New critical properties of the Nambu–Jona-Lasinio model with nonzero chemical potential

A. S. Vshivtsev[1)]

Moscow State Institute of Radioengineering, Electronics, and Automatic Systems, 117454 Moscow, Russia

V. Ch. Zhukovskiĭ[2)]

M. V. Lomonosov Moscow State University, 119899 Moscow, Russia

K. G. Klimenko[3)]

Institute of High-Energy Physics, 142284 Protvino, Moscow Region, Russia

(Submitted 26 June 1996)
Zh. Éksp. Teor. Fiz. **111**, 1921–1934 (June 1997)

It is shown for the first time that the massive-fermion state in the Nambu–Jona-Lasinio model with nonzero chemical potential μ is described by two different phases, the transition between which is second-order. It is proved that both first- and second-order phase transitions, depending on the values of the model parameters, restore the chiral symmetry of the model. Two tricritical points exist in the (μ,M) phase diagram (M is the dynamical fermion mass at $\mu=0$). © *1997 American Institute of Physics.* [S1063-7761(97)00106-6]

1. INTRODUCTION

The idea of spontaneous symmetry breaking plays an important role in the development of elementary-particle physics. At present there exist two well-known methods of spontaneous symmetry breaking. In the first one spontaneous symmetry breaking occurs in theories with auxiliary Higgs fields, where spontaneous breaking actually occurs at the classical-action level (the standard theory of electroweak interactions, grand unification theory, and others are based on this approach). The cost of this approach is the mandatory existence of as yet undiscovered Higgs bosons.

In the other approach the spontaneous symmetry breaking occurs in a dynamical manner, i.e., through radiative corrections to the classical action, and then there is no need to introduce Higgs fields. This mechanism of symmetry breaking was first discovered in models with a four-fermion interaction.[1,2] The simplest such model is described by a Lagrangian of the form

$$L=\sum_{k=1}^{N}\bar{\psi}_k i\hat{\partial}\psi_k+\frac{C}{2N}[(\bar{\psi}_k\psi_k)^2+(\bar{\psi}_k i\gamma_5\psi_k)^2] \tag{1}$$

and is called the Nambu–Jona-Lasinio model. In order to use the $1/N$ expansion, we shall study an N-fermion version of the model that is symmetric under the simplest continuous chiral transformations

$$\psi_k\rightarrow e^{i\theta\gamma_5}\psi_k \quad (k=1,...,N). \tag{2}$$

Theories with a four-fermion interaction find application for the explanation of superconductivity[3] and high-T_c superconductivity.[4] Lagrangians with a four-fermion interaction have been used before in discussions of superconductivity.[5–8] It should be noted that the electronic states near the Fermi surface are very important in the investigation of superconductivity (just as in other transport phenomena). As is well known, the Fermi energy for free electrons equals

$$\varepsilon_F=\frac{h^2(3\pi^2n_e)^{2/3}}{m_e}, \tag{3}$$

where m_e and n_e are the electron mass and density, respectively, and h is Planck's constant.

The potential interaction energy of two electrons separated by an average distance $\langle r\rangle\sim n_e^{-1/3}$ can be estimated as

$$\langle V_{\text{pot}}\rangle\approx\frac{e^2}{4\pi\varepsilon_0\langle r\rangle}=\frac{e^2n_e^{1/3}}{4\pi\varepsilon_0}. \tag{4}$$

The quantity

$$\alpha=\frac{\langle V_{\text{pot}}\rangle}{\varepsilon_F}=\frac{m_ee^2}{2\pi\varepsilon_0(3\pi^2)^{2/3}h^2n_e^{1/3}} \tag{5}$$

largely characterizes the properties of superconductors and is used for constructing different approximations. In metals it satisfies $\alpha\geqslant 1$, i.e., α is not small. In consequence, the interelectron Coulomb interaction as well as the electron–phonon interaction must be taken into account. The Coulomb interaction can result in both an exchange interaction and a correlation between the mutual displacements of the electrons. Such effects are important in calculations of the dispersion laws of electrons and in the construction of the Fermi surface.[9] On this basis, the phenomenological expressions obtained for regions $\alpha\geqslant 1$ are ordinarily used to determine these characteristics for specific metals.

In discussions of superconductivity (and other transport phenomena) the Fermi surface is ordinarily considered to be given, but the structure of the surface itself can be very complicated.[9] In the ground state all levels with energy $\varepsilon<\varepsilon_F$ are filled and all states with energy $\varepsilon>\varepsilon_F$ are free. Further, elementary excitations in real metals are assumed to be the same as in a system of noninteracting electrons with the same Fermi surface. Strictly speaking, this is not so, since the excited states in a system of strongly interacting particles can be substantially different from the excited states

1063-7761/97/061047-07$10.00

in a system of noninteracting particles. To give a complete description of the excited states it is necessary to solve the Coulomb-interaction problem, i.e., a self-consistent many-body problem.

