) = u_0(x) = \sum_i F_i\left(\frac{x - x_i}{\theta_i}\right). \tag{3}$$

Here θ_i is the width of the pulse and the point x_i determines the position of the maximum of the ith pulse. We shall assume that at the maximum $F_i(0) = 1$; then $0 \le F_i \le 1$. The average distance between the neighboring points x_i is l, and the ratio of the average width θ of the pulse to the distance between the pulses is

$$\gamma = \frac{\theta}{l} < 1. \tag{4}$$

The quantity γ is a characteristic parameter of the initial problem under study, and the last condition takes into account the fact that the pulses do not overlap.

The process of the temporal development of the initial function $u_0(x)$ (3) has been studied in [1] in accordance with the KdV dynamics. We shall recall its basic qualitative features. First, the highest-order derivatives in the KdV equation are not essential, and the solution in each of the pulses develops according to the Hopf dynamics. When the special toppling points are reached, single-mode (three-zone) fine-scale oscillations, described by Witham's equations [4, 10–12], arise near these points. The regions occupied by the oscillations gradually expand and then overlap. Overlapping occurs for $t \ge l$ (since the maximum velocity $u_{0\max} \sim 1$) and is the result of the superposition of waves generated in two neighboring pulses. As a result, regions of two-mode (five-zone) oscillations, which can be represented as two interacting nonlinear waves [13, 14], arise. As time passes, for $t > nl$, the waves which are now generated in a large number n of initial

pulses overlap. As a result, the number of forbidden and allowed zones on the Riemann surface increases, and the solution is now described by multimodal (multizone) nonlinear Witham waves [1, 2, 5, 15–18]. We note that the appearance of new modes in a Witham system is directly related with the number of maxima of the semiclassical initial function $u_0(x)$. In the asymptotic limit $t \longrightarrow \infty$, the number of modes is equal exactly to the number of initial maxima. Of course, the initial spectrum of the Schrödinger equation is completely preserved.

In our case the number P of initial pulses is infinite and correspondingly, with time the number of Witham modes grows with no limit. It is important that the width of the allowed zones decreases exponentially with time

$$\Delta_{bi} \propto \exp(-t/\tau_i).$$

As a result, each mode extends into a chain of solitons. The structure of the oscillations becomes extremely complicated, but once again they retain their dynamical character—the law of conservation of the number of waves holds for each mode. This is valid as long as the number n of modes is not too large so that the width Δ_{gi} of the forbidden zones is much greater than the dispersion parameter ε. In this case the solitons belonging to a given mode are separated by a distance that is much less than the scale of the oscillations of the initial function $u_0(x)$. This makes it possible, on account of the interaction of the solitons, to retain definite phase relations. However, according to the dynamics [1] of the oscillating initial function $u_0(x)$ defined on the entire axis, the widths of the forbidden zones decrease with time:

$$\Delta_{gi} \propto 1/t.$$

For this reason, for large values of $t \geq 1/\varepsilon$ the condition $\Delta_{gi} < \varepsilon$ is always satisfied. According to [3], the Witham approximation breaks down in this case.

We shall now discuss in greater detail the process leading to the destruction of the dynamics of the soliton phases. Even when the first oscillatory zone arises, phase interruption appears at singular points on its boundaries—for example, the position of the first soliton in the Witham system is not precisely determined [4, 2, 19]. This phase interruption arises at the moment oscillations are generated on passage of a singular point of the Hopf equation. The second mode is generated after the singularity in the single-mode solution is reached [14]. Phase interruption of the oscillations also arises at the moment of passage of this singularity. For this reason, although all phases are matched within each of the two interacting modes, a phase uncertainty is present between the modes. The same thing also happens with the generation of the third, fourth, fifth, and other modes. As a result, by the time the number n of modes exceeds the number of solitons on the characteristic scale of the oscillations of the initial function $u_0(x)$ ($n \sim t \gg 1/\varepsilon$), these solitons all belong to different modes, and therefore it can be assumed that their phases will be completely mismatched (random).

Thus, the analysis performed above shows that the development of an asymptotic solution of the KdV equation (1) and (2) for an oscillatory nonperiodic function $u_0(x)$ defined on the entire x axis can lead to the appearance of a stochastic state, i.e., a state with random, uniformly mixed soliton phases. The conditions for the appearance of this state have the form

$$\varepsilon \ll 1, \quad \varepsilon t \longrightarrow \infty, \quad \varepsilon P \longrightarrow \infty. \tag{5}$$

Here P is the number of humps in the initial function $u_0(x)$. Specifically, if the large-scale oscillations of the initial function $u_0(x)$ are themselves random, then the appearance of a stochastic state of a system of solitons in the asymptotic limit (5) is obvious. However, it can be inferred on the basis of the argument presented above that the same state also appears for a strictly deterministic initial function. The latter assumption requires detailed substantiation, which will be presented below on the basis of an investigation of the scattering problem for the Schrödinger equation.

We underscore that ordinarily solutions of the KdV equation are studied in the different limiting case of small dispersion [3, 10–18]: $\varepsilon \longrightarrow 0$ with the additional condition $\varepsilon t \longrightarrow 0$ or $\varepsilon P \longrightarrow 0$. This limit is of a purely dynamical character and therefore differs fundamentally from the statistical limit which we have examined and which arises for arbitrarily small but finite values of ε in the asymptotic limits $t \longrightarrow \infty$ and $P \longrightarrow \infty$.

3. INTEGRAL FORM OF THE CONSERVATION LAWS

We shall describe some general properties of the initial problem under study. Let the spectrum of the Schrödinger equation for a given initial semiclassical potential $-u_0(x)$ be known (as usual, we neglect the continuous part of the spectrum in the semiclassical approximation [2]). We shall assume that for a bounded oscillatory nonperiodic function $u_0(x)$ (the detailed conditions imposed on the class of ergodic functions $u_0(x)$ are presented in [1]) the spectrum can be described by the density $\rho(k)$ of the distribution over the wave numbers. Let n be the number of spectral levels $\lambda = -k^2$ per unit length. On account of the ergodic nature of the function $u_0(x)$, n does not depend on x. The levels present in the interval from k to $k + dk$ can be separated from these n levels:

$$dn = n\rho(k)dk. \tag{6}$$

It is convenient to fix the interval on which the wave numbers are defined, as $0 \leq k \leq 1$ and to normalize the spectral density according to the relation

$$\int_0^1 \rho(k)dk = 1. \tag{7}$$

We underscore that the exact determination of the spectrum of the Schrödinger equation for a nonperiodic oscillatory potential $-u_0(x)$ defined on the entire x axis is a very difficult mathematical problem [20]. In our case the problem simplifies because we are interested not in the detailed structure of the spectrum but only the completely discrete levels $\rho(k)$ in the semiclassical limit (2). In this case the density $\rho(k)$ can be determined as the spectrum of the potential $-u_0(x)$ on a finite segment $L \gg L_0$, where L_0 is the characteristic scale of uniformity of the function $u_0(x)$. Since the initial function is ergodic, it can be assumed that the normalized density $\rho(k)$ approaches a completely determined function in the limit $L \longrightarrow \infty$.

The spatial density of the discrete levels C is determined simultaneously with the spectral density $\rho(k)$. It is convenient to write it in a dimensionless form, normalized to the dispersion parameter ε,

$$C = \varepsilon\frac{N}{L}, \tag{8}$$

where N is the total number of levels on the characteristic scale L.

According to the semiclassical Bohr–Sommerfeld formula, the spectral density can be represented in the form

$$\rho(k) = \frac{k}{\pi C}\int\limits_{\xi_{i-}(k)}^{\xi_{i+}(k)}\frac{1}{L}\sum_{i=1}^{N}\frac{d\xi}{\sqrt{F_i(\xi)-k^2}}, \tag{9}$$
$$L \longrightarrow \infty, \quad N \longrightarrow \infty.$$

Here $\xi_{i\pm}(k)$ are the roots of the equation $F_i(\xi) = k^2$, and the constant C is determined by the normalization condition (7).

Specifically, if all pulses have the same form, for example,

$$F_i(\xi) = 1 - \frac{\xi^2}{(\theta_i/2)^2}, \tag{10}$$

then we find

$$\xi_{i\pm}(k) = \pm\frac{\theta_i}{2}\sqrt{1-k^2},$$

and, correspondingly, the spectral density of the levels is

$$\rho(k) = 2k. \tag{11}$$

The normalization (7) is used in the last expression. The dimensionless level density (8) is

$$C = \varepsilon\frac{N}{L} = \frac{\theta}{4l}. \tag{12}$$

The simple example of a nonperiodic function (3), (10) can be obtained by assuming that l is constant and modeling the width θ_i by a periodic function which is incommensurate with l:

$$\theta_i = \theta(1 + q\sin(\sigma x_i/l)),$$
$$q < 1, \quad \sigma \neq 2\pi\frac{n}{m}. \tag{13}$$

We shall consider this case below as a specific example. We note that the initial function (3), (10) has weak discontinuities at the points $x = x_i \pm \theta_i$. This feature affects the behavior of the spectrum $\rho(k) \propto k$ as $k \longrightarrow 0$. For a smooth function $u_0(x)$, which does not contain weak discontinuities, the asymptotic behavior in the limit $k \longrightarrow 0$ is different: $\rho(k) \propto k\ln(1/k)$.

We underscore that, as already indicated in [1], for hydrodynamic (semiclassical) initial problems the dimensionless (i.e., normalized to the dispersion parameter ε) soliton density is very low: $C \leq 0.3$. Thus, for the case of nonoverlapping parabolic pulses (10) considered above, the maximum value $\gamma = 1$ and, corresponding, $C_{\max} = 0.25$.

As is well known, the KdV equation has an infinite number of conservation laws [2]. However, for the initial problems of the semiclassical (hydrodynamic) type only three of these laws play a determining role. Specifically, Witham's system of equations has been obtained on the basis of three conservation laws [21]. The two laws,

$$u_t + (3u^2 + \varepsilon^2 u_{xx})_x = 0 \tag{14}$$

and

$$(u^2)_t + (4u^3 + \varepsilon^2(2uu_x - u_x^2))_x = 0, \tag{15}$$

are actually completely unrelated with small-scale oscillations. They express only the conservation of the average values $\langle u\rangle$ and $\langle u^2\rangle$ of a smooth initial large-scale function $u_0(x)$. Here the higherorder derivatives in the initial problem play only a small role because of the dispersion parameter ε is small. Conversely, the third conservation law describes the conservation of the number of small-scale oscillations which depends directly on the small parameter ε:

$$k_t + \omega_x = 0. \tag{16}$$

Here k is the wave number and ω is the frequency of the oscillations.

We note that all other higher-order KdV conservation laws can be represented in the form

$$\partial_t(u^n + O(\varepsilon^2)) + \partial_x P_n(u, u_x, \ldots) = 0. \tag{17}$$

For example, the fourth conservation law has the form

$$\partial_t(2u^3 - \varepsilon^2 u_x^2)$$
$$+ \partial_x(9u^4 + \varepsilon^2(6u^2u_{xx} + 2u_tu_x) + \varepsilon^4 u_{xx}^2) = 0.$$

For the problems of interest to us it is natural to switch to the integral form of the conservation laws.

Since initial functions $u_0(x)$ which are bounded and ergodic [22], i.e., uniform on a quite large scale L, are studied here, according to the conservation laws (14)–(16), the average value of the velocity, the mean-square value of the velocity, and the average density of discrete levels should be conserved at any moment in time t:

$$\langle u\rangle = \lim_{L\to\infty}\frac{1}{L}\int_{-L/2}^{L/2} u\,dx, \tag{18}$$

$$\langle u^2\rangle = \lim_{L\to\infty}\frac{1}{L}\int_{-L/2}^{L/2} u^2\,dx, \tag{19}$$

$$C = \varepsilon\left\langle\frac{k}{2\pi}\right\rangle = \lim_{L\to\infty}\frac{\varepsilon}{L}\int_{-L/2}^{L/2}\frac{k}{2\pi}dx. \tag{20}$$

Indeed, integrating equations (14)–(16) over dx in the interval $[-L/2, L/2]$ and dividing by L we obtain

$$\frac{1}{L}\int_{-L/2}^{L/2} u_t dx = \frac{1}{L}(3u^2+\varepsilon^2 u_{xx})_{-L/2}^{L/2},$$

$$\frac{1}{L}\int_{-L/2}^{L/2}(u^2)_t dx = \frac{1}{L}(4u^3+\varepsilon^2[2uu_x-u_x^2])_{-L/2}^{L/2},$$

$$\frac{\varepsilon}{L}\int_{-L/2}^{L/2}\left(\frac{k}{2\pi}\right)_t dx = \frac{\varepsilon}{2\pi L}(\omega)_{-L/2}^{L/2}.$$

Passing to the limit of infinite L and taking account of the bounded nature of the function $u(x, t)$ and its derivatives, we find that

$$\frac{d}{dt}\langle u\rangle = 0,\quad \frac{d}{dt}\langle u^2\rangle = 0,\quad \frac{d}{dt}C = 0,$$

and this means that the integrals (18)–(20) exist.

Thus, the average values C, $\langle u\rangle$, and $\langle u^2\rangle$ at any moment in time are determined by the form of the initial function $u_0(x)$. It is easy to show that all other integrals of motion (17) are determined initially to within a small correction of order ε^2 by the averages of powers of the velocity:

$$C_n = \langle u^n\rangle + Q(\varepsilon^2),\quad n = 3, 4, \ldots. \tag{21}$$

It is important to note that the entire collection of averages $\langle u^n\rangle$ is initially uniquely related with the average density C and the spectral density $\rho(k)$. Indeed, let us consider for simplicity a function $u_0(x)$ in the form of an infinite sequence of nonoverlapping pulses (3) and let us use the Bohr–Sommerfeld formula (9) for $\rho(k)$. To determine the relation between the level density (6) and the average values (18)–(21) let us consider an integral of the form

$$J_n = \int_0^1 k^{2n-1}\rho(k)dk.$$

Substituting the expression (9) into this integral we obtain

$$J_n = \pi^{-1}C^{-1}\int_0^1 k^{2n}dk\int_{\xi_-(k)}^{\xi_+(k)}\frac{1}{L}\sum_{i=1}^{N}\frac{d\xi}{\sqrt{F_i(\xi)-k^2}}.$$

Changing the limits of integration and performing the integration over dk, we obtain

$$J_n = \frac{1}{\pi CL}\sum_{i=1}^{N}\int_{-L/2}^{L/2}[F_i(\xi)]^n d\xi$$
$$\times\int_0^1(1-s^2)^{-1/2}s^{2n}ds = \frac{\langle u^n\rangle}{S_nC}, \tag{22}$$
$$S_n = \frac{2^{n+1}n!}{(2n-1)!!}.$$

The desired relation can be seen from the expression (22), specifically,

$$\langle u\rangle = 4C\int_0^1 k\rho(k)dk, \tag{23}$$

$$\langle u^2\rangle = \frac{16}{3}C\int_0^1 k^3\rho(k)dk. \tag{24}$$

For example, for the above-considered sequence of nonoverlapping parabolic pulses (10) of width θ with distance l between the centers we have

$$\langle u\rangle = \frac{1}{l}\int_{-\theta/2}^{\theta/2}\left[1-\left(\frac{2x}{\theta}\right)^2\right]dx = \frac{2\theta}{3l}, \tag{25}$$

$$\langle u^2\rangle = \frac{1}{l}\int_{-\theta/2}^{\theta/2}\left[1-\left(\frac{2x}{\theta}\right)^2\right]^2 dx = \frac{8\theta}{15l}. \tag{26}$$

Using equations (11) and (12) it is easy to show that the relations (23), (24), and (25), (26) are identical.

The opposite assertion is also valid: knowing all the averages, the spectral level density $\rho(k)$ can be reconstructed uniquely. Indeed, let us take account of the fact that the function $\rho(k)$ describes the density of the spectral levels of the Schrödinger equation $\lambda = -k^2$. For this reason, in accordance with equation (9), it is an odd

function of k.[2] Let us represent the function $\rho(k)$ in the form of a series in the odd Legendre polynomials:

$$\rho(k) = \sum_{n=0}^{\infty} a_n P_{2n+1}(k). \tag{27}$$

Multiplying equation (27) by k^{2m+1}, integrating over dk from –1 to 1, and comparing the expressions (18)–(22), we obtain a system of algebraic equations from which the coefficients a_n can be successively determined:

$$\begin{aligned} a_0 &= \frac{3}{4C}\langle u\rangle, \\ a_1 &= \frac{105}{32C}\left(\langle u^2\rangle - \frac{4}{5}\langle u\rangle\right), \ldots. \end{aligned} \tag{28}$$

The coefficients a_i are related, in addition, by the normalization condition (7). Thus, in the semiclassical limit there is a one-to-one correspondence between the spectral density of discrete levels and the KdV system of integral conservation laws.

We shall take account of one other interesting circumstance. The structure of the discrete levels does not change with time—this property reflects the conservation law for the soliton amplitudes when the solitons interact with one another. But, this means that the integral conservation laws can also be represented in the differential spectral form

$$dC_n = CS_n k^{2n-1}\rho(k)dk,$$

where C is the spatial density of discrete levels. Specifically,

$$d\langle u\rangle = 4Ck\rho(k)dk. \tag{29}$$

4. STATISTICAL LIMIT IN THE SCATTERING PROBLEM

We shall consider an initial function of the form (3) and investigate the scattering problem for the potential $U(x) = -u_0(x)$ in the semiclassical limit. To solve this problem we need to find the discrete spectrum k_n of the Schrödinger operator

$$L = -\varepsilon^2\partial_{xx} + U(x)$$

(the parameters k_n are related with the eigenvalues λ_n by the relation $k_n = \sqrt{-\lambda_n}$) and the shift parameters ("phases") s_n. Since we are studying the semiclassical potential (3), in which there are no resonances because the form of the wells varies quasiperiodically,

$$\lambda_i \neq \lambda_j, \tag{30}$$

the spectrum is, to exponential accuracy, a sum of the spectra of each well which are found from the Bohr–Sommerfeld quantization conditions:

$$\begin{aligned} \cos\omega_j = 0, \quad \omega_j &= \frac{1}{\varepsilon}\int_{x_{2j-1}}^{x_{2j}} |p(x)|dx, \\ p(x) &= \sqrt{\lambda - U(x)}. \end{aligned} \tag{31}$$

Here, as usual, x_r are the turning points determined by the vanishing of the momentum in each well:

$$p(x_r) = 0, \quad r = 1, \ldots, 2N.$$

To determine the region of applicability of our analysis we shall calculate the correction λ_n^1 to the eigenvalues (31) which is due to the interaction with other wells. According to the general rules of quantum mechanics [23], we have

$$\lambda_n^1 = V_{nn}, \quad V_{mn} = \int \psi_m^{(0)*} V(x)\psi_n^{(0)}dx,$$

where $V(x)$ is a sum of all wells except the well to whose level we are seeking a correction, and $\psi_n^{(0)}$ is the eigenfunction in the absence of a perturbation. We shall estimate the correction λ_n^1 for each well. For this, we take account of the fact that the unperturbed eigenfunction decays exponentially in a classically inaccessible region:

$$\psi_n^{(0)} \propto \exp(-k_n x/\varepsilon).$$

Hence we obtain the estimate

$$\lambda_n^1 \approx \sum_{j=1}^{N} U\theta_j \exp\left(-\frac{k_n l_j}{\varepsilon}\right) \approx U\theta \exp\left(-\frac{k_n l}{\varepsilon}\right).$$

Here θ_j is the characteristic width of the jth well, l_j is the distance between the jth well and the well under study, l and θ are the characteristic spacing and the average width of the wells (see equations (3) and (4)). The correction $\psi^{(1)}$ for the wave function has the form [23]

$$\psi_n^{(1)} = \sum_{m\neq n} \frac{V_{mn}}{\lambda_n - \lambda_m}\psi_m^{(0)}.$$

Thus, in our case the corrections to the spectrum and the wave function are small. As is well known, the condition of applicability of perturbation theory is determined by the relation [23]

$$|V_{nm}| \ll |\lambda_n - \lambda_m|. \tag{32}$$

Since

$$|\lambda_n - \lambda_m| \propto \varepsilon,$$

and the matrix element

$$|V_{nm}| \propto \exp[-(k_m + k_n)l/\varepsilon],$$

[2] We note that the even continuation of the function $\rho(k)$ into the region $k < 0$ leads to a singularity at the point $k = 0$.

the relation (32) is always satisfied well in our case when

$$l \gg \frac{\varepsilon|\ln\varepsilon|}{k_m + k_n}.$$

This means that the condition (32) holds for all levels except a small fraction of the low-energy levels (of the order of ε), for which

$$k_n < K_0 \approx \varepsilon|\ln\varepsilon|. \tag{33}$$

We shall now determine the shift parameters [2]

$$s_n = \frac{\varepsilon}{2k_n}\ln|b_n|, \tag{34}$$

where $b_n = C_{n+}/C_{n-}$ is the ratio of the coefficients in the asymptotic eigenfunctions ψ_n as $x \longrightarrow \pm\infty$. We shall first find the parameters s_n for one potential well, and then we shall determine their shift as a result of the effect of N wells. We shall use the semiclassical solution in which the lower limit of integration is chosen at the turning point x_r:

$$\psi_r^{\pm} \propto \frac{1}{\sqrt{p(x)}}\exp\left(\pm\frac{i}{\varepsilon}\int_{x_r}^{x} p(x)dx\right).$$

We choose the branch of the root around a circuit in the upper half-plane in accordance with the scheme shown in Fig. 1a. The figure also shows the Stokes lines on which the asymptotic behavior of the exponentially decreasing solution must be changed. We choose to the left of the point x_1 a solution in the form

$$\psi(x) = \psi_1^-(x) = \frac{1}{\sqrt{p(x)}}\exp\left(-\frac{i}{\varepsilon}\int_{x_1}^{x} p(x)dx\right)$$

$$= \frac{e^{-i\pi/4}}{\sqrt{|p(x)|}}\exp\left(\frac{1}{\varepsilon}\int_{x_1}^{x}|p(x)|dx\right).$$

This solution approaches zero as $x \longrightarrow -\infty$. Passing into the region (x_2, x_3) from the right-hand side of the first well and using the well-known rules for crossing the Stokes lines, we obtain

$$\psi(x) = e^{-i\omega_1}\psi_2^-(x) - 2i\cos(\omega_1)\psi_2^+(x).$$

To satisfy the condition $\psi(+\infty) = 0$, the coefficient of the WBK solution ψ_2^+, which grows as $x \longrightarrow +\infty$, must vanish. This condition gives the quantization rule (31). Comparing the asymptotic behaviors

$$\psi(x) = \psi_1^-(x) \text{ at } x \longrightarrow -\infty$$

and

$$\psi(x) = e^{-i\omega_1}\psi_2^- \text{ at } x \longrightarrow +\infty,$$

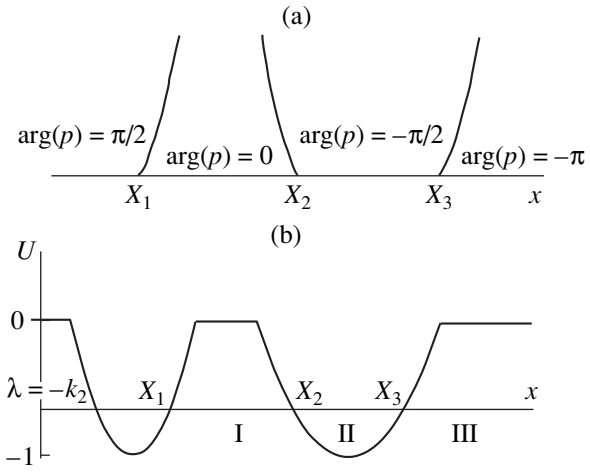

Fig. 1. (a) Scheme for choosing the branches of the root on a circuit in the upper half-plane. The Stokes lines are also shown. (b) Three regions singled out to clarify the change in phase on passage of the jth well.

we arrive at a relation for the shifts (as a simplification, we omit the index n):

$$b = \frac{C_+}{C_-}$$

$$= (-1)^n\exp\frac{1}{\varepsilon}\left[k(x_1 + x_2) + \left(\int_{x_1}^{-\infty} + \int_{x_2}^{+\infty}\right)(k - |p|)dx\right],$$

$$s = \frac{\varepsilon}{2k}\ln|b| = \frac{1}{2}(x_1 + x_2)$$

$$+ \frac{1}{2k}\left(\int_{x_1}^{-\infty} + \int_{x_2}^{+\infty}\right)(k - |p|)dx.$$

We shall now determine the phase shift for the general case of N wells. For this, we shall see how the phase of the wavefunction, localized in the well under study changes on passage of the jth well to which the turning points x_r and x_{r+1} $(r = 2j - 1)$ refer. For passage of this well, we single out three regions, as shown in Fig. 1b. In the region I we have a solution with amplitude A_{r-1} which decays exponentially from the initial wave function:

$$\psi_{\mathrm{I}} = \frac{A_{r-1}}{\sqrt{p}}\exp\left(-\frac{1}{\varepsilon}\int_{X_1}^{x}|p|dx\right). \tag{35}$$

In the region II the wave function has the form

$$\psi_{\mathrm{II}} = \frac{A_r}{\sqrt{p}}\exp\left(\frac{i}{\varepsilon}\int_{X_1}^{x}|p|dx\right) + \frac{B_r}{\sqrt{p}}\exp\left(-\frac{i}{\varepsilon}\int_{X_1}^{x}|p|dx\right). \tag{36}$$

Passing in the complex plane above the point X_2, we find from equations (35) and (36) the following relation between the coefficients:

$$A_r = A_{r-1}\exp\left(i\frac{\pi}{4} - \frac{1}{\varepsilon}\int_{X_1}^{X_2}|p|dx\right).$$

Hence, right up to the point X_3, we have the solution

$$\psi_{\mathrm{II}}^{+} = A_{r-1}\exp\left(i\frac{\pi}{4} - \frac{1}{\varepsilon}\int_{X_1}^{X_2}|p|dx\right)\exp\left(\frac{i}{\varepsilon}\int_{X_2}^{x}|p|dx\right).$$

Similarly, passing around the point X_3, we obtain

$$\begin{aligned} A_{r+1}\exp\left(-\frac{1}{\varepsilon}\int_{X_3}^{x}|p|dx\right) &= A_{r-1}\exp\left(i\frac{\pi}{4} - \frac{1}{\varepsilon}\int_{X_1}^{X_2}|p|dx\right) \\ &\times \exp\left(\frac{i}{\varepsilon}\int_{X_2}^{X_3}|p|dx\right)\exp\left(-\frac{1}{\varepsilon}\int_{X_3}^{x}|p|dx\right). \end{aligned} \tag{37}$$

To determine the shifts we are interested only in the modulus of the amplitude ratio. Proceeding from equation (37), we obtain a relation between the amplitudes on passage of the jth potential well, whose levels, according to equation (30), do not coincide with those of the first well:

$$\left|\frac{A_{r+1}}{A_{r-1}}\right| = \exp\left(-\frac{1}{\varepsilon}\int_{X_1}^{X_2}|p|dx\right).$$

Repeating this procedure N times and taking account of, in addition, the integrals from the "terminal" segments $(-\infty, x_1)$ and $(x_N, +\infty)$, we obtain

$$\begin{aligned} \left|\frac{C_+}{C_-}\right| = \exp\left\{\frac{k}{\varepsilon}(x_1 + x_N) - \frac{1}{\varepsilon}\sum_{j=1}^{N-1}\int_{x_{2j}}^{x_{2j+1}}|p|dx \right. \\ \left. + \frac{1}{\varepsilon}\left(\int_{x_1}^{-\infty} + \int_{x_N}^{+\infty}\right)(k - |p|)dx\right\}. \end{aligned} \tag{38}$$

The integrals in the expression (38) extend over the regions where the wavefunction decays exponentially. Using equations (38) and (34), we obtain finally

$$\begin{aligned} s_n(k, N) = \frac{1}{2}(x_1 + x_N) - \frac{1}{2k}\sum_{j=1}^{N-1}\int_{x_{2j}}^{x_{2j+1}}|p|dx \\ + \frac{1}{2k}\left(\int_{x_1}^{-\infty} + \int_{x_N}^{+\infty}\right)(k - |p|)dx. \end{aligned} \tag{39}$$

It is important to underscore that the minus sign in front of the sum in the expressions (38) and (39) is determined by the fact that we have found corrections to the shifts from the wells located to the right of the well under study. If a similar procedure is performed for wells on the left side, then the sign changes to plus.

We shall now calculate the shifts s_n for a parabolic potential of the form (3), (10). Simple but quite laborious calculations of the integrals in equation (39) yield

$$\begin{aligned} s_m(N) &= \chi_m\sum_{j=1}^{N}\theta_j, \\ \chi_m &= \chi(k_m) \\ &= \left\{\frac{1}{2} - \sqrt{1-k_m^2} + \frac{1}{2}(1-k_m^2)\ln\frac{1+k_m}{\sqrt{1-k_m^2}}\right\}. \end{aligned} \tag{40}$$

Here θ_j are the half-widths of the parabolic pulses. We shall assume that they are modulated according to the law (13). To simplify the formulas we set $\sigma = 1$, since the result depends on the modulation parameter σ trivially. Let us examine the limit $N \longrightarrow \infty$. For this, we first investigate the "symmetric" limit, i.e., we add N wells to the right of the potential well under study and N wells to the left. We renumber the wells from left to right. Then, using equation (40), we find

$$s_m = \chi_m q\theta\left\{\sum_{j=N+1}^{2N}\sin j - \sum_{j=1}^{N}\sin j\right\}.$$

Calculating the sums, we obtain, finally,

$$\begin{aligned} s_m(N) &= q\theta\chi_m\left[2\sin\left(\frac{1}{2}\right)\right]^{-1}\sin\left(\frac{N}{2}\right)\sin\left(\frac{N+1}{2}\right) \\ &= q\theta\chi_m\left[2\sin\left(\frac{1}{2}\right)\right]^{-1}\left[\cos\left(\frac{1}{2}\right) - \cos\left(N+\frac{1}{2}\right)\right]. \end{aligned} \tag{41}$$

It is evident from equation (41) that the phase has no limit as $N \longrightarrow \infty$: as N increases, the phase s_m densely fills the interval

$$s_m \in \left(q\theta\chi_m\left[2\sin\left(\frac{1}{2}\right)\right]^{-1}\left[\cos\left(\frac{1}{2}\right) - 1\right],\right.$$
$$\left. q\theta\chi_m\left[2\sin\left(\frac{1}{2}\right)\right]^{-1}\left[\cos\left(\frac{1}{2}\right) + 1\right]\right).$$

The process of filling this interval is completely analogous to the process leading to the appearance of chaos in dynamical systems [24].

We now consider the asymmetric passage to the limit $N \longrightarrow \infty$, i.e., the situation where the number of wells on the left-hand side differs by M from the number of wells on the right-hand side. It is easy to show

that all calculations presented above are preserved, and the interval over which the phase is spread will be centered at the point

$$X^{(0)} = \frac{q\theta\chi_m}{2}\cot\left(\frac{1}{2}\right) + \chi_m\sum_{j=1}^{M}\theta_j.$$

Since the parameters θ_j are incommensurate with the period l on which the minima of the potential wells lie, for arbitrary M the phase uniformly covers the entire period $[0, l]$.

Thus, for an infinite number of potential wells the scattering phase fills any segment $[l_j, l_{j+1}]$ randomly and uniformly.

The result obtained can be extended to any sequence of initial pulses of the form (3), including pulses of variable height. At the same time, we underscore that the narrow range of small values $k_n < K_0$ (33) does not conform to the indicated laws and can form a separate structure.

In concluding this section we call attention to the correspondence between the scattering problem studied here and the process of temporal mixing of a deterministic initial state, described in Section 2. Each potential well generates an individual mode of a Witham system. The appearance of a new mode, as noted above, leads to phase interruption in a soliton chain, and this interruption corresponds to the contribution of a new potential well in the scattering problem. It is important that the phase interruption does not depend on the time when a new mode appears. This corresponds completely to the fact that the contribution to the scattering phase does not depend on the distance between two given potential wells and the number of wells in the interval between them. Thus, the above investigation of the scattering problem with an infinite number of potential wells can be taken as a proof of the fact that the function $u_0(x)$, according to KdV dynamics, decomposes asymptotically as $t \longrightarrow \infty$ into a system of solitons with random phases.

5. MAXIMIZER OF THE *N*-SOLITON SOLUTION WITH RANDOM PHASES

We shall now describe an infinite system of solitons with random phases. According to the methods of statistical physics [25], this requires separating an arbitrary subsystem from N solitons on some segment and then averaging over an ensemble of states with random phases s, assuming the number N and the segment length to be increasing with no limit in proportion to one another (the statistical limit). It is natural to use the well-known exact solution [26–28] for N interacting solitons:

$$u_N(x, t) = 2\varepsilon^2\partial_{xx}\ln\sum_{\mu}\exp[\Phi_\mu(x, t)]. \qquad (42)$$

Here $\mu = (\mu_1, \mu_2, \ldots, \mu_N)$, $\mu_n = 0$ or 1, and

$$\sum_{\mu} = \sum_{\mu_1=0}^{1}\cdots\sum_{\mu_N=0}^{1}. \qquad (43)$$

The argument of the exponential is a linear function of x and t:

$$\Phi_\mu(x, t) = 2\varepsilon^{-1}\sum_{n=1}^{N}\mu_n k_n(s_n + 4tk_n^2 - x) - \sum_{m=1}^{N}\sum_{n=1}^{N}\mu_m\mu_n l_{mn}, \qquad (44)$$

where

$$l_{mn} = \ln|(k_m + k_n)/(k_m - k_n)| \quad \text{at } m \neq n, \quad l_{nn} = 0.$$

The function $u_N(x, t)$ depends on $2N$ parameters: the wave vectors $k_1, \ldots, k_N$ and the phases $s_1, \ldots, s_N$.