For ordinary low-T_c superconductors the difference between the excited states is small. This has led to good agreement between the results of numerical calculations for the superconducting parameters and experiments. In the case of high-T_c superconductivity, where n_e is lower than in ordinary metals, the Coulomb interaction (Coulomb correlations) apparently can cause the excited states of the high-T_c superconductor to differ substantially from the corresponding states for the situation when the interaction is neglected. As a result of both technical and mathematical difficulties, to describe high-T_c superconductivity theoreticians are compelled to use well-proven field-theoretic models, taking account of the different values of the physical parameters. In addition, these model problems, like the results obtained with them, must be treated with caution. This is because different types of interactions are sometimes taken into account very formally in these models by including an interaction of the corresponding type in the Lagrangian, and this makes it difficult to identify the model results with the physical data. With respect to the Lagrangian (1), it should be noted that only one aspect of the question is studied—the breaking of chiral symmetry in the presence of a finite chemical potential (we note that superconductivity is ordinarily attributed to the breaking of $U(1)$ symmetry of this model). The introduction of a formal chemical potential can be substantiated, for example, in a generalized Nambu–Jona-Lasinio model with vector particles (ρ mesons). The drawbacks inherent in this model—nonrenormalizability and existence of a cutoff parameter $\Lambda \sim 4\pi f_\pi \approx 1.2$ GeV—are largely compensated by its predictive power in the low-energy approximation of hadron physics. For example, this model has been used actively to describe the low-energy physics of mesons,[10] construct alternative models of the electroweak interaction,[11] and so on. Recently, it was discovered on the basis of these models that external magnetic fields can catalyze spontaneous symmetry breaking[12] (see also Ref. 13). All this shows why interest in Nambu–Jona-Lazinio type models has remained undiminshed for more than 30 years with special attention being devoted to investigations of the structure and the critical properties of the vacuum in the presence of a surrounding medium, i.e., taking into account factors such as the temperature and chemical potential,[14,15] different external fields,[16] and the nontrivial topology and curvature of space-time.[17]

On this basis, in the present paper we investigate in detail the phase structure of the four-dimensional Nambu–Jona-Lasinio model with a nonzero chemical potential μ. In contrast to previous works on this subject,[14,15] we have discovered a richer phase structure of the model as well as first- and second-order chiral phase transitions, depending on the values of the initial parameters (a more detailed discussion of the results is contained in the concluding section).

2. PHASE STRUCTURE OF THE MODEL WITH $\mu = 0$

First, we recall the well-known properties of the model with a zero chemical potential μ. To investigate the properties of the vacuum of the Nambu–Jona-Lasinio model it is convenient to employ instead of the Lagrangian (1) the auxiliary Lagrangian

$$\overline{L} = \overline{\psi} i\hat{\partial}\psi - \overline{\psi}(\sigma_1 + i\gamma_5\sigma_2)\psi - \frac{N}{2G}(\sigma_1^2 + \sigma_2^2) \tag{6}$$

(here, to simplify the equations, the index k labelling the Fermi field is dropped). This auxiliary Lagrangian for the equations of motion for the fields $\sigma_{1,2}$ is equivalent to the initial Lagranian (1) of the theory.

The effective action of this model to leading order in the parameter $1/N$ is

$$\exp(iS_{\rm eff}(\sigma_{1,2})) = \int D\overline{\psi}D\psi \exp\left(i\int \overline{L}d^4x\right),$$

where

$$\frac{1}{N}S_{\rm eff}(\sigma_{1,2}) = -\int d^4x\,\frac{\sigma_1^2+\sigma_2^2}{2G} - i\ln\det(i\hat{\partial} - \sigma_1 - i\gamma_5\sigma_2). \tag{7}$$

Assuming here that the fields $\sigma_{1,2}$ do not depend on the space–time coordinates, we have by definition

$$S_{\rm eff}(\sigma_{1,2}) = -V_{\rm eff}(\sigma_{1,2})\int d^4x, \tag{8}$$

$$\frac{1}{N}V_{\rm eff}(\sigma_{1,2}) = \frac{\Sigma^2}{2G} + 2i\int\frac{d^4p}{(2\pi)^4}\ln(\Sigma^2 - p^2) \equiv \frac{1}{N}V_0(\Sigma), \tag{9}$$

where $\Sigma = \sqrt{\sigma_1^2+\sigma_2^2}$. Switching in Eq. (9) to the Euclidean metric ($p_0 \to ip_0$) and introducing an Lorentz-invariant cutoff of the integration region ($p^2 \leqslant \Lambda^2$) we obtain

$$\frac{1}{N}V_0(\Sigma) = \frac{\Sigma^2}{2G} - \frac{1}{16\pi^2}\left\{\Lambda^4\ln\left(1+\frac{\Sigma^2}{\Lambda^2}\right) + \Lambda^2\Sigma^2 - \Sigma^4\ln\left(1+\frac{\Lambda^2}{\Sigma^2}\right)\right\}. \tag{10}$$

The condition that the function (10) be stationary is

$$\frac{1}{N}\frac{\partial V_0(\Sigma)}{\partial\Sigma} = 0,$$

$$\frac{\Sigma}{4\pi^2}\left\{\frac{4\pi^2}{G} - \Lambda^2 + \Sigma^2\ln\left(1+\frac{\Lambda^2}{\Sigma^2}\right)\right\} \equiv \frac{\Sigma}{4\pi^2}F(\Sigma) = 0. \tag{11}$$

Hence it is obvious that for $G < G_c = 4\pi^2/\Lambda^2$ Eq. (11) has no solutions except $\Sigma = 0$, i.e., the fermions are massless and chiral symmetry (2) is not broken.