The expressions (42)–(44) establish a unique correspondence between a set of points in S space with the coordinates

$$s_n(t) = s_n + 4tk_n^2,$$

which move uniformly, i.e., they do not interact, and a set of interacting solitons. It is natural to characterize the first set by its average density

$$C_s = \varepsilon N/L_s,$$

where L_s is the length of the interval containing the points s_n. Thus we arrive at the problem of the statistical properties of the function (42)–(44), in which the shift parameters $s_1, \ldots, s_N$ are random quantities which uniformly and independently distributed in an interval of length L_s, where L_s approaches infinity in proportion to N so that the ratio $N/L_s = \varepsilon^{-1}C_s$ is constant. We shall consider the interval $(-L_s/2, L_s/2)$, so that the possible edge effects due to the ends of the interval would have no effect (at $L_s \longrightarrow \infty$) in any predetermined bounded region on the x axis. The velocity $v_s = 4k^2$ of the points in S space is determined by the parameter k, which varies over the same interval as for solitons $0 \le k \le 1$. The spectrum of solitons over k is given by the known function $\rho(k)$, and the distribution $\rho_s(k)$ of the points of the S-set over k is unknown—we must find this distribution. The normalization of the function $\rho_s(k)$ is determined, just as in equation (7), by the condition

$$\int_0^1\rho_s(k)dk = 1. \qquad (45)$$

This means that the densities C_s and C are related by the relation

$$C_s = C\frac{L}{L_s}, \quad CL = \varepsilon N, \tag{46}$$

where L is the scale on the x axis, on which the N interacting solitons are distributed. Here the law of conservation of the number of solitons (20) and the unique correspondence between the solitons and the points of the S-set is taken into account.

The expression (42)–(44) is exact: it takes account of the interaction of all N solitons with one another. This is why the formulas (42)–(44) are so complicated and the number of terms is so large ($\sim 2^N$). However, it is obvious that most of them are unessential: only a small number of terms, including the maximum terms, determining the interaction of the group of closest solitons, plays the main role in the sum. Our problem is to find the maximum of the function Φ_μ by varying the numbers μ_n (44) with a fixed set of values of k_n and s_n for the N-soliton solution and then passing to the continuous limit. This method is essentially similar to the method developed previously by Lax and Levermore [3] (see also Mazur [29]). The fundamental difference lies in the fact that in [3, 29] the maximizer was constructed under the conditions of a uniquely prescribed functional relation between the values of the vectors k_n and their phases s_n and was a completely determinate dynamical problem. In our case, however, there is no dynamical relation between k_n and s_n: The parameters $s_1, \ldots, s_n$ are, on the contrary, a set of random numbers. Statistical methods must now be used to determine the maximizer. We note that the N-soliton solution (42)–(44) is valid for a potential decreasing rapidly as $|x| \longrightarrow \infty$ [3]. For this reason, the very possibility of using this solution to describe a statistical subsystem with a nondecreasing potential $-u_0(x)$ is a hypothesis which will be substantiated below.

In analyzing the problem of finding the maximum of Φ_μ, we shall employ an analogy with the simplified model $\Phi_{\mu 0}$, obtained by dropping in equation (44) the terms which are quadratic in μ and which describe the interaction of the solitons. The maximum of the remaining part (linear in μ) can be found exactly and very easily: the quantity

$$\varepsilon\Phi_{\mu 0}(x, t) = 2\sum_{n=1}^{N} \mu_n k_n (s_n + 4tk_n^2 - x) \tag{47}$$

reaches a maximum for $\mu = \hat{\mu}$ such that

$$\hat{\mu}_n = \eta(s_n - a_n), \tag{48}$$

where $a_n = x - 4tk_n^2$, and the function $\eta(\xi) = 0$ for $\xi < 0$ and $\eta(\xi) = 1$ for $\xi > 0$. In other words, in the sum (47) it is sufficient to take $\mu_n = 0$ for $s_n < a_n$ and $\mu_n = 1$ for $s_n > a_n$, where $n = 1, \ldots, N$. The maximizer $\hat{\mu}$ depends on $s_1, \ldots, s_N$; i.e., it is a random quantity.

We shall use a construction similar to equation (48) (a step function of s_n) to find the maximizer of the complete expression (44), including the quadratic interaction terms. It turns out that this makes it possible to find an approximate (in the probabilistic sense) maximizer $\tilde{\mu}$, and the accuracy of this expression increases with N, which makes it possible in the limit $N \longrightarrow \infty$ to replace the exact maximizer by the approximate one.

We seek $\tilde{\mu}$ in the form

$$\tilde{\mu}_n = \eta(s_n - b_n), \tag{49}$$

where $-L_s/2 \le b_n \le L_s/2$ (there is no point in considering $b_n < -L_s/2$ or $b_n > L_s/2$, since s_n stays within the range $(-L_s/2, L_s/2)$). The average value of $\tilde{\mu}_n$ is equal to the probability that $\tilde{\mu}_n = 1$:

$$\bar{\tilde{\mu}}_n = (L_s/2 - b_n)/L_s = 1/2 - b_n/L_s. \tag{50}$$

Changing in the sequence $\tilde{\mu}$ one element $\tilde{\mu}_n$ gives the sequence μ

$$\mu_\nu = \tilde{\mu}_\nu + (1 - 2\tilde{\mu}_n)\delta_{n\nu}.$$

The function $\varepsilon\Phi_\mu$ with the indicated elementary change in μ acquires the increment

$$\Delta_n \equiv \varepsilon(\Phi_\mu - \Phi_{\tilde{\mu}}) = 2\sigma_n\left[k_n(s_n - a_n) - \varepsilon\sum_{\nu=1}^{N} l_{n\nu}\tilde{\mu}_\nu\right], \tag{51}$$

where $\sigma_n = 1 - 2\tilde{\mu}_n = (-1)^{\tilde{\mu}_n}$.

For the example considered above without an interaction, there is no sum over ν, and we have

$$\Delta_n = 2k_n\sigma_n(s_n - a_n). \tag{52}$$

Setting $b_n = a_n$ in equation (49), we obtain $\sigma_n = -\operatorname{sgn}(s_n - a_n)$ and, therefore,

$$\Delta_n = -2k_n|s_n - a_n|; \tag{53}$$

i.e., all Δ_n are negative, as should be for a maximum.

Separating from $\sum l_{n\nu}\tilde{\mu}_\nu$ its average value (which, according to equation (50), can be simply expressed in terms of $b_1, b_2, \ldots$), equation (51) can be put into the form

$$\Delta_n = 2k_n\sigma_n \times\left[s_n - a_n - \frac{\varepsilon}{k_n}\sum_{\nu=1}^{N} l_{n\nu}\left(\frac{1}{2} - \frac{b_\nu}{L_s}\right) + \frac{\varepsilon}{k_n}\sum_{\nu=1}^{N} l_{n\nu}(\bar{\tilde{\mu}}_\nu - \tilde{\mu}_\nu)\right]. \tag{54}$$

Comparing this relation with equation (52) shows that if the numbers b_n are made to satisfy the conditions

$$a_n + \frac{\varepsilon}{k_n}\sum_{\nu=1}^{N} l_{n\nu}\left(\frac{1}{2} - \frac{b_\nu}{L_s}\right) = b_n, \quad n = 1, \ldots, N, \tag{55}$$

which comprise a system of N linear equations for N unknowns b_n, the expression (54) becomes (compare equation (53))

$$\Delta_n = -2k_n|s_n - b_n| + 2\varepsilon\sigma_n\sum_{\nu=1}^{N} l_{n\nu}(\tilde{\bar{\mu}}_\nu - \tilde{\mu}_\nu). \tag{56}$$

Hence it is evident that the unremovable random correction, which depends on all s_ν with $\nu \neq n$ and is zero on the average, makes it impossible to satisfy the conditions for a maximum $\Delta_n \leq 0$.

An estimate of this correction for $N \longrightarrow \infty$ is presented below. It shows that in the probabilistic sense the correction is negligibly small. Thus, the multi-index $\tilde{\mu}$, which can be efficiently calculated in analytical form, is a good approximation (as $N \longrightarrow \infty$) for the exact maximizer $\hat{\mu}$, which cannot be found analytically.

An important aspect of the calculation of $\tilde{\mu}$ is solving the system of linear equations (55). It is better to interpret this system as a system for the unknowns $y_n = 1/2 - b_n/L_s \equiv \tilde{\bar{\mu}}_n$:

$$y_n = \frac{1}{2} - \frac{x}{L_s} + \frac{4t}{L_s}k_n^2 - \frac{\varepsilon}{L_s k_n}\sum_{\nu=1}^{N} l_{n\nu}y_\nu. \tag{57}$$

We also note the fact that although the formal solution of this system of equations can be found for any x and t and is a linear function of these variables, it makes sense only as long as $0 \leq y_n \leq 1$ for all n (the corresponding limitation for the system (55) is that $-L_s/2 \leq b_n \leq L_s/2$). Outside the limits of the corresponding region of the xt plane the system (57) must be modified somewhat.

In passing to the limit $N \longrightarrow \infty$ ($L_s = \varepsilon N C_s^{-1}$) in the system (57), $x/L_s = \xi$ and $t/L_s = \tau$ must be fixed. In this limit the unknowns y_n are made continuous, $y_n = y(k_n)$, and equation (57) gives an integral equation of the second kind for the function $y(k)$:

$$y(k) = \frac{1}{2} - \xi + 4\tau k^2 - \frac{C_s}{k}\int_0^1 l(k, k')\rho_s(k')y(k')dk', \tag{58}$$

where

$$l(k, k') = \ln|(k + k')/(k - k')|$$

and $\rho_s(k)$ is the level density in the S set. To obtain the general solution of this equation for arbitrary values of the parameters τ and ξ, it is sufficient to solve two specific integral equations

$$y_\alpha(k) = k^\alpha - \frac{C_s}{k}\int_0^1 l(k, k')\rho_s(k')y_\alpha(k')dk', \quad \alpha = 0, 2 \tag{59}$$

and form the linear combination

$$y(k) = \left(\frac{1}{2} - \xi\right)y_0(k) + 4\tau y_2(k). \tag{60}$$

The solution obtained makes sense only for ξ and τ such that $0 \leq y(k) \leq 1$.

We shall now clarify the sense in which $\tilde{\mu}$ is an approximate maximizer. According to equation (56), the increment Δ_n is a sum of two independent random quantities: a negative quantity

$$\Delta_n^{(\mathrm{neg})} = -2k_n|s_n - b_n|,$$

which is distributed piecewise-uniformly, adjoining zero in a region of length $\sim L_s$ on the left-hand side, with the probability density $(k_n L_s)^{-1}$, and a sign-alternating correction

$$\tilde{\Delta}_n = 2\varepsilon\sigma_n\sum_{\nu=1}^{N} l_{n\nu}(y_\nu - \tilde{\mu}_\nu).$$

The latter, being a sum of $N - 1$ independent random quantities, is distributed (in the limit $N \longrightarrow \infty$) according to a Gaussian distribution with zero mean. Its variance is

$$d_n = 4\varepsilon^2\sum_{\nu=1}^{N} (l_{n\nu})^2 y_\nu(1 - y_\nu).$$

Passing to the continuous limit, we rewrite d_n in the form

$$d_n = \varepsilon^2 N d(k_n),$$

$$d(k) = 4\int_0^1 [l(k, k')]^2\rho_s(k')y(k')[1 - y(k')]dk'.$$

Thus, while

$$\left|\Delta_n^{(\mathrm{neg})}\right| \sim L_s \sim \varepsilon N,$$

the correction $\tilde{\Delta}_n \propto \sqrt{d_n}$ has, on the average, the much smaller value $\sim\varepsilon\sqrt{N}$.

We are interested in positive values of Δ_n, which should not exist for deviations from an exact maximum. In the region $\Delta > 0$ the distribution density of Δ_n is given by the expression (a convolution of a step distribution $\Delta_n^{(\mathrm{neg})}$ and a Gaussian distribution $\tilde{\Delta}_n$)

$$w_n(\Delta) = \frac{1}{k_n L_s \sqrt{2\pi d_n}} \int_{-\infty}^{0} \exp[-(z-\Delta)^2/2d_n]dz,$$

whence it is easy to obtain the probability of a positive value of Δ_n

$$\int_0^\infty w_n(\Delta)d\Delta = \frac{\sqrt{d_n}}{k_n L_s \sqrt{2\pi}} = \frac{C_s}{\sqrt{2\pi N}} \frac{\sqrt{d(k_n)}}{k_n}.$$

This quantity approaches zero as $N \longrightarrow \infty$ and the average number of "incorrect" elements in $\tilde{\mu}$ (i.e., elements $\tilde{\mu}_n$ such that replacing them by $1 - \tilde{\mu}_n$ increases Φ_μ) is equal to

$$\frac{C_s}{\sqrt{2\pi}} \sqrt{N} \int_0^1 k^{-1} \sqrt{d(k)} \rho_s(k) dk$$

that is only an infinitesimal fraction of all N of its elements.

A "correction" at $\tilde{\mu}$ "incorrect" zeros and units gives a better approximate maximizer $\tilde{\mu}^{(1)}$, in which their number is of the order of $N^{1/4}$, and so on. Proceeding in this manner we arrive at the exact maximizer $\hat{\mu}$ (for finite N, in a finite number of steps). Based on this procedure of successive approximations, it can be shown that (in the limit as $N \longrightarrow \infty$)

$$\overline{\hat{\mu}_n} = \overline{\tilde{\mu}_n}. \tag{61}$$

Thus, the expressions (49) and (50) for sufficiently large N can be used as the maximizer and its average value.

We recall, however, that we still do not know the density C_s of points in S space or the spectral density $\rho_s(k)$. As a result, the expression for the maximizer is also undetermined. To find it, a relation must be established between C_s, $\rho_s(k)$ and C, $\rho(k)$, which are determined by the initial potential $-u_0(x)$.

6. CONTINUOUS SPECTRUM OF THE *S*-SET

The conservation laws (18)–(21) play an important role in the determination of the relation between the initial and phase-mixed states of the system. As already mentioned above, because of the conservation law for the number of solitons (20), the densities C and C_s are related by the relation (46). For this reason, we shall calculate the average $\bar{u}$ and the mean-square $\overline{u^2}$ in the stochastic state, using the expression obtained above for the maximizer.

It is convenient to express the averages $\bar{u}$ and $\overline{u^2}$ in terms of the derivatives ∂_x, ∂_t of the average quantity $w = \varepsilon\partial_x\Lambda$, where Λ is the logarithm in the expression (42):

$$\Lambda = \ln \sum_\mu \exp(\Phi_\mu).$$

Evidently,

$$\bar{u} = 2\varepsilon\partial_x\bar{w} \tag{62}$$

and, because of the conservation law (14),

$$\overline{u^2} = -\frac{1}{3}(2\varepsilon\partial_t\bar{w} + \varepsilon^2\partial_{xx}\bar{u}). \tag{63}$$

As we shall see below, $\partial_{xx}\bar{u} = 0$ in the region of the xt plane of interest to us. The calculation of the average of $w = \varepsilon\partial_x\Lambda$ reduces to averaging the quantity

$$\partial_x\hat{\Phi}(x) = \partial_x \max_\mu \Phi_\mu(x),$$

because, as shown below, $\overline{\partial_x\Lambda} = \overline{\partial_x\hat{\Phi}}$ in the limit $N \longrightarrow \infty$. However, the average of $\overline{\partial_x\hat{\Phi}}$ can be easily found, since the approximate maximizer $\tilde{\mu}$ found in the preceding section can be used for this.

The difference

$$D(x) = \Lambda - \hat{\Phi} = \ln\left(1 + \sum_{\mu \neq \hat{\mu}} e^{-(\hat{\Phi} - \Phi_\mu)}\right)$$

is a positive continuous piecewise-smooth function of x, uniformly bounded above as $N \longrightarrow \infty$. For this reason, if $L_s \propto N \longrightarrow \infty$, then the derivative $\partial_x D$ can be averaged over an interval $x_1 < x < x_2$ whose length approaches ∞, comprising only a small part of the entire length L_s. In the limit we obtain

$$\overline{\partial_x D} = \frac{1}{x_2 - x_1} \int_{x_1}^{x_2} \partial_x D dx = \frac{D(x_2) - D(x_1)}{x_2 - x_1} \longrightarrow 0.$$

Thus,

$$\bar{w} = \varepsilon\overline{\partial_x\hat{\Phi}} \tag{64}$$

in the limit as $N \longrightarrow \infty$.