For $G > G_c$ a nontrivial solution $\Sigma_0(G,\Lambda) \neq 0$ of Eq. (11) such that $F(\Sigma_0) = 0$ appears. It follows from Eq. (11) that the derivative of the function $V_0(\Sigma)$ is negative for $0 \leqslant \Sigma < \Sigma_0$ and positive for $\Sigma > \Sigma_0$. Therefore the potential $V_0(\Sigma)$ has a global minimum at the point $\Sigma_0 \neq 0$, signifying spontaneous breaking of the chiral symmetry of the model and the appearance of a fermion mass

$$M \equiv \Sigma_0(G,\Lambda). \tag{12}$$

It is obvious that the fermion mass depends on the coupling constant G and the cutoff parameter Λ.

In what follows, for $G>G_c$, we shall use instead of G the mass M (12) of the Fermi particles as the independent parameter of the theory. In this case, the constant G can be expressed in terms of H and Λ by means of the stationarity equation (11) and the following equivalent expression can be derived for $V_0(\Sigma)$:

$$\frac{16\pi^2}{N}V_0(\Sigma)=\Sigma^2\Lambda^2-2\Sigma^2M^2\ln\left(1+\frac{\Lambda^2}{M^2}\right)-\Lambda^4\ln\left(1+\frac{\Sigma^2}{\Lambda^2}\right)+\Sigma^4\ln\left(1+\frac{\Lambda^2}{\Sigma^2}\right). \tag{13}$$

We note once again that for all $G>0$ the function (10) is the effective potential of the model. However, in the phase in which chiral symmetry is broken spontaneously (i.e., for $G>G_c$) the expression (13) can be used for V_{eff}. Then the fermion mass M is a free parameter and the constant G can be found from the equation $F(M)=0$ (the function $F(x)$ is given in Eq. (11)).

3. TAKING ACCOUNT OF THE FINITENESS OF THE CHEMICAL POTENTIAL

Let us assume temporarily that, together with the chemical potential μ, a thermostat with temperature T acts on the system described by the Nambu–Jona-Lasinio Lagrangian (1). In this case, to obtain the effective potential $V_{\mu T}(\Sigma)$, the following transformation of the integration measure must be made in Eq. (9):

$$\int\frac{dh_0}{2\pi}\to iT\sum_{n=-\infty}^{\infty},\quad p_0\to i\pi T(2n+1)+\mu.$$

After summing over n, i.e., over the Matsubara frequencies in the expression obtained, we obtain the effective potential for $T,\mu\neq 0$

$$\frac{1}{N}V_{\mu T}(\Sigma)=\frac{1}{N}V_0(\Sigma)-2T\int\frac{d^3p}{(2\pi)^3}\ln\{|1+\exp(-\beta(\sqrt{\Sigma^2+\mathbf{p}^2}+\mu))||1+\exp(-\beta(\sqrt{\Sigma^2+\mathbf{p}^2}-\mu))|\}, \tag{14}$$

where $\beta=1/T$, and $V_0(\Sigma)$ is given in Eqs. (10)–(13). We now let T in Eq. (14) pass to zero:

$$\frac{1}{N}V_\mu(\Sigma)=\frac{1}{N}V_0(\Sigma)-2\int\frac{d^3p}{(2\pi)^3}\,\theta(\mu-\sqrt{\Sigma^2+\mathbf{p}^2})\times(\mu-\sqrt{\Sigma^2+\mathbf{p}^2}). \tag{15}$$

This expression, which is the effective potential in the Nambu–Jona-Lasinio model with a nonzero chemical potential, will be investigated below for a global minimum in order to obtain the phase structure of the model and hence the properties of the vacuum as a function of the values of μ, M, and Λ. First, however, we integrate in Eq. (15) over the momentum variables

$$\frac{1}{N}V_\mu(\Sigma)=\frac{1}{N}V_0(\Sigma)-\frac{\theta(\mu-\Sigma)}{16\pi^2}\left\{\frac{10}{3}\mu(\mu^2-\Sigma^2)^{3/2}-2\mu^3\sqrt{\mu^2-\Sigma^2}+\Sigma^4\ln\frac{(\mu+\sqrt{\mu^2-\Sigma^2})^2}{\Sigma^2}\right\}. \tag{16}$$

Starting with this formula, we assume everywhere below that $\Sigma\geqslant 0$. There is no loss of generality in making this restriction, since (15) is odd as a function of the variable Σ.

The condition for the function $V_\mu(\Sigma)$ to be stationary has the form

$$\frac{1}{N}\frac{\partial V_\mu(\Sigma)}{\partial\Sigma}=0=\frac{\Sigma}{4\pi^2}\left\{\frac{4\pi^2}{G}-\Lambda^2+\Sigma^2\ln\left[1+\frac{\Lambda^2}{\Sigma^2}\right]+\theta(\mu-\Sigma)\left[2\mu\sqrt{\mu^2-\Sigma^2}-2\Sigma^2\ln\frac{\mu+\sqrt{\mu^2-\Sigma^2}}{\Sigma}\right]\right\}. \tag{17}$$

Let $G<G_c=4\pi^2/\Lambda^2$. In this case, for $\Sigma>\mu$ Eq. (17) has the same form as the equation of stationarity (11) with $\mu=0$ and therefore it has no solutions. If $\Sigma<\mu$, then Eq. (17) reduces to the equation

$$\frac{\Sigma}{4\pi^2}\left\{\frac{4\pi^2}{G}-\Lambda^2+\Sigma^2\ln\left[1+\frac{\Lambda^2}{\Sigma^2}\right]+2\mu\sqrt{\mu^2-\Sigma^2}-2\Sigma^2\ln\frac{\mu+\sqrt{\mu^2-\Sigma^2}}{\Sigma}\right\}=0, \tag{18}$$

which, as one can easily see, has no solutions other than $\Sigma=0$.

In summary, for values of the coupling constant G less than G_c the Nambu–Jona-Lasinio theory lies in the massless chirally symmetric phase irrespective of the chemical potential $\mu\geqslant 0$.

4. PHASE STRUCTURE OF THE MODEL FOR $G>G_C$ AND $\mu>0$

In this section we shall examine the phase structure of the Nambu–Jona-Lasinio model for $\mu\neq 0$ and $G>G_c$. As already noted in Sec. 2, in this case there is a one-to-one correspondence between the values of the constant G and the dynamically arising fermion mass M. Specifically,

$$\frac{4\pi^2}{G}-\Lambda^2=-M^2\ln\left(1+\frac{\Lambda^2}{M^2}\right). \tag{19}$$

Let us consider the set of all pairs of nonnegative values of the parameters (μ,M). Our problem is to indicate the phase of the Nambu–Jona-Lasinio model for each element of this set.

A) Let $\mu<M$.

Let us investigate the equation of stationarity (17) in this case. It is obvious that for $\Sigma>\mu$ it has the solution $\Sigma_1=M$. For $\Sigma<\mu$ Eq. (17) assumes the form

$$\Sigma f_\mu(\Sigma)\equiv\Sigma\left\{2\mu\sqrt{\mu^2-\Sigma^2}-M^2\ln\left(1+\frac{\Lambda^2}{M^2}\right)\right.$$

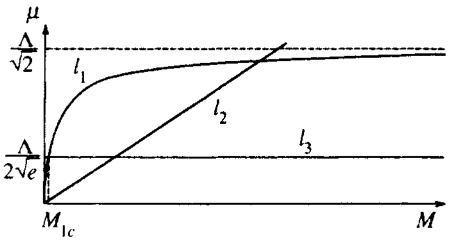


FIG. 1.

$$+\Sigma^2 \ln \frac{\Sigma^2+\Lambda^2}{(\mu+\sqrt{\mu^2-\Sigma_2})^2}\Bigg\} = 0. \tag{20}$$

The domain of the function $f_\mu(\Sigma)$ is $\Sigma \in (0,\ \mu)$. Figure 1 displays three curves l_1, l_2, and l_3:

$$l_1:\quad \mu=\left[\frac{M^2}{2}\ln\left(1+\frac{\Lambda^2}{M^2}\right)\right]^{1/2}\equiv\mu_{1c}(M),$$

$$l_2:\quad \mu=M,$$

$$l_3:\quad \mu=\frac{\Lambda}{2\sqrt{e}}. \tag{21}$$

The quantities $f_\mu(0)$ and $f_\mu(\mu)$ vanish on the curves l_1 and l_2, respectively, and the function $\varphi_\mu(\Sigma)$ vanishes on the curve l_3 (see Eq. (A5) in the Appendix) at the point $\Sigma=0$. These and other properties of the function $f_\mu(\Sigma)$ are presented in the Appendix, whence it follows that if the point (μ,M) lies below the curves l_1 and l_2, then $f_\mu(\Sigma)$ as a function of Σ is negative on the interval $(0,\mu)$. This situation is reflected in Fig. 2 by two different plots of the function $f_\mu(\Sigma)$: Curve *1* corresponds to values of the chemical potential $\mu<\Lambda/2\sqrt{e}$ and for curve *2* we have $\mu>\Lambda/2\sqrt{e}$. Therefore, in this case Eq. (20) has only the trivial solution $\Sigma_2=0$. If the point (μ,M) lies below l_2 but above l_1, then the equation $f_\mu(\Sigma)=0$ will have the unique solution $\Sigma_3(\mu)$ (see curve *3* in Fig. 2, which qualitatively represents the behavior of the function $f_\mu(\Sigma)$ in this case). However, in $\Sigma_3(\mu)$ the effective potential $V_\mu(\Sigma)$ assumes a value greater than at the point $\Sigma_2=0$ (because the function $f_\mu(\Sigma)$ is positive on the interval $(0,\Sigma_3(\mu))$, i.e., on the same interval where the derivative $dV_\mu(\Sigma)/d\Sigma$ is positive and the function $V_\mu(\Sigma)$ increases monotonically).