The coefficient of x in the linear dependence (44) is

$$-2\varepsilon^{-1}\sum_{n=1}^{N} \mu_n k_n,$$

whence it follows that

$$\varepsilon\partial_x\hat{\Phi}(x) = -2\sum_{n=1}^{N}\hat{\mu}_n(x)k_n.$$

Using equation (61), we find that as $N \longrightarrow \infty$

$$\varepsilon\overline{\partial_x\hat{\Phi}} \approx -2\sum_{n=1}^{N}\bar{\bar{\mu}}_n k_n$$

$$\approx\left(\frac{2x}{L_s}-1\right)\sum_{n=1}^{N}k_n(y_0)_n+\frac{8t}{L_s}\sum_{n=1}^{N}k_n(y_2)_n.$$

Using this relation and equation (64), we obtain from the general expressions (62) and (63)

$$\begin{aligned}\bar{u} &= 4\varepsilon L_s^{-1}\sum_{n=1}^{N}k_n(y_0)_n,\\ \overline{u^2} &= \frac{16\varepsilon}{3}L_s^{-1}\sum_{n=1}^{N}k_n(y_2)_n.\end{aligned} \tag{65}$$

Here the fact that we confine our attention to the region of x and t where $0 \le y_i \le 1$ and therefore $(y_0)_n$ and $(y_2)_n$ do not depend on x and t is taken into account. Hence, passing to the continuous limit $N \longrightarrow \infty$ ($\varepsilon N/L_s = C_s$) we obtain

$$\bar{u} = 4C_s\int_0^1 k\rho_s(k)y_0(k; C_s)dk. \tag{66}$$

This average does not depend on x, so that $\partial_{xx}\bar{u} = 0$ and therefore equations (63) and (64) give

$$\overline{u^2} = -\frac{2}{3}\varepsilon\partial_t\bar{w} = \frac{16}{3}C_s\int_0^1 k^3\rho_s(k)y_0(k; C_s)dk. \tag{67}$$

The derivation of the last expression also employed the identity

$$\int_0^1 k\rho_s(k)y_2(k; C_s)dk = \int_0^1 k^3\rho_s(k)y_0(k; C_s)dk,$$

which is obtained by multiplying equation (59) with $\alpha = 0$ and $\alpha = 2$ by $\rho_s(k)$ and $k^2\rho_s(k)$, respectively, and then integrating the difference of the obtained relations over k from 0 to 1.

We now separate in the sum (65) for $\bar{u}$ a narrow region Δk in a neighborhood of any value of k_n chosen in the interval (0, 1). Passing to the continuous limit as $N \longrightarrow \infty$, just as in equation (66), we find

$$d\bar{u} = 4C_s k\rho_s(k)y_0(k; C_s)dk. \tag{68}$$

The possibility of separating the differential contribution $d\bar{u}$ is due to the fact that the amplitude of the solitons is preserved in an interaction so that there is no mixing of the spectrum in regions with different values of k. This makes it possible to use the differential conservation law (29). Comparing equations (29) and (68) we shall establish a relation between the spectral densities of the S-set and the set of solitons:

$$C_s y_0(k, C_s)\rho_s(k) = C\rho(k). \tag{69}$$

It is obvious that this relation agrees with the integral conservation laws (compare equations (23) and (66) and equations (24) and (67)).

Next, we shall take into account the normalization condition (45). This makes it possible to find immediately the solution of the first integral equation (59)

$$y_0(k, C) = 1 - \frac{C}{k}\int_0^1 l(k, k')\rho(k')dk' \tag{70}$$

and the density of the gas noninteracting particles in the statistical limit

$$C_s = C\int_0^1\frac{\rho(k)dk}{y_0(k, C)}. \tag{71}$$

Thus, the use of a maximizer, together with the initial requirement that the conservation laws be satisfied, completely determines the relation between the soliton spectrum and the S-set with an asymptotic transition into a stochastic state:

$$C \longrightarrow C_s(C), \quad \rho(k) \longrightarrow \rho_s(k) = \frac{C\rho(k)}{C_s(C)y_0(k, C)}. \tag{72}$$

Such a transformation is physically completely obvious. It determines the correspondence between a one-dimensional gas of free, uniformly moving, noninteracting points of the S-set and a gas of interacting solitons. Both sets are related by the relations (42)–(44) and are in a statistical state, i.e., the phases s_n are random. We recall that the density C is proportional to the dispersion parameter ε. For this reason, the fact that in our problem the parameter ε is a finite but small quantity plays the main role in the possibility of the transformation (72).

The spectrum $\rho_s(k)$ for various values of the density C for the example of parabolic pulses (3) and (10) which was considered above is presented in Fig. 2. It is evident that there is a large deformation of the spectrum: ρ_s decreases for large k and increases for small k. This drift in the direction of small k becomes increasingly stronger as C increases. The density C_s also increases substantially with C, and as C approaches the maximum value $C_m = 0.25$, in general, C_s approaches infinity.[3] At the same time, the scale L_s on which the

[3] We note that a singularity arises as $C \longrightarrow C_m$, and the small modulations of the pulse widths θ_i (see equation (13)) also become substantial. For this reason, this case must be analyzed separately.

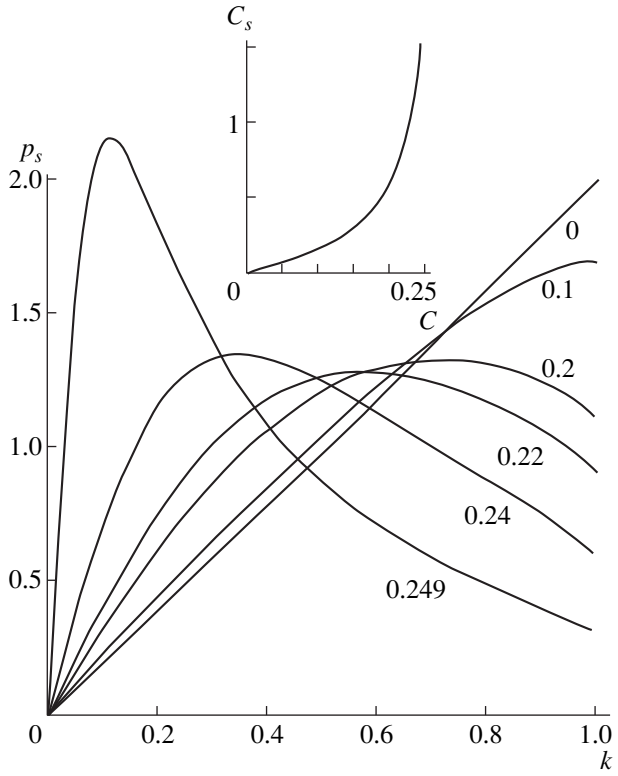

Fig. 2. Spectral density $\rho_s(k)$ for the indicated values of C and the dependence of C_s on C.

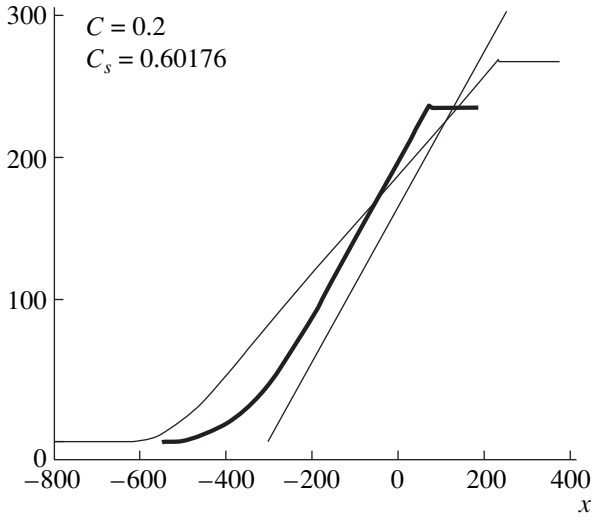

Fig. 3. Integral u from $-\infty$ to x for $C = 0.2$.

points of the S-set are given, according to equation (46), decreases with increasing C:

$$L_s = L\frac{C}{C_s}.$$

We note that such a stretching of the space

$$s \longrightarrow Kx$$

is possible only for an ergodic initial state defined on the entire x axis. Physically, the result obtained is completely understandable: The points of the S-set do not interact, while solitons repel one another and therefore the "volume" which they occupy increases. It is because of the repulsion that the soliton density C cannot exceed a finite value C_m even as $C_s \longrightarrow \infty$. From Fig. 2, just as from equations (69)–(71), it is also evident that as the density of the soliton gas decreases, the difference between ρ_s and ρ decreases, and in the limit $C \longrightarrow 0$ we have $\rho_s(k) = \rho(k)$. This degenerate case was studied in [1].

The necessity of transforming the spectrum (72) when passing from solitons to the S-set is illustrated in Fig. 3. The figure shows the integral of the function $u_N(x)$ from $-\infty$ to x. The calculation was performed using the maximizer for $N = 100$ solitons and the set of parameters k_n, s_n corresponding $\rho_s(k)$, C_s according to the transformation (72). To smooth the fluctuations, averaging was performed over an ensemble of 50 random realizations of the function $u_N(x)$. The thin straight line shows the slope of the integral, determined by equation (23) for $\langle u \rangle$. It is evident that the agreement with the numerical calculation is good. The thin line represents the same integral but with the parameters k_n, s_n corresponding to the initial $\rho(k)$ and C. Here, conversely, there is no agreement with the formula (23).

It should be especially underscored that the maximizer plays an important role, making it possible to develop a very efficient algorithm for performing numerical calculations using the formula for the N-soliton solution (42). For large values of N, finding the maximizer $\hat{\mu}$ of the argument Φ_μ of the exponential beforehand makes it possible to retain in the sum (42) a comparatively small number of terms corresponding to variations near the maximizer. The number of terms included depends on the accuracy required of the calculation and especially strongly on the parameter C_s. Actually, for small C_s ($C_s \sim 0.2$ and smaller), only several (2–5) terms are required to obtain an accuracy of 10^{-2}–10^{-3}. But, even for $C_s \sim 1$, when hundreds of terms are required in order to obtain the same accuracy, the number of terms is still incomparably less than 2^N. This makes the calculation radically faster than summing all 2^N terms. In addition, the approximate maximizer (49), which can be easily calculated even for large N, is helpful in the numerical calculation of $u_N(x)$, being a good starting point for finding the exact maximizer by the "steepest rise" method. Ordinarily, it is sufficient to change in 2–4 elements in $\tilde{\mu}$ (with a total of the order of 100 elements) so that further changes would only decrease Φ_μ. The indicated methods make it possible to perform very efficiently numerical calculations for all necessary characteristics in the statistical limit under consideration.

Thus, we have constructed an algorithm for finding the asymptotic solution of the KdV equation in the semiclassical limit (2) for initial conditions which result in stochastization of the phases. The basis of the algorithm lies in establishing a direct relation between the density and spectrum of the *S*-set, C_s and $\rho_s(k)$, and the set of interacting solitons C and $\rho(k)$. The latter are determined according to a prescribed initial potential $u_0(x)$, after which C_s and $\rho_s(k)$ are found using the relations (70)–(72) and then a direct calculation of the required characteristics is performed using the maximizer constructed here for the *N*-soliton solution in the statistical limit. In what follows, we shall present additional examples of the application of this method.

The characteristics of the statistical *S*-set, determining the asymptotic properties of the solution of the KdV equation (1) and (2) for a prescribed initial function $u_0(x)$, were found above. An inverse formulation of the problem is also possible: to find the initial distribution from prescribed C_s and $\rho_s(k)$. As follows from equation (59), $y_0(k, C_s)$ is then determined by the integral equation

$$y_0(k, C_s) = 1 - \frac{C_s}{k}\int_0^1 l(k, k')\rho_s(k')y_0(k', C_s)dk'.$$

Having solved this equation, we can find, according to equations (69) and (7), the density

$$C = C_s\int_0^1 \rho_s(k)y_0(k, C_s)dk$$

and the spectrum

$$\rho(k) = \frac{\rho_s(k)y_0(k, C_s)}{\int_0^1 \rho_s(k)y_0(k, C_s)dk},$$

as well as the average velocity (23), the mean-square velocity (24), and all other averages (21). It is obvious that an entire class of initial functions $u_0(x)$ leads to the same values of the density C and spectrum $\rho(k)$. But greater information about $u_0(x)$ cannot be obtained on the basis of prescribed C_s and $\rho_s(k)$. This is one of the manifestations of the irreversibility of the problem under consideration.

The density C and the spectrum $\rho(k)$, which are determined by the initial function $u_0(x)$, are preserved in the statistical limit found. According to the results of Section 3, this means that all conservation laws of the KdV equation are satisfied in our solution. The method developed, where an *N*-soliton solution with random phases is used, can therefore be regarded as a modification of the inverse problem of scattering for an oscillating nonperiodic semiclassical initial potential $-u_0(x)$ defined on the entire x axis.

7. QUASISOLITONS

Solitons are not free in the stochastic solution which we found for the KdV equation: They are constantly interacting with other solitons. This changes their velocity, which, just as the form of the soliton, fluctuates constantly. However, in a homogeneous stochastic state, the average velocity $\bar{v}_s(k)$, the average amplitude $a_s(k)$, and the effective average form $\bar{u}_s(x)$ of a soliton are constant and determined by the single parameter k. Therefore it is natural to introduce the concept of a "quasisoliton," i.e., a fluctuating soliton, corresponding to this parameter k. In the limit $C \longrightarrow 0$, a quasisoliton becomes an ordinary classical soliton. The deviations from an ordinary soliton for average quantities and for fluctuations grow with increasing density C.

A quasisoliton is, seemingly, an ordinary soliton moving in a fluctuating soliton vacuum. Its average velocity $\bar{v}(k)$, as follows from equation (60), is

$$\bar{v} = \frac{4y_2(k)}{y_0(k)}. \tag{73}$$

The first correction obtained hence with respect to the parameter C to the soliton velocity

$$\delta v(k) = \frac{4C}{k}\int l(k, k')(k^2 - k'^2)\rho(k')dk' \tag{74}$$

is identical to the correction found by Zakharov [8].

We shall investigate the changes in the form and amplitude of a quasisoliton for a numerical example. Figure 4 shows a "test" soliton against the background consisting of a stochastic soliton gas with density $C = 0.2$. The background solitons have, as usual, wave numbers in the interval $0 \le k \le 1$ (i.e., the amplitude $0 \le a \le 2$) and their spectrum is determined according to equation (11). In the calculation, the numerical method described above was used, taking account of the parameters of the *S*-set, which were determined according to equations (70)–(72). The amplitude of the test quasisoliton was assumed to be large, $a_0 = 2k_0^2 = 3$, so that it can be singled out among the background solitons. It is evident that the test soliton joins the background curve $u_N(x)$ almost exactly, if it falls into the region with a low perturbation level in $2k_0^2$ (and its height remains almost unchanged, close to $2k_0^2$). For a high level of $u_N(x)$, however, the interaction has a large effect and leads to "repulsion" of close maxima of $u_N(x)$, and the height of the test soliton decreasing appreciably.

Figure 5 compares for the same value $C = 0.2$ the form of the test quasisoliton with amplitude $2k_0^2 = 2.5$ with a "free" soliton with the same value of k_0, i.e., with the function $S_0(x) = 2k_0^2 \cosh^{-2}(k_0 x)$. The wave numbers of the background solitons $k_1, \ldots, k_N$ were distrib-

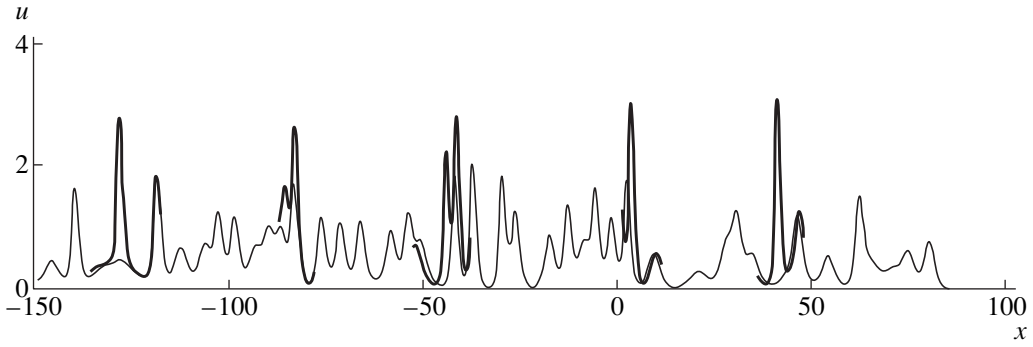

Fig. 4. Trial soliton with amplitude $2k_0^2 = 3$ (thick line) ($0 < k_n \le 1$) for $C = 0.2$.

uted in the interval (0, 1) with density $\rho(k) = 2k$. The form of the quasisoliton was calculated by averaging over 150 shifts of the parameter s_0: the functions $u_{N+1}(x; k_0, \ldots, k_N, s_0, \ldots, s_N)$ for various values of s_0 were shifted so that their highest maximum would shift into the point $x = 0$, and their arithmetic mean was calculated. The result of the averaging was used to calculate the average value of $u_N(x)$, which, according to equation (23), is $(8/3)C$ for $\rho(k) = 2k$. It is evident that a quasisoliton differs substantially from a free soliton by the maximum height, a general downward shift, and a more complicated form (minima appear). As k_0 increases, the difference between the quasisoliton and a free soliton decreases, just as expected. Conversely, as k_0 decreases, a larger change in the form of a quasisoliton can be expected.