Therefore the global minimum of the effective potential for $\mu<M$ lies at one of the two points $\Sigma_1=M$ and $\Sigma_2=0$. The first point corresponds to the massive phase. The interphase boundary for $\mu<M$ is determined by the equation

$$V_\mu(0)=V_\mu(M). \tag{22}$$

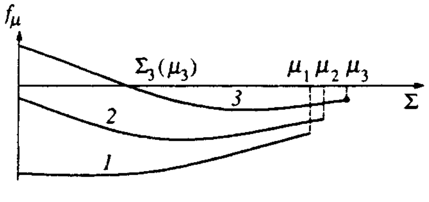


FIG. 2.

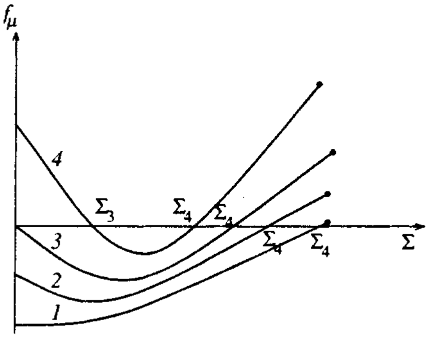


FIG. 3.

Figure 4 shows the interphase boundary in the form of the critical curve $\mu=\mu_{2c}(M)$, which can readily be represented, using Eqs. (13), (16), and (22), in the analytical form

$$\mu_{2c}(M)=\left[\frac{3}{4}\left(M^4\ln\left(1+\frac{\Lambda^2}{M^2}\right)\right.\right.$$
$$\left.\left.+\Lambda^4\ln\left(1+\frac{M^2}{\Lambda^2}\right)-M^2\Lambda^2\right)\right]^{1/4} \tag{23}$$

for all $M\geqslant M_{2c}$ $(\mu_{2c}(M_{2c})=M_{2c})$, where

$$M_{2c}=\Lambda/\sqrt{4.895676}. \tag{24}$$

In the theory a first-order phase transition occurs from the massive phase into the massless phase and vice versa on passing through the critical point $\mu_{2c}(M)$, since here the value of the order parameter (fermion mass) changes abruptly.

B) Let $\mu>M$.

In this case, the equation of stationarity (17) obviously has no solutions for $\Sigma>\mu$. For $\Sigma<\mu$ Eq. (17) reduces to Eq. (20) and possesses the explicit solution $\Sigma_2=0$. Let us assume now that the point (μ,M) falls between the curves l_1 and l_2 (see Fig. 1) and, moreover, that $M<M_{1c}$, where M_{1c} is the point of intersection of l_1 and l_3 (18). M_{1c} is the root of the equation

$$\frac{M_{1c}^2}{2}\ln\left(1+\frac{\Lambda^2}{M_{1c}^2}\right)=\frac{\Lambda^2}{4e}. \tag{25}$$

On the basis of the results obtained in the Appendix, it is obvious that for such points (μ,M) the function $f_\mu(\Sigma)$ increases monotonically on the interval $(0,\mu)$ and possesses a

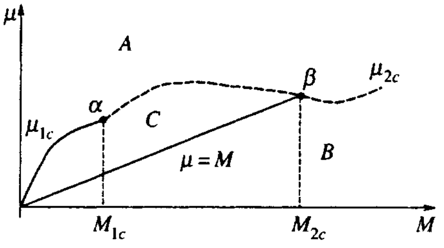


FIG. 4.

unique zero $\Sigma_4(\mu,M)$ where the global minimum of the potential $V_\mu(\Sigma)$ will lie. The curve *1* in Fig. 3 corresponds to this case. The point is that $\Sigma_2=0$ is a local minimum of the potential in this case, since for all $\Sigma \in (0,\Sigma_4(\mu,M))$ the derivative of the function $V_\mu(\Sigma)$ with respect to the variable Σ is negative and therefore $V_\mu(\Sigma)$ is a decreasing function on this interval, i.e., $V_\mu(0)>V_\mu(\Sigma_4(\mu,M))$. The solution $\Sigma_4(\mu,M)$ possesses the following properties (for $0<M<M_{1c}$):

$$\Sigma_4(\mu,M)\to M, \quad \text{if } (\mu,M)\to l_2,$$

$$\Sigma_4(\mu,M)\to 0, \quad \text{if } (\mu,M)\to l_1. \tag{26}$$

In the case when the point (μ,M) lies above the curve l_1 (for $0<M<M_{1c}$) the equations of stationarity (17) and (20) will have a unique solution $\Sigma_2=0$ where the potential $V_\mu(\Sigma)$ possesses a global minimum.

Therefore, in this range of the parameters μ and M (i.e. $\mu>M$ and $M<M_{1c}$) the curve l_1 is a critical curve of second-order phase transitions, since on l_1 the order parameter is a continuous function of the variables μ and M (see Eq. (26)). The region corresponding to the massive phase lies below the curve l_1 and the massless phase lies above l_1; an analytic expression relating the chemical potential μ and M on l_1 is given in Eq. (21).