Figure 6 shows a spatial-temporal picture of the motion of a test soliton among N solitons with randomly selected initial values of s_n. As an example, $N = 30$ values were chosen, $C_s = 0.4$, and k_0 was chosen to be above the maximum background solitons in the ratio 10/9. The figure shows the coordinates of the maxima of the function $u_{N+1}(x, t)$ in short time intervals ($\Delta t = 0.1$). The points of the maximum which exceed 90% of the free amplitude of the test soliton $2k_0^2$ are singled out. It is noteworthy that even the points of the maximum on the "track" of the test soliton which were not singled out fall practically on the same straight line. This indicates that the maximum moves with the same (and even somewhat higher) velocity, even though its height is lower. It is clearly seen that the average velocity of a quasisoliton is higher than the velocity of a free soliton $4k_0^2$ (the dashed curve in the figure corresponds to uniform motion with this velocity). The reason for this acceleration is simple: a high-amplitude soliton undergoes rapid additional shifts at the moments when collisions occur. The value of the average velocity of the test soliton agrees well with equation (73). Figure 6 shows the tracks and some background solitons whose amplitude is less than that of the test soliton and which move, naturally, more slowly. We note that according to equation (73), the solitons with a small amplitude are not accelerated, but rather they move more slowly than the free soliton with the same amplitude: their average velocity decreases because of the interaction with the background.

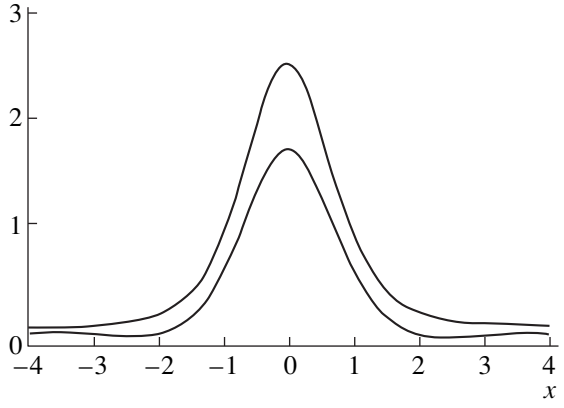

Fig. 5. Comparison of a quasisoliton averaged over the shifts s_0 (with the average $\bar{u}$ subtracted out) and a free soliton with the same wave number k_0.

8. PROBABILITY DENSITY

A general method for determining the probability density in the asymptotic limit $t \longrightarrow \infty$ was developed in [1]:

$$f(u) = \langle \delta(u - u(x, t)) \rangle + O\left(\frac{1}{t}\right), \qquad (75)$$

where the average $\langle\rangle$ is taken over the characteristic scale of the uniformity of the problem. However, it was found that $f(u)$ could be calculated only in the limiting case of free solitons $C \longrightarrow 0$, since the relation between the phase-mixed set of solitons and the S-set was not

studied in [1]. In the present paper this relation is established and it is shown that it is unique. This makes it possible to indicate a direct algorithm for calculating the probability density.

The N-soliton formula (42) is used to make a specific calculation of the uniform probability density $f(u)$ distributed uniformly over the phases. First, according to equations (46) and (70)–(72) obtained above, the scale, density, and spectrum of the initial distribution are transformed as $L \longrightarrow L_s$, $C \longrightarrow C_s$, and $\rho(k) \longrightarrow \rho_s(k)$, respectively, for the transition to the S-set. Then, according to equation (75), we have

$$f(u) = \frac{1}{L}\int_{-L/2}^{L/2} \delta(u - u_N(x, t))dx = \left\langle \frac{1}{L}\sum_i \left|\frac{du}{dx}\right|_{x_i(u)} \right\rangle_{s_1 \ldots s_N}. \quad (76)$$

The averaging in the last expression was first performed over all roots of the inverse function $x_i = x_i(u)$. The brackets $\langle\,\rangle_{s_1 \ldots s_N}$ denote a further averaging over an entire ensemble of uniformly distributed phases $s_1 \ldots s_N$. We note that according to equation (76), the function $f(u)$ is "bad" from the standpoint of continuity, since $u(x)$ undergoes many oscillations ($\sim N$), and each extremum gives in equation (76) a singularity of the type $(u - u_{\min})^{-1/2}$ or $(u - u_{\max})^{-1/2}$. Averaging over many realizations does not remove these singularities: their number only increases, though the coefficient for each one correspondingly decreases. For this reason, an additional smoothing procedure is used to calculate $f(u)$. Of course, the computational results do not depend on this smoothing procedure.

The calculations were performed using quite long N-soliton realizations. Since the fluctuations are substantial even for $N \sim 100$, to obtain objectively significant results it is necessary to perform an additional averaging over an ensemble consisting of a quite large number of realizations (which, because of ergodicity, is equivalent to increasing the length of a realization, i.e., the number N).

Figure 7 demonstrates the presence of a quite large random spread in the functions $f(u)$, calculated over 20 specific realizations of the function $u_N(x)$ with $N = 100$. The next figure shows the averages over ensembles of a much larger number of realizations of the function $f(u)$ for two values of the parameter C: for $C = 0.1$ the averaging was performed over 500 realizations, while for $C = 0.2$ the averaging was performed over 400 realizations. It is evident that as C increases, the structure of $f(u)$ deforms: a maximum of the probability density stands out clearly at a finite value of u. The maximum shifts in the direction of large values of u as $C \longrightarrow C_m$. Conversely, for low densities C the function $f(u)$ increases for small values of u according to a power law:

$$f(u) \approx Au^{-\alpha}.$$

The dependence of the exponent α on the soliton density C is also shown in Fig. 8. The straight line $\alpha = 1 - 2C$ shows the limit $C \longrightarrow 0$, which was established in [1].

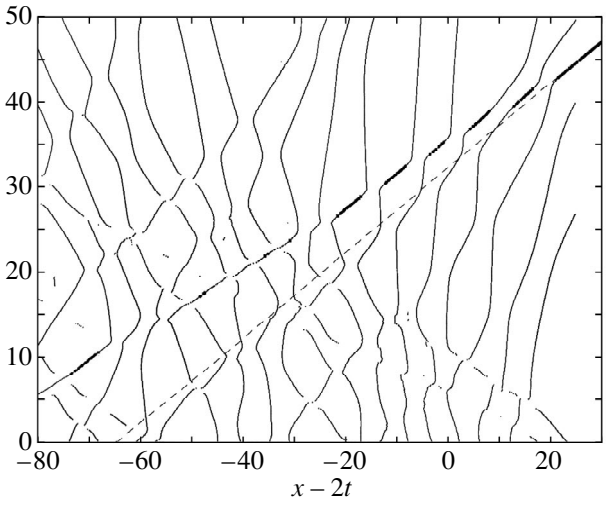

Fig. 6. Motion of a test soliton with amplitude $2k_0^2$ among background solitons.

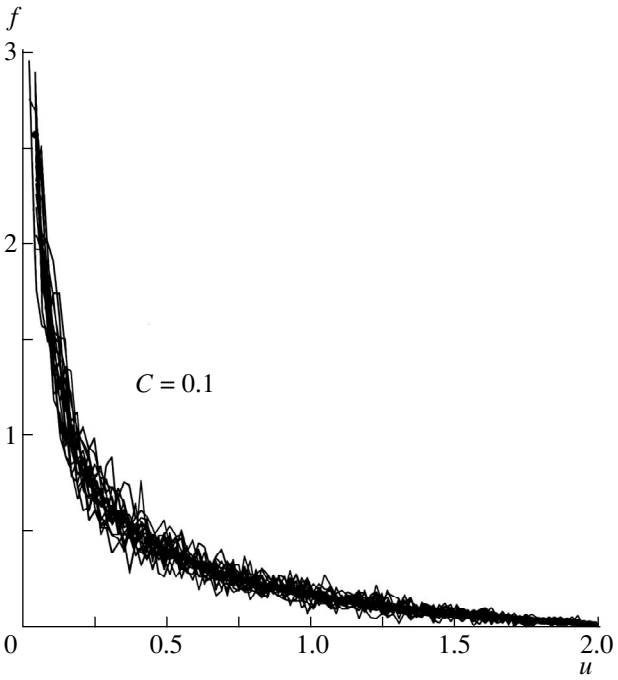

Fig. 7. Probability density $f(u)$ with $C = 0.1$. Fluctuations for various random realizations of the function $u_N(x)$ are shown.

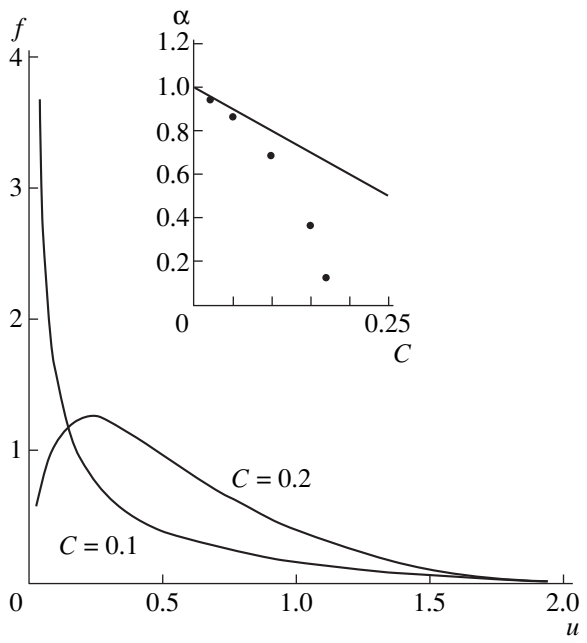

Fig. 8. Probability density $f(u)$, averaged over ensembles of realizations of the function $u_N(x)$, for two values of C. The dependence of the exponent on C in the asymptotic limit $f(u) \approx Au^{-\alpha}$ as $u \longrightarrow 0$; straight line æ theoretical value as $C \longrightarrow 0$ [1].

We note that for very widely spaced pulses, when $C \longrightarrow 0$, the expression (42)–(44) for $u_N(x, t)$ becomes a sum of single-soliton solutions $u_s(x)$. As a result, the expression for the equilibrium distribution function $f(u)$ reduces to the expression which we obtained previously in [1]:

$$f(u) = \langle \delta(u - u_s(x)) \rangle = \frac{2}{L}\left\langle \frac{1}{|du_s/dx|} \right\rangle$$

$$= \frac{2}{L}\int_u^2 \rho(a)\left|\frac{du_s}{dx}\right|^{-1} da.$$

Here $\rho(a)$ is the amplitude distribution of the solitons.

Thus, the general algorithm constructed here made it possible to find, for a specific example, the probability density $f(u)$ for the initial function $u_0(x)$, defined on the entire x axis and generating a system of interacting solitons with arbitrary density. The correlation function $K(s)$ can be found in a completely similar manner.

In conclusion, we note that the theory constructed in [1] and the present paper describes the appearance and development of a turbulent state in one-dimensional dispersion hydrodynamics. The successive elimination of the small dispersion parameter ε in the semiclassical (weak) limit considered makes it possible to describe a flow using a system of quasilinear nondissipative equations of hydrodynamic (Euler) type: Hopf's equation and Witham's equation for single-mode and multimode systems. The singular points arising in a hydrodynamic flow are decisive for the appearance and structure of fine-scale oscillations. New modes of fine-scale oscillations always develop near singular points. The increase of the number of modes in time ultimately results in complete stochastization of the small-scale oscillations, i.e., the development of turbulence.

The statistical properties of developed small-scale turbulence are completely determined by the integral characteristics of the initial large-scale hydrodynamic flow. A method making it possible to determine the characteristics of turbulence according to a prescribed initial state was developed. The finite (though arbitrarily small) value of the dispersion parameter ε plays a fundamental role in developed turbulence. Of course, the theory constructed makes substantial use of the integrability properties of the KdV equations.

ACKNOWLEDGMENTS

We thank S. P. Novikov and I. M. Krichever for detailed discussions of this work and for valuable remarks. This work was supported by the Russian Foundation for Basic Research (project no. 96-15-96556) and INTAS.

REFERENCES

1. A. V. Gurevich, K. P. Zybin, and G. A. Él', Zh. Éksp. Teor. Fiz. **115**, 333 (1999) [JETP **88**, 182 (1999)].
2. S. Novikov, S. V. Manakov, L. P. Pitaevskii, and V. E. Zakharov, *Theory of Solitons: the Inverse Scattering Method* (Nauka, Moscow, 1980; Consultants Bureau, New York, 1984).
3. P. D. Lax and C. D. Levermore, Commun. Pure Appl. Math. **36**, 253 (1983); **36**, 571 (1983); **36**, 809 (1983).
4. A. V. Gurevich and L. P. Pitaevskiĭ, Zh. Éksp. Teor. Fiz. **65**, 590 (1973) [Sov. Phys. JETP **38**, 291 (1973)].
5. B. A. Dubrovin, V. B. Matveev, and S. P. Novikov, Usp. Mat. Nauk **31**, 59 (1976).
6. J. N. Elgin, Phys. Lett. A **110**, 441 (1985).
7. F. Kh. Abdulaev, Phys. Rep. **179**, 1 (1989).
8. V. E. Zakharov, Zh. Éksp. Teor. Fiz. **60**, 993 (1971) [Sov. Phys. JETP **33**, 538 (1971)].
9. Y. Matsuno, Phys. Lett. A **64**, 14 (1972); Physica A **104**, 71 (1980).
10. A. V. Gurevich, A. L. Krylov, and N. G. Mazur, Zh. Éksp. Teor. Fiz. **95**, 1674 (1989) [Sov. Phys. JETP **68**, 966 (1989)].
11. A. V. Gurevich, A. L. Krylov, and G. A. Él', Zh. Éksp. Teor. Fiz. **101**, 1797 (1992) [Sov. Phys. JETP **74**, 957 (1992)].
12. F. R. Tian, Commun. Pure Appl. Math. **46**, 1093 (1993).
13. B. A. Dubrovin, Usp. Mat. Nauk **36**, 215 (1981).
14. T. Grave, PhD Thesis ISAS (Trieste, 1998).
15. H. Flaschka, G. Forest, and D. W. McLaughlin, Commun. Pure Appl. Math. **33**, 739 (1979).
16. I. M. Krichever, Funk. An. Pril. **22**, 37 (1988).

17. B. A. Dubrovin and S. P. Novikov, Usp. Mat. Nauk **44**, 35 (1989).
18. S. Venakides, Am. Math. Soc. Trans. **31**, 189 (1987).
19. E. Ya. Khruslov and V. P. Kotlyarov, Usp. Mat. Nauk **19**, 129 (1994).
20. Ya. G. Sinaĭ and E. I. Dinaburg, private communication (1998).
21. G. B. Witham, Proc. R. Soc. London, Ser. A **283**, 238 (1965).
22. I. P. Kornfel'd, Ya. G. Sinaĭ, and S. V. Fomin, *Ergodic Theory* (Nauka, Moscow, 1980).
23. L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-Relativistic Theory* (Nauka, Moscow, 1986, 4th ed.; Pergamon Press, Oxford, 1977, 3rd ed.).
24. G. M. Zaslavsky, *Chaos in Dynamical Systems* (Nauka, Moscow, 1984; Harwood, Chur, 1985).
25. L. D. Landau and E. M. Lifshitz, *Statistical Physics* (Nauka, Moscow, 1976, 3rd ed.; Pergamon Press, Oxford, 1980, 3rd ed.), Vol. 1.
26. I. Kay and H. E. Moses, J. Appl. Phys. **27**, 1503 (1956).
27. P. D. Lax, C. D. Levermore, and S. Venakides, in *Important Developements in Soliton Theory*, Ed. by A. S. Fokas and V. E. Zakharov (Springer, New York, 1993), p. 205.
28. *Solitons*, Ed. by R. K. Bullough and P. J. Caudrey (Springer-Verlag, Berlin, 1980; Mir, Moscow, 1983).
29. N. G. Mazur, Teor. Mat. Fiz. **106**, 44 (1996).

Translation was provided by AIP

Journal of Experimental and Theoretical Physics, Vol. 90, No. 4, 2000, pp. 714–723.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 117, No. 4, 2000, pp. 818–828.

MISCELLANEOUS

Relativistic Quantum Bit Commitment in Real-Time

S. N. Molotkov* and S. S. Nazin

Institute of Solid-State Physics, Russian Academy of Sciences, Chernogolovka, Moscow oblast, 142432 Russia
**e-mail: molotkov@issp.ac.ru*

Received November 2, 1999

Abstract—A relativistic quantum exchange protocol making it possible to implement a bit commitment scheme is realized. The protocol is based on the idea that in the relativistic case the propagation of a field into a region of space accessible for measurement requires, in contrast to the nonrelativistic case, a finite time that depends on the structure of the states. The protocol requires one classical and several quantum communication channels. It turns out that it is possible in principle to preserve the secret bit for as long a period of time desired and with probability as close to 1 as desired. *© 2000 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The conventional bit commitment scheme includes two spatially separated participants A and B and reduces to the following [1]. Let us assume that at a certain time t_0 the participant A selects one bit (0 or 1) and transmits to B part of the information about this bit, and this information is insufficient to determine reliably which bit A chose. Further, after receiving the information sent by participant A participant B can ask A, at any time at the second stage of exchange (commitment stage), for the remaining part of the information about the bit, and B must be confident that A has not changed his decision about the choice of bit, received at the time t_0, i.e., A should not be able to fool B.