Let us now assume that we are in the region where $\mu>M$ and $M>M_{1c}$ hold. Here, also, $\Sigma_2=0$ is the solution of the equations of stationarity (17) and (20) and $\Sigma_1=M$ is not their solution. In this case the behavior of the function $f_\mu(\Sigma)$ depends strongly on where the point (μ,M) lies. If this point lies below the curve l_3 (see Fig. 1), then a plot of the function $f_\mu(\Sigma)$ looks like the curve *1* in Fig. 3; if (μ,M) lies above l_3 but below l_1, then $f_\mu(\Sigma)$ corresponds to the curve *2* in Fig. 3. For $(\mu,M) \in l_1$ the curve $f_\mu(\Sigma)$ has the form *3*. Finally, for (μ,M) lying above but ''near'' l_1 the curve *4* in Fig. 3 describes the qualitative behavior of $f_\mu(\Sigma)$. For even larger values of the chemical potential the function $f_\mu(\Sigma)$ is strictly greater than zero in its entire domain $\Sigma \in (0,\mu)$, and for this reason the only solution $\Sigma_2=0$ of Eq. (20) in this case corresponds to the massless symmetric phase of the model.

The following conclusions can be drawn from Fig. 3 (recall that $\mu>M$ and $M>M_{1c}$). For points located no higher than l_1 (see Fig. 1), the corresponding equation of stationarity (20) has a unique nontrivial solution $\Sigma_4(\mu,M) \neq 0$ (in Fig. 3 these are the points where the curves *1–3* intersect the Σ axis), at which the potential evidently possesses a global minimum. The phase of the massive-fermion model with mass $\Sigma_4(\mu,M)$ lies here.

In the case when (μ,M) lies above the curve l_1 the plot of the function $f_\mu(\Sigma)$ can cross the Σ axis (see Fig. 3) at a point, in addition to Σ_4, $\Sigma_3(\mu)$ where the potential $V_\mu(\Sigma)$ possesses a local minimum (we discussed a similar situation in paragraph A).

Therefore, in the range of parameters μ and M under study an absolute minimum of the potential can lie either at $\Sigma_2=0$ or $\Sigma_4(\mu,M)$ (the latter point is the solution of Eq. (20) which equals M in the limit $\mu\to M$) and the interphase boundary is given by the equation

$$V_\mu(0)=V_\mu(\Sigma_4(\mu,M)). \tag{27}$$

It follows from the preceding discussion that the line $\mu=\mu_{3c}(M)$ lies no lower than l_1. (The curve l_1 for $\mu>M$ and $M>M_{1c}$ is most likely no longer a critical curve, i.e., the boundary between the massive and massless phases, as in the case $\mu>M_{1c}$. If, however, phase transitions still occur on l_1, then only first-order transitions occur, since on l_1 the order parameter $\Sigma_4(\mu,M) \neq 0$ in this case will be a continuous function of the variables μ and M.) Furthermore, it is obvious that

$$\mu_{3c}(M_{1c})=\mu_{1c}(M_{1c})=\Lambda/2\sqrt{e},$$

$$\mu_{3c}(M_{2c})=\mu_{2c}(M_{2c})=M_{2c}=\Lambda/\sqrt{4.89.}\ .$$

and the critical point $\mu_{3c}(M)$ is a curve of first-order phase transitions. The results of this section are presented in Fig. 4, where critical curves of the second $(-\mu_{1c}\equiv l_1)$ and first (μ_{2c},μ_{3c}) kinds, which separate the massless symmetric phase *A* of the Nambu–Jona-Lasinio model from the (μ,M) domain corresponding to the massive state of the model, are plotted. We have actually shown that the point α in Fig. 4 is a tricritical point, since second- and first-order phase transitions occur in an arbitrarily small neighborhood of this point.

5. NEW PHASE OF THE NAMBU–JONA-LASINIO MODEL

In the present section we shall show that the Nambu–Jona-Lasinio model can exist in the massive state in two different phases. In Fig. 4 the letter *B* designates the region where fermions possess a mass M, and in the region *C* the fermion mass equals $\Sigma_4(\mu,M)$; in addition, on the line separating *B* in *C* regions, i.e., for $\mu=M$, the fermion mass is a continuous function of the parameters μ and M since $\Sigma_4(M,M)=M$.

We wish to show that the line $\mu=M$ is indeed the boundary between the two different phases. As is well known,[18] the criterion for a phase transition is that a jump be present in some partial derivative of one of the thermodynamic potentials of the system at the interphase boundary. In our case it is most convenient to employ the thermodynamic potential $\Omega(\mu)$, which is the value of the effective potential of the model at the point of the global minimum. It follows from Eqs. (13)–(16) that in the region *B*

$$\Omega_B(\mu)=V_\mu(M)=V_0(M)=\frac{N}{16\pi^2}\left\{M^4-M^4\right.$$
$$\left.\times\ln\left(1+\frac{\Lambda^2}{M^2}\right)-\Lambda^4\ln\left(1+\frac{M^2}{\Lambda^2}\right)\right\}. \tag{28}$$

It is clear that Ω_B does not depend on μ, so that its derivatives with respect to μ all vanish identically in the region *B* as well as on its boundary $\mu=M$.

In the region *C* the thermodynamic potential of the model equals the effective potential (16) at the point $\Sigma_4(\mu,M)$, which is a solution of Eq. (20) and possesses the property $\Sigma_4(M,M)=M$, i.e.,

$$\Omega_C(\mu)=V_\mu(\Sigma_4(\mu,M)). \tag{29}$$

Obviously, $\Omega_B(M)=\Omega_C(M)$. The first derivative of the function $\Omega_C(\mu)$ is

$$\frac{d\Omega_C(\mu)}{d\mu}=\left\{\frac{\partial V_\mu(\Sigma)}{\partial\mu}+\frac{\partial V_\mu(\Sigma)}{\partial\Sigma}\frac{\partial\Sigma}{\partial\mu}\right\}\Bigg|_{\Sigma=\Sigma_4(\mu,M)}. \tag{30}$$

Since Σ_4 (in what follows we shall often employ the abbreviated notation $\Sigma_4=\Sigma_4(\mu,M)$) is a solution of the stationarity equation (20), the second term in braces in Eq. (30) vanishes. Taking account of Eq. (16), we obtain

$$\frac{d\Omega_C(\mu)}{d\mu}=\frac{\partial V_\mu(\Sigma)}{\partial\mu}\Bigg|_{\Sigma=\Sigma_4}=\frac{N}{3\pi^2}(\mu^2-\Sigma_4^2)^{3/2}. \tag{31}$$

The expression (31) vanishes as $\mu\to M_+$, i.e., on the curve $\mu=M$, since here $\Sigma_4\to M$, and hence the first derivative of the thermodynamic potential of the model is a continuous function on this curve. To calculate the higher-order derivatives of $\Omega_C(\mu)$ we shall require the following relations for the derivatives of the function $\Sigma_4(\mu,M)$, which can be easily obtained from Eq. (20) which gives Σ_4 as an implicit function of μ:

$$\frac{d\Sigma_4}{d\mu}\equiv\Sigma_4'=\left\{\frac{\partial f_\mu(\Sigma)}{\partial\mu}\left[\frac{\partial f_\mu(\Sigma)}{\partial\Sigma}\right]^{-1}\right\}\Bigg|_{\Sigma=\Sigma_4}$$

$$=\frac{-2\sqrt{\mu^2-\Sigma_4^2}}{\Sigma_4\{\ln[(\Sigma_4^2+\Lambda^2)/(\mu+\sqrt{\mu^2-\Sigma_4^2})^2]-\Lambda^2/(\Sigma_4^2+\Lambda^2)\}}, \tag{32}$$

$$\frac{d^2\Sigma_4}{d\mu^2}\equiv\Sigma_4''=-\frac{2(\mu-\Sigma_4\Sigma_4')}{\Sigma_4\sqrt{\mu^2-\Sigma_4^2}\{\ldots\}}+O(\mu^2-\Sigma_4^2), \tag{33}$$

where it is understood that the expression in braces in Eq. (33) is the same as the expression in braces in Eq. (30). It is obvious that $\Sigma_4'\to 0$ as $\mu\to M$, but the second derivative satisfies $\Sigma_4''\to-\infty$ as $\mu\to M$. Let us now find the second and third derivatives of the potential $\Omega_C(\mu)$. Using Eqs. (32) and (33), we obtain easily from Eq. (31)

$$\frac{d^2\Omega_C(\mu)}{d\mu^2}=-\frac{N}{\pi^2}\sqrt{\mu^2-\Sigma_4^2}(\mu-\Sigma_4\Sigma_4'), \tag{34}$$

$$\frac{d^3\Omega_C(\mu)}{d\mu^3}=-\frac{N}{\pi^2}\frac{(\mu-\Sigma_4\Sigma_4')^2}{\sqrt{\mu^2-\Sigma_4^2}}-\frac{N}{\pi^2}\sqrt{\mu^2-\Sigma_4^2}\times(1-(\Sigma_4')^2-\Sigma_4\Sigma_4''). \tag{35}$$

The expression (34) vanishes as $\mu\to M$, but the third derivative of the thermodynamic potential becomes infinite as $\mu\to M$, as one can easily see from Eqs. (35) and (33).

Therefore at the points $\mu=M$ (for $M<M_{2c}$) the third derivative of the thermodynamic potential $\Omega(\mu)$ changes abruptly at the transition from the region B into C and vice versa. This fact strictly indicates that the line $\mu=M$ is a critical curve of second-order phase transitions, i.e., the regions B and C in Fig. 4 correspond to different massive phases of the Nambu–Jona-Lasinio model. The main physical characteristic distinguishing the phases B and C is the particle number density

$$n=-\partial\Omega(\mu)/\partial\mu. \tag{36}$$

From Eqs. (36) and (28) we easily obtain $n_B=0$ in the phase B. In the phase C (see Eq. (31)) we have

$$n_C=\frac{N}{3\pi^2}(\mu^2-\Sigma_4^2(\mu,M))^{3/2}\neq 0. \tag{37}$$

In Fig. 4 the solid and dashed lines represent the critical curves of the first- and second-order phase transitions, respectively. Furthermore, it follows from what has been said above that in this figure, in addition to the point α, the point β is also a tricritical point of the exact model, since both first- and second-order phase transitions can occur in an arbitrarily small neighborhood of this point.