The classical bit commitment schemes, where A and B can exchange information only through a classical communication channel, are based on the unproved complexity of computing, for example, a discrete logarithm. If A and B can also use a quantum communication channel in addition to a classical communication channel, then such a scheme is called a quantum bit commitment scheme. In this case the information carriers are quantum states. Various quantum bit commitment protocols have been proposed [2–4]. Subsequently it was determined that the ideal quantum bit commitment scheme is impossible, since A can always fool B, using an Einstein–Podolsky–Rosen (EPR) pair and delaying his measurements (actually delaying the choice of bit himself) to the second stage of the protocol [5, 6].

A classical bit commitment protocol that takes account of the existence of a maximum propagation velocity of the signal (information) was recently proposed [7, 8]. A relativistic classical protocol [7, 8] is definitely secret (i.e., the secrecy is based only on the fundamental laws of nature) and makes it possible, in principle, to delay the second stage of the protocol, i.e., to preserve information about the secret bit, conceived by A, as long as desired, but to implement the protocol each participant A and B can monitor two spatially separated units.

We shall propose here a relativistic quantum protocol that makes it possible to implement bit commitment in a finite time interval. More accurately, we are talking about a protocol in which participant B could not reconstruct in a finite time t_c, on the basis of the information sent to him, the bit chosen by the participant A (this means that the probability of correctly identifying the chosen bit during as long as desired, but finite, prescribed time t_c does not exceed 1/2, i.e., the probability of randomly guessing this bit by as small as desired prescribed value ε). At any time $t < t_c$ the participant B can ask A for the bit transmitted to him, and at a certain time $T > t_c$ he can check which bit participant A chose at time t_0. We note that in reality the finite time t_c is implicitly also present in classical bit commitment protocols based on the computational complexity of particular problems, where it is determined by the computational resources of the participant B and must be chosen so that these resources would be inadequate to find the selected bit on the basis of information obtained in time t_c. In our protocol, however, the impossibility of determining in a time less than t_c the bit chosen by A is based not on the limitation of the technical capacity of B but rather by the fundamental laws of nature, just as in the classical relativistic protocol [7, 8]. However, the use of quantum states (more accurately, single-particle states of the field) as carriers of information makes it possible to construct a protocol in which each participant controls only his own unit. In addition, the existence of a limiting propagation velocity for states of the field makes it possible to construct, in contrast to nonrelativistic quantum mechanics, a secret protocol based on orthogonal states.

The quantum-mechanical measurements used in the protocol are discussed in Section 2. A graphic model example of measurements in the one-dimensional case

is presented in section 3. The protocol is described in Section 4.

2. QUANTUM-MECHANICAL MEASUREMENTS USED IN THE PROTOCOL

All quantum cryptographic protocols use the following two circumstances in one form or another. First, there is the so-called theorem [9] stating that it is impossible to copy an unknown quantum state, i.e., the assertion that the process

$$|A\rangle|\psi\rangle \longrightarrow U(|A\rangle|\psi\rangle) = |B_\psi\rangle|\psi\rangle|\psi\rangle,$$

where $|A\rangle$ and $|B_\psi\rangle$ are states of the apparatus before and after copying and U is a unitary operator, is impossible. Such a process is forbidden on the basis of the linearity and unitarity of evolution in quantum mechanics. Second, the even weaker process of obtaining any information about one of two nonorthogonal states without perturbing them is impossible [10], i.e., a process of the form

$$\begin{aligned}|A\rangle|\psi_1\rangle &\longrightarrow U(|A\rangle|\psi_1\rangle) = |A_{\psi_1}\rangle|\psi_1\rangle,\\ |A\rangle|\psi_2\rangle &\longrightarrow U(|A\rangle|\psi_2\rangle) = |A_{\psi_2}\rangle|\psi_2\rangle\end{aligned}$$

with $|A_{\psi_1}\rangle \neq |A_{\psi_2}\rangle$, if $\langle\psi_1|\psi_2\rangle \neq 0$, is impossible, which means that it is impossible to distinguish nonorthogonal states reliably. There is no such prohibition for orthogonal states. For this reason, almost all cryptographic protocols employ nonorthogonal states as information carriers. An exception is a protocol proposed in [11].

In nonrelativistic quantum mechanics there are no restrictions on reliably distinguishing one of a pair of orthogonal states without perturbing the states, and there are no restrictions on doing this instantaneously. At the present time, there is every reason to believe that in relativistic quantum theory orthogonal states also can be distinguished reliably without perturbing them. However, the existence of a maximum propagation velocity of the field (and information) introduces an additional restriction. States cannot be reliably distinguished instantaneously: orthogonal states can be distinguished reliably only in a finite time, which depends on the structure of the states themselves.

This is due to the fact that such measurements are nonlocal. We shall elucidate what we have in mind.

Let us consider the nonrelativistic case first. Let two orthogonal states $|\psi_1\rangle$ and $|\psi_2\rangle$ of a one-dimensional nonrelativistic particle be given,

$$\langle\psi_1|\psi_2\rangle = 0. \tag{1}$$

In the momentum representation of the Hilbert space we have

$$|\psi_{1,2}\rangle = \int_{-\infty}^{\infty}\psi_{1,2}(p)|p\rangle dp. \tag{2}$$

Let the wave functions $|\psi_{1,2}\rangle$ possess nonoverlapping supports

$$\begin{aligned}&\text{supp}\ \psi_1(p)\bigcap\text{supp}\ \psi_2(p) = \emptyset,\\ &\langle\psi_1|\psi_2\rangle = \int_{-\infty}^{\infty}\psi_1^*(p)\psi_2(p)dp = 0.\end{aligned} \tag{3}$$

The orthogonality of the states is, generally speaking, a nonlocal property in the sense that the values of the wave functions at all points in space appear in the scalar product. If the states are orthogonal in the complete space, then their projections on a subspace are not necessarily orthogonal. To demonstrate this we shall examine the same states in the coordinate representation of the Hilbert space:

$$\begin{aligned}|\psi_{1,2}\rangle &= \int_{-\infty}^{\infty}\psi_{1,2}(x)|x\rangle dx,\\ \psi_{1,2}(x) &= \int_{-\infty}^{\infty}e^{ipx}\psi_{1,2}(p)dp.\end{aligned} \tag{4}$$

The orthogonality of the states is preserved, generally speaking, only in the complete space, and with a restriction on some subspace the states are not necessarily orthogonal, i.e.,

$$\begin{aligned}&\langle\psi_1|\psi_2\rangle = \int_{\Omega^{(x)}}\psi_1^*(x)\psi_2(x)dx \neq 0,\\ &\text{where}\quad \Omega^{(x)} \neq (-\infty, \infty).\end{aligned} \tag{5}$$

Generally speaking, in order to be able to distinguish states reliably, access is required to the entire region of the coordinates space where these states are present. In nonrelativistic quantum mechanics there are no restrictions on the instantaneous access to the entire coordinate space, so that there are no restrictions on performing instantaneous nonlocal measurements. Let us clarify what we have in mind. According to the nonrelativistic quantum-mechanical theory of measurements, the statistics of the outcomes of any measurement refers to a state of the system at a certain definite moment in time. In the nonrelativistic case information about the outcome of a certain measurement which touches upon spatially separated points (for example, the measurement apparatus can "extract" information about a quantum state at the same moment in time at different spatial points) can be collected into a single point for the observer instantaneously (at the same moment in time), since there are

no restrictions on the propagation velocity of information.

In the relativistic case, in contrast to the nonrelativistic quantum mechanics, there is no systematic theory of measurements. Apparently, nonlocal instantaneous measurements in the sense mentioned above for the nonrelativistic case are impossible. Even if the measurement apparatus performs measurements on a state at the same moment in time (in some reference system) at different spatial points $\mathbf{x}_1$ and $\mathbf{x}_2$, a finite time $t = |\mathbf{x}_1 - \mathbf{x}_2|/2c$ is required to collect the data obtained at one point for the observer. For this reason, nonlocal measurements require a finite time that depends on the structure of the state.

Thus, in relativistic quantum field theory a finite time is required to obtain information about the outcome of a nonlocal measurement. In addition, when the measurement apparatus has access to a bounded region D of space, a measurement of a nonstationary state will give information about this state only if the spatial support of the state intersects D at the moment the measurement is performed. If at time t_0 the support of the state is located outside D, then complete information about the state can be obtained by a measurement localized in D no sooner than in a time $t_1 \approx L/2c$, where L is the characteristic size of the support of the state, since the support of the state cannot be wholly in D for a time less than t_1 because the field propagation velocity is finite.

Before describing the protocol, we shall discuss in greater detail the measurements used. In what follows, we shall be considering massless particles (photons). The operators of the four-dimensional vector potential have the form [12]

$$\begin{aligned} A_n^{\pm}(\hat{x}) &= \frac{1}{(2\pi)^{3/2}}\int e^{\pm i\hat{k}\hat{x}} A_n^{\pm}(\mathbf{k})\frac{d\mathbf{k}}{\sqrt{2k^0}} \\ &= \frac{1}{(2\pi)^{3/2}}\int e^{\pm i\hat{k}\hat{x}} e_n^m(\mathbf{k}) a_m^{\pm}(\mathbf{k})\frac{d\mathbf{k}}{\sqrt{2k^0}}, \end{aligned} \tag{6}$$

where $\hat{k}\hat{x} = k^0x^0 - \mathbf{kx}$. The operators $A_n^{\pm}(\hat{x})$ of the four-dimensional vector potential are required to satisfy the Bose commutation relation [12]

$$[A_m^-(\hat{x}_1), A_n^+(\hat{x}_2)]_- = ig_{mn}D_0^-(\hat{x}_1 - \hat{x}_2), \tag{7}$$

where g^{nm} is the metric tensor ($g^{00} = -g^{11} = -g^{22} = -g^{33} = 1$), D_0^- is the negative-frequency commutator function for a field with zero mass

$$\begin{aligned} D_0^-(\hat{x}) &= -i\frac{1}{(2\pi)^{3/2}}\int d\hat{k}\,\delta(\hat{k}^2)\theta(-k^0)e^{i\hat{k}\hat{x}} \\ &= \frac{i}{(2\pi)^3}\int \frac{d\mathbf{k}}{2|\mathbf{k}|}\exp(-ix^0|\mathbf{k}| + i\mathbf{xk}), \end{aligned} \tag{8}$$

which is different from zero and possesses a singularity on the light cone [12]

$$D_0^-(\hat{x}) = \frac{1}{4\pi}\varepsilon(x^0)\delta(\lambda), \quad \lambda^2 = (x^0)^2 - \mathbf{x}^2. \tag{9}$$

In what follows, we assume $c = 1$. The creation and annihilation operators $a_m^{\pm}(\mathbf{k})$ describe four types of photons: two transverse, "time-like," and longitudinal. The latter are fictitious and are introduced in order to preserve the four-dimensional structure of the vector potential. The commutation relations have the form

$$[a_m^-(\mathbf{k}), a_n^+(\mathbf{k}')]_- = -g^{nm}\delta(\mathbf{k} - \mathbf{k}'),$$

$$(\mathbf{e}^{\alpha}\cdot\mathbf{e}^{\beta}) = \delta_{\alpha,\beta} \quad (\alpha, \beta = 1, 2, 3),$$

$$e_0^{\alpha} = 0, \quad \mathbf{e}^3 = \mathbf{k}/|\mathbf{k}|.$$

When working with four types of photons, it is necessary to use an indefinite metric. For our purposes it will be sufficient to confine our attention to one type of photon with a definite helicity, so that we shall work immediately in the subspace of single-particle states with the standard Hermitian scalar product (see details in [12]).

The single-photon states $|\psi_1\rangle$ and $|\psi_2\rangle$, referring to 0 and 1, respectively, can be chosen in the form

$$\begin{aligned} |\psi_{1,2}\rangle &= \int \psi_{1,2}(\mathbf{k})a^+(\mathbf{k}, s)|0\rangle\frac{d\mathbf{k}}{\sqrt{2k^0}} \\ &= \int \psi_{1,2}(\mathbf{k})|\mathbf{k}, s\rangle\frac{d\mathbf{k}}{\sqrt{2k^0}}, \\ \langle\mathbf{k}, s|\mathbf{k}', s\rangle &= \delta(\mathbf{k} - \mathbf{k}'), \end{aligned} \tag{10}$$

where $a^+(\mathbf{k}, s)$ is the creation operator for a photon with momentum $\mathbf{k}$ and helicity s. We assume that the amplitudes of the states have nonoverlapping supports

$$\begin{aligned} \text{supp}\ \psi_1(\mathbf{k}) \bigcap \text{supp}\ \psi_2(\mathbf{k}) &= \emptyset, \\ \Omega_i^{(k)} &= \text{supp}\ \psi_i(\mathbf{k}). \end{aligned} \tag{11}$$

Then the states $|\psi_1\rangle$ and $|\psi_2\rangle$ are orthogonal:

$$\langle\psi_1|\psi_2\rangle = \int \psi_1^*(\mathbf{k})\psi_2(\mathbf{k})\frac{d\mathbf{k}}{2k^0} = 0. \tag{12}$$

A measurement making it possible to distinguish reliably a pair of orthogonal states $|\psi_1\rangle$ and $|\psi_2\rangle$ is given by a decomposition of unity:

$$\mathcal{P}_{\psi_1} + \mathcal{P}_{\psi_2} + \mathcal{P}_{\perp} = I, \quad I = \int |\mathbf{k}, s\rangle\langle\mathbf{k}, s|\frac{d\mathbf{k}}{2k^0}, \tag{13}$$

$$\mathcal{P}_{\psi_{1,2}} = \left(\int\limits_{\Omega_{1,2}^{(k)}} \psi_{1,2}(\mathbf{k})|\mathbf{k}, s\rangle \frac{d\mathbf{k}}{\sqrt{2k^0}}\right) \times \left(\int\limits_{\Omega_{1,2}(k)} \langle \mathbf{k}', s|\psi_{1,2}^*(\mathbf{k}')\frac{d\mathbf{k}'}{\sqrt{2k^{0'}}}\right),$$

$$\mathcal{P}_\perp = I - \mathcal{P}_{\psi_1} - \mathcal{P}_{\psi_2}.$$

The probability of various outcomes on an input state $|\psi_1\rangle$ is given by the relations

$$\begin{aligned} \Pr_1\{\mathcal{P}_{\psi_1}\} &= \mathrm{Tr}\{|\psi_1\rangle\langle\psi_1|\mathcal{P}_{\psi_1}\} \\ &= \left(\int\limits_{\Omega_1^{(k)}} |\psi_1(\mathbf{k})|^2 \frac{d\mathbf{k}}{2k^0}\right)\left(\int\limits_{\Omega_1^{(k)}} |\psi_1(\mathbf{k}')|^2 \frac{d\mathbf{k}'}{2k^{0'}}\right) \equiv 1, \\ \Pr_{\psi_1}\{\mathcal{P}_{\psi_2,\perp}\} &= \mathrm{Tr}\{|\psi_1\rangle\langle\psi_1|\mathcal{P}_{\psi_2,\perp}\} \equiv 0 \end{aligned} \tag{14}$$

and analogously for the input state $|\psi_2\rangle$.

To demonstrate explicitly the nonlocal nature of the measurement (13) we shall examine an auxiliary measurement in the coordinate representation, which on expanding the spatial region up to the entire space becomes a measurement making it possible to distinguish a state with nonoverlapping supports.

We shall consider an auxiliary measurement making it possible to distinguish reliably not a specific pair of states but rather any states with nonoverlapping supports. Such a measurement is very similar to the measurement (13):

$$\begin{aligned} &\mathcal{P}_1 + \mathcal{P}_2 + \mathcal{P}_\perp = I, \quad I = \int |\mathbf{k}, s\rangle\langle \mathbf{k}, s| \frac{d\mathbf{k}}{2k^0}, \\ &\mathcal{P}_{1,2} = \int\limits_{\Omega_{1,2}^{(k)}} |\mathbf{k}, s\rangle\langle \mathbf{k}, s| \frac{d\mathbf{k}}{2k^0}, \quad \mathcal{P}_\perp = I - \mathcal{P}_1 - \mathcal{P}_2. \end{aligned} \tag{15}$$

The probability of various outcomes with an input state $|\psi_1\rangle$ is given by

$$\begin{aligned} \Pr_1\{\mathcal{P}_1\} = \mathrm{Tr}\{|\psi_1\rangle\langle\psi_1|\mathcal{P}_1\} &= \int\limits_{\Omega_1^{(k)}} |\psi_1(\mathbf{k})|^2 \frac{d\mathbf{k}}{2k^0} \equiv 1, \\ \Pr_1\{\mathcal{P}_{2,\perp}\} &= \mathrm{Tr}\{|\psi_1\rangle\langle\psi_1|\mathcal{P}_{2,\perp}\} \equiv 0 \end{aligned} \tag{16}$$

and similarly for the input state $|\psi_2\rangle$.

According to the quantum-mechanical theory of measurements, any quantum-mechanical measurement gives a decomposition of unity on a Hilbert space of states of the system, more accurately, by prescribing a positive operator-valued measure M on a certain measured set of results $\mathcal{U}$ that satisfies the conditions [13]

$$\begin{aligned} &1)\ M(\emptyset) = 0, \quad M(\mathcal{U}) = I, \quad M(\mathcal{U}_i) \geq 0, \\ &2)\ M(\mathcal{U}_1) \leq M(\mathcal{U}_2), \quad \mathcal{U}_1 \subseteq \mathcal{U}_2, \\ &3)\ M\left(\bigcup_i \mathcal{U}_i\right) = \Sigma_i M(\mathcal{U}_i), \quad \mathcal{U}_i \bigcap \mathcal{U}_j = \emptyset. \end{aligned} \tag{17}$$

We shall take for the set of measurement results a discrete set consisting of three points denoted by the indices $\{1, 2, \perp\}$:

$$\begin{aligned} M_{1,2} = &\int\limits_{\Omega^{(x)}} d\mathbf{x}\left(\int\limits_{\Omega_{1,2}^{(k)}} \exp(i\hat{k}(\hat{x} - \hat{x}_0))|\mathbf{k}, s\rangle \frac{d\mathbf{k}}{\sqrt{2k^0}}\right) \\ &\times \left(\int\limits_{\Omega_{1,2}^{(k)}} \exp(-i\hat{k}'(\hat{x} - \hat{x}_0))\langle \mathbf{k}', s|\frac{d\mathbf{k}'}{\sqrt{2k^{0'}}}\right), \end{aligned} \tag{18}$$

$$\begin{aligned} M_\perp = &\iiint\limits_{\mathcal{X}\mathcal{X}\mathcal{X}} d\mathbf{x}\left(\exp(i\hat{k}(\hat{x} - \hat{x}_0))|\mathbf{k}, s\rangle \frac{d\mathbf{k}}{\sqrt{2k^0}}\right) \\ &\times \left(\exp(-i\hat{k}'(\hat{x} - \hat{x}_0))\langle \mathbf{k}', s|\frac{d\mathbf{k}'}{\sqrt{2k^{0'}}}\right) - M_1 - M_2, \end{aligned} \tag{19}$$

where

$$\begin{aligned} \mathcal{X} &= \{\mathbf{x}\colon \mathbf{x} \in (-\infty, \infty)\}, \\ \mathcal{H} &= \{\mathbf{k}\colon \mathbf{k} \in (-\infty, \infty)\}. \end{aligned}$$

The positive operator-valued metric M_i gives a decomposition of unity on the space of results $\{1, 2, \perp\}$:

$$\begin{aligned} M_1 + M_2 + M_\perp + &\iiint\limits_{\mathcal{X}\mathcal{X}\mathcal{X}} d\mathbf{x}\left(\exp(i\hat{k}(\hat{x} - \hat{x}_0))|\mathbf{k}, s\rangle \frac{d\mathbf{k}}{\sqrt{2k^0}}\right) \\ &\times \left(\exp(-i\hat{k}'(\hat{x} - \hat{x}_0))\langle \mathbf{k}', s|\frac{d\mathbf{k}'}{\sqrt{2k^{0'}}}\right) \\ &= \int\limits_{\mathcal{H}} |\mathbf{k}, s\rangle\langle \mathbf{k}, s| \frac{d\mathbf{k}}{2k^0} = I. \end{aligned} \tag{20}$$

We note first that a measurement, as follows from equations (19) and (20), with the spatial region $\Omega^{(x)}$ expanded up to its coordinate space $\mathcal{X}$ transforms into an orthogonal decomposition of unity (15), which makes it possible to distinguish uniquely states with nonoverlapping supports.

For example, the probability of obtaining the result 1 with an input state $|\psi_1\rangle$ is given by the relations

$$\Pr_1\{M_1\} = \mathrm{Tr}\{|\psi_1\rangle\langle\psi_1|M_1\}$$

$$\times \int\limits_{\Omega^{(x)}} d\mathbf{x}\left(\int\limits_{\Omega_1^{(k)}} \exp(i\hat{k}(\hat{x}-\hat{x}_0))\psi_1^*(\mathbf{k})\frac{d\mathbf{k}}{2k^0}\right)$$
$$\times\left(\int\limits_{\Omega_1^{(k)}} \exp(-i\hat{k}'(\hat{x}-\hat{x}_0))\psi_1(\mathbf{k}')\frac{d\mathbf{k}'}{2k^{0'}}\right) \quad (21)$$
$$= -(2\pi)^6 \int\limits_{\Omega^{(x)}} d\mathbf{x}\left|\psi_1\left(-i\frac{\partial}{\partial\mathbf{x}}\right)\right|^2 D_0^-(\hat{x}-\hat{x}_0)D_0^+(\hat{x}_0-\hat{x}).$$

Since the amplitude $\psi_1(\mathbf{k})$ possesses a finite support, i.e., it is zero outside the region $\Omega_1^{(k)}$, the region of integration over $\mathbf{k}$ and $\mathbf{k}'$ in equation (21) can be expanded up to the entire space $\mathcal{H}$. We shall assume that the amplitude $\psi_1(\mathbf{k})$ is a sufficiently smooth function so that the argument $\mathbf{k}$ can be replaced by $i\partial/\partial\mathbf{x}$ and the function can be removed from the integrand in the integral over $\mathbf{k}$. Next, taking account of the definition of the function $D_0^-(\hat{x})$ (8), we obtain the final expression. We recall that $D_0^-(\hat{x}) = -D_0^+(-\hat{x})$. The probability of obtaining the result 2 with input state $|\psi_1\rangle$ is identically zero (and similarly for outcome 1 with input state $|\psi_2\rangle$):

$$\Pr_1\{M_2\} = \mathrm{Tr}\{|\psi_1\rangle\langle\psi_1|M_2\}$$
$$= \int\limits_{\Omega^{(x)}} d\mathbf{x}\left(\int\limits_{\Omega_2^{(k)}} \exp(i\hat{k}(\hat{x}-\hat{x}_0))\psi_1^*(\mathbf{k})\frac{d\mathbf{k}}{2k^0}\right) \quad (22)$$
$$\times\left(\int\limits_{\Omega_2^{(k)}} \exp(-i\hat{k}'(\hat{x}-\hat{x}_0))\psi_1(\mathbf{k}')\frac{d\mathbf{k}'}{2k^{0'}}\right) \equiv 0,$$

since the amplitudes $\psi_1(\mathbf{k})$ and $\psi_2(\mathbf{k})$ have nonoverlapping supports. Finally, the probability of obtaining the result $\perp$ with input state $|\psi_1\rangle$ (and similarly for $|\psi_2\rangle$) is

$$\Pr_1\{M_\perp\} = \mathrm{Tr}\{|\psi_1\rangle\langle\psi_1|M_\perp\}$$
$$= \int\limits_{\mathcal{H}\setminus\Omega^{(x)}} d\mathbf{x}\left(\int\limits_{\Omega_1^{(k)}} \exp(i\hat{k}(\hat{x}-\hat{x}_0))\psi_1^*(\mathbf{k})\frac{d\mathbf{k}}{2k^0}\right)$$
$$\times\left(\int\limits_{\Omega_1^{(k)}} \exp(-i\hat{k}'(\hat{x}-\hat{x}_0))\psi_1(\mathbf{k}')\frac{d\mathbf{k}'}{2k^{0'}}\right) \quad (23)$$
$$= -(2\pi)^6 \int\limits_{\mathcal{H}\setminus\Omega^{(x)}} d\mathbf{x}\left|\psi_1\left(-i\frac{\partial}{\partial\mathbf{x}}\right)\right|^2 D_0^-(\hat{x}-\hat{x}_0)D_0^+(\hat{x}_0-\hat{x}).$$

The measurement (19)–(23) has a simple meaning. The formula (21) describes the probability of detecting in a spatial region $\Omega^{(x)}$ a state which possesses a support in $\Omega_1^{(k)}$ (and similarly for $|\psi_2\rangle$).

To detect reliably states with support in $\Omega_1^{(k)}$ it is necessary to have access to the entire spatial region where the state is present. In addition, as one can see from equation (21), because the commutator functions $D_0^\pm(\hat{x}-\hat{x}_0)$ are different from zero only on the light cone, only the points of space-time which are related with one another by a cause-effect coupling, i.e., $(\hat{x}-\hat{x}_0)^2 = 0$, $|\mathbf{x}_0-\mathbf{x}| = c|t_0-t|$, make a contribution. Even though the commutator functions in equations (21) and (22) possess a singularity on the light cone, their product always exists as a generalized function, since the Fourier transforms of the D^- functions possess support in the forward part of the light cone (for a detailed discussion, see, for example, [14]).

The formula (23) describes the probability of detecting a state with support in $\Omega_1^{(k)}$ (and similarly for $|\psi_2\rangle$) in the residual spatial region $\mathcal{H}\setminus\Omega^{(x)}$ on account of the "tails" of the state $|\psi_1\rangle$, which did not "fit into" the spatial region $\Omega^{(x)}$. When the accessible region $\Omega^{(x)}$ is expanded up to a size so that the state $|\psi_1\rangle$ "fits" into it entirely, this measurement becomes a measurement corresponding to orthogonal projectors in (15), which make it possible to distinguish a pair of states with probability 1 (reliably). Correspondingly, the probability of the outcome $\perp$ approaches zero.

The formula (21) has an especially transparent meaning for a state that is strongly localized in coordinate space. In this case, the amplitude of the state in the momentum representation is strongly delocalized (in the limit $|\psi_1(\mathbf{k})|^2) \longrightarrow$ const, the support $\Omega_1^{(k)} \longrightarrow \mathcal{H}$ and, correspondingly, $|\psi(-i\partial/\partial\mathbf{x})|^2$ does not depend on $\mathbf{x}$). The formula (21) becomes

$$\Pr_1\{M_1\} = (2\pi)^6 \int\limits_{\Omega^{(x)}} d\mathbf{x}\left|D_0^-(\hat{x}-\hat{x}_0)\right|^2. \quad (24)$$

If it is now recalled that $-iD_0^-(\hat{x}-\hat{x}_0)$ describes the amplitude of the propagation of a single-particle state of the field created at the point $\hat{x}_0$ into the point $\hat{x}$

$$\langle 0|\psi_1^-(\hat{x})\psi_1^+(\hat{x}_0)|0\rangle = -iD_0^-(\hat{x}-\hat{x}_0), \quad x_0 > x_0^0, \quad (25)$$

then the formula (21) describes the probability of detecting a single-particle state of the field, having support in $\Omega_1^{(k)} \longrightarrow \mathcal{H}$ in the spatial region $\Omega^{(x)}$. It is evident from equation (24) that if the region of integration $\Omega^{(x)}$ does not encompass the point of creation $\hat{x}_0$ of the single-particle state, then the probability of detection is

zero. For a state that is strongly localized in space, a spatial region of arbitrarily small size $\Omega^{(x)} \longrightarrow 0$ can be chosen for reliable detection. An arbitrarily short time is required for the state of the field to fill this region (of course, minus the time required to reach $\Omega^{(x)}$ from the point of creation $\hat{x}_0$ of the field).

However, if the state is not strongly localized in space (in this case $|\psi_1(\mathbf{k})|^2 \neq \text{const}$ and, therefore, $|\psi_1(-i\partial/\partial\mathbf{x})|^2 \neq \text{const}$), then all points of the region $\Omega^{(x)}$ contribute. In addition, the more strongly the state is localized in $\mathbf{k}$ space, the more strongly it is delocalized in coordinate space, and the larger the region $\Omega^{(x)}$ required to detect a given state with probability 1 is. The state of the field cannot fill this region more rapidly than in time $t \approx L/c$, where L is the characteristic size of the region.

For the bit commitment protocol it will be important that the probability of detecting states increases with time (as the field fills the region accessible for measurements). After a finite time T, which depends on the structure of the states themselves, has elapsed, the states become reliably different because of their orthogonality. The states are effectively nonorthogonal (reliably indistinguishable) for times $0 < t < T$. In addition, there are no fundamental restrictions on making them effectively nonorthogonal (reliably indistinguishable) in a preassigned arbitrarily long time interval T, choosing their supports $\Omega^{(k)}_{1,2}$ to be increasingly more localized.

For the protocol, because of the "symmetry" of 0 and 1, it is more convenient to choose states so that the region $\Omega^{(x)}$ is identical (the probability of correct identification as a function of time will then be the same for 0 and 1, i.e., unbiased for 0 and 1). This can always be done for photons, choosing, for example, a pair of narrow-band states with the same frequency width of the spectrum, but with different support frequencies.

We underscore that equations (21)–(23) are of a statistical character. If a small spatial region is accessible for measurements or, equivalently, a determination must be made in a short time interval as to which state $|\psi_1\rangle$ or $|\psi_2\rangle$ is present, then the probability of a correct answer is low (correspondingly, the probability of an error is high), since the main outcomes of the measurements will occur in the channel $M_\perp$ irrespective of the initial state $|\psi_1\rangle$ or $|\psi_2\rangle$. But the probabilities of outcomes in the channels M_1 and M_2 are low to the extent that the region accessible for detection is small. However, if a detector has been triggered in channel M_1 or M_2, then this is already sufficient to say unequivocally what the state is; the probability of such detection is low simply because the accessible region is small.

We shall also present a measurement that depends on the dimensions of the region accessible for measurements and makes it possible to distinguish reliably a pair of specific orthogonal states $|\psi_1\rangle$ and $|\psi_2\rangle$ (and not only states which have nonoverlapping supports) when the spatial region $\Omega^{(x)}$ is expanded up to the entire coordinate space:

$$M_{1,2} = \left(\int\limits_{\Omega^{(x)}} d\mathbf{x} \iint\limits_{\mathcal{H}\mathcal{H}} \psi_{1,2}(\mathbf{k}) \exp(i(\hat{k} - \hat{k}')(\hat{x} - \hat{x}_0)) \right.$$

$$\left. \times |\mathbf{k}\rangle \frac{d\mathbf{k}\,d\mathbf{k}'}{(2k^0 2k^{0'})^{1/4}} \right) \left(\int\limits_{\Omega^{(x)}} d\mathbf{x} \iint\limits_{\mathcal{H}\mathcal{H}} \langle \mathbf{k}| \psi^*_{1,2}(\mathbf{k}) \right. \tag{26}$$

$$\left. \times \exp(-i(\hat{k} - \hat{k}')(\hat{x} - \hat{x}_0)) \frac{d\mathbf{k}\,d\mathbf{k}'}{(2k^0 2k^{0'})^{1/4}} \right),$$

$$M_\perp = I - M_1 - M_2. \tag{27}$$

When the accessible region is expanded up to the entire space ($\Omega^{(x)} \longrightarrow \mathcal{X}$), this measurement transforms into the orthogonal projectors (13).

3. EXAMPLE OF MEASUREMENTS IN THE ONE-DIMENSIONAL CASE

To obtain more specific estimates and to determine the qualitative picture, we shall examine a one-dimensional model situation, since in the three-dimensional case it is necessary to specify the geometry of the spatial regions. Moreover, experimental implementations employ optical fibers, which are a quasi-one-dimensional system, for the quantum channel. Such model one-dimensional schemes are often used in problems of quantum optics.

Let

$$\begin{gathered} |\psi_{1,2}\rangle = \int\limits_0^\infty \psi_{1,2}(k)|k\rangle dk, \\ \rho_{1,2} = |\psi_{1,2}\rangle\langle\psi_{1,2}|, \\ k > 0, \quad \langle k|k'\rangle = \delta(k - k') \end{gathered} \tag{28}$$

be a pair of orthogonal single-photon packets, where $|k\rangle = a_k^+|0\rangle$ is a single-photon monochromatic Fock state (we are considering particles moving in one direction). In addition, in the one-dimensional case the energy is $k^0 = k$ ($c = 1$ is the speed of light).

The states possess nonoverlapping supports with the same spectral width

$$\begin{gathered} E_{1,2} = \operatorname{supp} \psi_{1,2} \\ = \{k\colon k \in (-\Delta/2 + k_{1,2}, k_{1,2} + \Delta/2)\}, \\ E = \{k\colon k \in [0, \infty)\}, \\ k_1 - k_2 \geq \Delta, \quad \int\limits_{E_{1,2}} |\psi_{1,2}(k)|^2 dk = 1. \end{gathered} \tag{29}$$

The states with nonoverlapping supports are orthogonal:

$$\langle\psi_1|\psi_2\rangle = \int_0^\infty \psi_1^*(k)\psi_2(k)dk = 0. \quad (30)$$

A measurement that makes it possible to distinguish reliably the states (28) is given by a decomposition of unity, similar to equation (15),

$$\mathcal{P}_1 + \mathcal{P}_2 + \mathcal{P}_\perp = I, \quad I = \int_0^\infty |k\rangle\langle k|dk,$$

$$\mathcal{P}_{1,2} = \int_{E_{1,2}} |k\rangle\langle k|dk, \quad (31)$$

$$\mathcal{P}_\perp = \int_{E\setminus(E_1 \cup E_2)} |k\rangle\langle k|dk.$$

The coordinate and the time parameter do not yet appear in the decomposition of unity (31). This implicitly implies that the spatial region accessible for measurements is $x \in (-\infty, \infty)$.

In the one-dimensional case a measurement M_i analogous to equations (19) and (20) has the same form but with the three-dimensional region $\Omega^{(x)}$ replaced by a one-dimensional region, which, for brevity, we shall denote by $\mathcal{X} = \{x: x \in (-X, X)\}$. The one-dimensional case is especially clear, since the coordinate and the time parameter are present in the combination $x - ct$. This is actually due to the properties of the fundamental solution of the wave equation in the one-dimensional case, which, as is well known, has the form $\mathcal{E}(x, t) = \propto \theta(ct - |x|)$, in contrast to the three-dimensional case ($\mathcal{E}(x, t) \propto \theta(t)\delta(c^2t^2 - |\mathbf{x}|^2)$) [15]. For this reason, in the one-dimensional case an increase in the spatial region X accessible for measurement can be effectively obtained (with fixed X) by increasing the accessible time interval T. In what follows, with these stipulations, we shall say briefly that the measurements are performed in an accessible time window $(-T, T)$.

It is convenient to write the probability of obtaining a result in a set 1 on the input density matrix ρ_1 in the one-dimensional case in the form (performing first the integration over x)

$$\Pr_1\{M_1\} = \mathrm{Tr}\{\rho_1 M_1\}$$

$$= \int_{E_1}\int_{E_1} \psi_1^*(k)\psi_1(k')\frac{\sin[(k-k')T]}{(k-k')}dkdk', \quad (32)$$

$$T \equiv ct - |X|.$$

The formula (32) is the probability of detecting systems with a density matrix whose support lies in E_1 in the time window $(-T, T)$.

If the time window during which a measurement can be made is small ($T\Delta \ll 1$), then from equation (32) we have

$$\Pr_1\{M_1\} \approx T\Delta \ll 1 \approx 0, \quad (33)$$

i.e., the probability of detecting a state is proportional to the size of the time window. For a large time window ($T\Delta \gg 1$) we find

$$\Pr_1\{M_1\} \approx \frac{1}{\pi}\int_{-T}^{T}\frac{\sin\xi}{\xi}d\xi = 1 - \frac{\cos(T\Delta)}{T\Delta} \approx 1, \quad T\Delta \gg 1. \quad (34)$$

If the input state is ρ_2, then irrespective of the size T of the time window the probability of obtaining a result in channel 1 is zero (and similarly for ρ_1 in channel 2)

$$\Pr_2\{M_1\} = \mathrm{Tr}\{\rho_2 M_1\} = \Pr_1\{M_2\} = \mathrm{Tr}\{\rho_1 M_2\} = 0. \quad (35)$$

The formula (32) can be rewritten in the more familiar form

$$\Pr_1\{M_1\} = \int_{-T}^{T} |\psi_1(\tau)|^2 d\tau,$$

$$\psi_1(\tau) = \int_0^\infty \psi_1(k)e^{-ik\tau}dk, \quad (36)$$

which agrees with the classical intuitive ideas. If a function is localized in the frequency representation, then it is smeared in the time representation, so that a large time window into which $\psi(\tau)$ "fits" completely is required for reliable detection of the signal.

Here we shall make a stipulation. If a classical state (a classical function $\psi(x)$ whose support in the k representation is unknown in advance (supp $\psi(k) = \Delta$)) is presented, then in order to determine the support it is necessary to know the values of the function in a region of coordinate space of size not less than $L_x \approx 1/\Delta$. However, if there are two classical functions $\psi_1(k)$ and $\psi_2(k)$ with nonoverlapping supports and it is only necessary to distinguish one from the other, then it is sufficient to have the value at one of the points x, where $\psi_1(x) \neq \psi_2(x)$. For this reason, to distinguish two classical functions with probability 1, a single point where the functions are not equal is sufficient (if the functions are not specially pathological, i.e., it is assumed that they coincide only on a set of points of measure zero).

To distinguish two orthogonal quantum states $|\psi_1\rangle$ and $|\psi_2\rangle$ with probability 1 it is necessary to have access to the entire spatial region where the states are different from zero (we assume that the states are different from zero in the same spatial region).

For a small time window ($T \ll 1/\Delta$, the reciprocal of the width of the spectrum) the probability of a result is

low to the extent that $T\Delta \ll 1$ is small. If the result is obtained in channels 1 or 2, then this is sufficient to distinguish $|\psi_1\rangle$ and $|\psi_2\rangle$ reliably. However, for a small time window the outcomes on the input state $|\psi_1\rangle$ and on the state $|\psi_2\rangle$ will occur in the channel $\perp$ with an overwhelming probability, i.e., the states are indistinguishable with overwhelming probability (they are effectively nonorthogonal).

The results of measurements in the channel $\perp$ will arise in the time window $(-\infty, \infty)$ on account of states whose supports do not lie in E_1 or E_2 and, moreover, in the time intervals $(-\infty, -T)$ and (T, ∞) on account of the "tails" of states with supports in $E_{1,2}$, which were not detected in the channels 1 and 2 in $(-T, T)$.

The probability of detecting $\rho_{1,2}$ in the channel $\perp$ has the form

$$\Pr_{1,2}\{\perp\} = \int\limits_{E_{1,2}}\int\limits_{E_{1,2}} \left\{\delta(k-k') - \frac{\sin[(k-k')T]}{(k-k')}\right\} \times \psi^*_{1,2}(k)\psi_{1,2}(k')dkdk'. \quad (37)$$

For $T\Delta \gg 1$ (large time window) we have

$$\Pr_{1,2}\{\perp\} \approx \cos(T\Delta)/T\Delta \longrightarrow 0 \quad (38)$$

and for $T\Delta \ll 1$ (small time window)

$$\Pr_{1,2}\{\perp\} \approx \sin(T\Delta)/T\Delta \longrightarrow 1. \quad (39)$$

The formulas (32)–(39) actually mean that if a state with ρ_1 or ρ_2 is presented and it is required to identify the state in a time T, then for $T \ll 1/\Delta$ the probability of a correct answer is $p_+ \approx T\Delta \ll 1$ (correspondingly, the probability of an incorrect answer is $p_- \approx 1 - T\Delta \sim 1$), since the probability of detection in the interval T is low. However, if measurements in a large time window ($T\Delta \gg 1$) are allowed, then the probability of a correct answer is $p_+ \approx 1 - 1/T\Delta \approx 1$ (and the probability of an error is $p_- \approx 0$) and the states are distinguishable reliably ($T\Delta \gg 1$).

For a small time window the states are effectively nonorthogonal (reliably indistinguishable). The effective angle α between ρ_1 and ρ_2 in the interval $(-T, T)$ is small

$$\langle\psi_1|\psi_2\rangle \approx \cos\alpha \approx 1 - T\Delta, \quad \alpha \approx T\Delta \ll 1. \quad (40)$$

4. RELATIVISTIC QUANTUM BIT COMMITMENT PROTOCOL

We shall now describe the bit commitment protocol. The participants in the protocol agree beforehand as to which states $|\psi_1\rangle$ and $|\psi_2\rangle$ correspond to 0 and 1. It is always possible to choose a pair of orthogonal states so that their extent would be much greater than the length of the communication channel, and the rate of decay at infinity would make possible virtually reliable identification in a certain finite time T. For photons it is possible to choose two states in a quite narrow energy band such that the effective extent $L \approx c/\Delta\omega$ ($\Delta\omega$ is the energy width of the spectrum) would be much greater than the length of the communication channel ($L \gg L_{ch}$, L_{ch} is the length of the communication channel). In this case the length of the communication channel can be assumed to be effectively zero. Formally, this means that the participants A and B in the protocol can monitor only neighborhoods of points x_A and x_B and cannot monitor all space outside these points, i.e., outside the neighborhoods x_A and x_B anything can be done that is not forbidden by the laws of relativistic quantum mechanics. The number N of quantum states transmitted by A is also agreed upon beforehand.

The participant A conceives his bit a (0 or 1), which is a parity bit of N quantum states ($a = a_1 \oplus a_2 \oplus \ldots \oplus a_N$). Next, at time $t = 0$ A immediately transmits N states. More accurately, A switches on N sources which form the states. As the states are formed, they start to propagate in the communication channel. The states $|\psi_1\rangle$ and $|\psi_2\rangle$ are, in the general case, nonstationary and by fixing the time $t = 0$ it is possible to "tune" the measurement (13) with a corresponding phase shift for any spatial region which the field reaches as it propagates. The participant B performs the measurement (13) for each state individually. Since the states are orthogonal, they become distinguishable by the time T, when they become completely accessible. In a time $0 < t < T$ the states are completely inaccessible (effectively nonorthogonal), so that they are not distinguishable reliably. The probability of correctly identifying the states is an increasing function of time ($p(t = 0) = 0$ and ($p(t = T) = 1$); the specific form of $p(t)$ depends on the structure of the states and is not essential for us. For individual measurements, the probability of correct identification of a parity bit is $P_+(t) = p^N(t) \ll 1$ at times when $p(t) \ll 1$. In collective measurements, when measurements of all states are performed immediately in order to determine the parity bit, the probability of correct identification is $P_+^{\text{collect}}(t) = \sqrt{p^N(t)} \ll 1$ [16], and it can also be made as small as desired by appropriate choice of the states. It is always possible to choose states so that for any as small as desired $\delta \ll 1$, known beforehand, and as long as desired time t_c (t_c is the conservation of the secret bit conceived by the participant A) the probability $P_+^{\text{collect}}(t)$ of correct identification of the parity bit in time $t < t_c$ can be made as small as desired. This can be achieved by increasing the effective extent of the states (decreasing the width of the spectrum).

The arguments presented above concern the case where the participant B obtains information about states only from quantum-mechanical measurements. We note that the probabilities $p(t)$ appearing above are the probability of detecting states (triggering a classical device), which in times $t < T$ is less than 1 because the normalized states are not completely accessible. However, if detection has occurred, this uniquely makes it

possible to distinguish the states. For this reason, as long as the probability of detection itself is $p_i(t) < 1/2$, the participant B need not make any measurements at all, and he simply can guess what was transmitted along each channel. At times when $p_i(t) > 1/2$, measurements give more information than simply guessing without performing any measurements. As $t \longrightarrow T$ the measurements give almost reliable information about the states.

Thus, the probability of preserving the secret bit $P^{\text{store}}(t)$ conceived by the participant A is a decreasing function of time ($P^{\text{store}}(t = 0) = 1$ and $P^{\text{store}}(t = T) = 0$).

After the quantum part of the protocol is completed for $t > T$, when B now possesses complete access to the states and can distinguish them reliably, the user A reports through the classical channel what he has sent in each of the N quantum channels. If it is discovered that the classical information is inconsistent with the quantum-mechanical measurements in at least one of the channels, the protocol is terminated. The fact that participant A must report classical information after quantum-mechanical measurements and the fact that the states are individually reliably distinguishable (orthogonal) make deception impossible. For this reason, participant A cannot, for example, send mixed states of the type $\rho = |\psi_1\rangle\langle\psi_1| + |\psi_2\rangle\langle\psi_2|$, since this will cause (for large N) the quantum measurements to be inconsistent with the classical information. Likewise, the participant A cannot send some states other than $|\psi_1\rangle$ and $|\psi_2\rangle$, once again because they are orthogonal (reliably distinguishable). Any other states will give (for large N), when using measurements $\mathcal{P}_{1,2,\perp}$, nonzero outcomes in the channels where they should not be present. Because of the existence of a maximum propagation velocity of the states of the field, the user A also cannot, for example, delay sending his states, since nonzero outcomes can arrive in a measurement (13) which "is tuned" to the states $|\psi_1\rangle$ and $|\psi_2\rangle$. Formally, a delay can be described as a translation in space-time. This results in the addition of the phase factor $e^{i\hat{k}\hat{x}_0}$ in the integrand in equation (10). A "shifted" state, for example, $|\psi_1\rangle$, will no longer give an outcome of probability 1 in the channel $\mathcal{P}_1$.

In the nonrelativistic case the use of EPR states by participant A would enable him to delay the choice of his bit before measurements are performed by the participant B [5, 6]. In the relativistic case an EPR attack does not work because of the existence of a maximum propagation velocity of the states of the field. Even though EPR correlations are preserved in the relativistic case, for measurements on entangled two-particle states of the field at points separated by space-like intervals [17], a field cannot propagate from an EPR source faster than the speed of light. For this reason, the use of an EPR pair by participant A for purposes of delay before the second stage of the protocol, in contrast to nonrelativistic quantum mechanics, does not work.

We note once again that for the protocol it is important that the orthogonal states are quantum states. For this reason a finite spatial region is required in order to distinguish one state from another with probability 1. Two classical orthogonal states (functions) can be distinguished with probability 1 on the basis of one point.

The time T required to distinguish states reliably (the time during which the secret parity bit is preserved) is determined by the width $\Delta|\mathbf{k}| \approx \Delta\omega/c$ of the spectrum of photons used in the protocol and is estimated as $T \approx 1/\Delta\omega$. We note that although there are no fundamental reasons why the width of the photon spectrum cannot be made as small as desired, this problem is technically very difficult.

5. CONCLUSIONS

We shall make one remark in conclusion. The possibility of identifying states with unit probability one during a finite time interval T depends on whether or not for particles of a given type there exist states with a finite support in space. If the particles are photons, then today only states with exponential localization of energy and detection velocity are known [18]. The latter means, formally, that a result with unit probability can be obtained only over an infinite time. This is not too important for the protocol, since the time interval can be chosen to be long enough so as to obtain a probability of distinguishing two orthogonal states exponentially close to one.

The experimental implementation is very simple (at least speculatively). It is sufficient to have N wide-band sources. The wide-band sources mean that when they are switched on, the sources emit signals which are localized in space and time, i.e., signals with a wide frequency spectrum can be prepared at the time $t = 0$. Before entering the communication channel the signals are passed through narrow-band filters and then through attenuators in order to reach the single-photon level. Transmission through a narrow-band filter requires, effectively, either a long time or a great deal of space. Detection is performed with narrow-band photodetectors.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project no. 99-02-18127) and project no. 02.04.5.2.40.T.50 of the program "Promising Technologies and Devices for Micro- and Nanoelectronics" as well as the program "Physics of Quantum and Wave Processes" (subprogram "Quantum Computations").

REFERENCES

1. C. H. Bennett and G. Brassard, in *Proceeding of the International Conference on Computers, Systems, and Signal Processing, New York, 1984* (IEEE, New York, 1984), p. 175.

2. G. Brassard and C. Crépeau, Lecture Notes in Computer Science **537**, 49 (1991).
3. G. Brassard, C. Crépeau, R. Jozsa, and D. Langlois, in *Proceedings of the 34th Annual IEEE Symposium*, 1993 (Computer Society Press, Los Alamitos, 1993), p. 362.
4. M. Ardehali, quant-ph/9603019.
5. H.-K. Lo and H. F. Chau, Phys. Rev. Lett. **78**, 3410 (1997).
6. D. Mayers, Phys. Rev. Lett. **78**, 3414 (1997).
7. A. Kent, Phys. Rev. Lett. **83**, 1447 (1999).
8. A. Kent, E-print archive, quant-ph/9906103 (1999).
9. W. K. Wootters and W. H. Zurek, Nature **299**, 802 (1982).
10. C. H. Bennett, Phys. Rev. Lett. **68**, 3121 (1992).
11. L. Goldenberg and L. Vaidman, Phys. Rev. Lett. **75**, 1239 (1995).
12. N. N. Bogolyubov and D. V. Shirkov, *Introduction to the Theory of Quantized Fields* (Nauka, Moscow, 1984, 4th ed.; Wiley, New York, 1980, 3rd ed.).
13. A. S. Kholevo, *Probability and Statistical Aspects of Quantum Theory* (Nauka, Moscow, 1980).
14. N. N. Bogolyubov, A. A. Logunov, and I. T. Todorov, *Introduction to Axiomatic Quantum Field Theory* (Nauka, Moscow, 1969; Benjamin, New York, 1975).
15. V. S. Vladimirov, *Equations of Mathematical Physics* (Nauka, Moscow, 1971, 2nd ed.; Dekker, New York, 1971).
16. Tal Mor, E-print archive, quant-ph/9906073 (1999).
17. S. N. Molotkov and S. S. Nazin, Pis'ma Zh. Éksp. Teor. Fiz. **70**, 53 (1999) [JETP Lett. **70**, 54 (1999)].
18. I. Bialynicki-Birula, Phys. Rev. Lett. **80**, 5247 (1998).

Translation was provided by AIP