6. CONCLUSIONS

In the present paper we have investigated to leading order in the $1/N$ expansion the phase structure of the Nambu–Jona-Lasinio model with nonzero values of the chemical potential. The main results are presented in Fig. 4, which displays the phase diagram, studied in detail above, of the model.

In the past it was believed that the massive-fermion state of the Nambu–Jona-Lasinio theory is the only phase with spontaneously broken chiral symmetry. Our analysis here has demonstrated the possibility that a more complicated situation arises in this model. It may be possible to compare such situations with specific physical realizations and to relate them to observable effects (for example, high-T_c superconductivity) which are described on the basis of this model. Here, however, skepticism and caution must be excercised, since this model does not include many effects which are characteristic for this phenomenon. It was previously assumed that the transition from the massive to the massless phase is a second-order transition;[14] this immediately led to an equation for the critical curve $f_\mu(0)=0$, i.e., to the expression $\mu=\mu_{1c}(M)$ (21) for all values of the mass M. We showed that this curve $\{l_1:\mu=\mu_{1c}(M)\}$ is an interphase boundary only for sufficiently small $M<M_{1c}$ (see Eq. (25)). For $M>M_{1c}$ chiral symmetry is restored via first-order phase transitions on the curves $\mu_{2c}(M)$ and $\mu_{3c}(M)$, which are different from $\mu_{1c}(M)$.

We also showed that two massive phases B and C (see Fig. 4), the transition between which is a first-order phase transition, exist in the present model. For comparison, we note that in contrast to the four-dimensional case the phase diagrams of four-fermion models in two- $(D=2)$ and three-dimensional $(D=3)$ spaces do not contain the phase C at all. (It is possible that an effect due to the dimensionality of the space is manifested in the present model, and in view of the attempts made by some theoreticians to associate high-T_c superconductivity with the bulk properties of materials, this possibility is not excluded.) We note that in the Gross–Neveu model with $D=2$ the A and B phases are separated by a critical curve $\mu=M/\sqrt{2}$ of first-order phase transitions[19] and for $D=3$ the curve of first-order transitions $\mu=M$ is also the boundary separating the A and B phases.[20]

Finally, it is important to note that the existence of the tricritical points α and β (Fig. 4) in the phase diagram (μ,M) of the Nambu–Jona-Lasinio model is an absolutely new result with no analogs in the literature.

Considering the importance of this model not only for elementary particle physics but for wide applications of the model in solid-state phyics, our analysis could be of interest for a wide circle of physicists.

We thank D. Ebert, R. N. Faustov, and N. O. Agasyan for helpful discussions and A. K. Klimenko and V. A. Vshivtsev for performing numerical calculations and preparing the manuscript for publication.

This work was supported in part by the Russian Fund for Fundamental Research under Grant No. 95-02-037004-a.

7. APPENDIX

We present here some properties of the function $f_\mu(\Sigma)$ defined in Eq. (17). It is easily seen that

$$f_\mu(0)=2\mu^2-M^2\ln(1+\Lambda^2/M^2), \tag{A1}$$

$$f_\mu(\mu)=-M^2\ln(1+\Lambda^2/M^2)+\mu^2\ln(1+\Lambda^2/\mu^2), \tag{A2}$$

$$\frac{d}{d\mu}[f_\mu(\Sigma)|_{\Sigma=0}]>0,\quad \frac{d}{d\mu}[f_\mu(\Sigma)|_{\Sigma=\mu}]>0. \tag{A3}$$

From Eq. (A3) it is obvious that $f_\mu(\mu)=0$ holds only for $\mu=M$ (curve l_2 in Fig. 1); $f_\mu(\mu)<0$ holds for $\mu<M$ and $f_\mu(\mu)>0$ for $\mu>M$. It follows from Eq. (A1) that $f_\mu(0)=0$ holds on the curve l_1 (see Fig. 1) where the parameters μ and M are related by the relation (18):

$$\mu=\mu_{1c}(M)=\sqrt{\frac{1}{2}M^2\ln(1+\Lambda^2/M^2)}. \tag{A4}$$

Since $f_\mu(0)$ is a monotonically increasing function of the parameter μ (see Eq. (A3)), we have $f_\mu(0)<0$ for points (μ,M) lying below the curve l_1 and $f_\mu(0)>0$ for $\mu>\mu_{1c}$.

We define a function $\varphi_\mu(\Sigma)$ as follows:

$$\frac{df_\mu(\Sigma)}{d\Sigma}\equiv 2\Sigma\varphi_\mu(\Sigma)=2\Sigma\left[\ln\frac{\Sigma^2+\Lambda^2}{(\mu+\sqrt{\mu^2-\Sigma^2})^2}-\frac{\Lambda^2}{\Sigma^2+\Lambda^2}\right]. \tag{A5}$$

Hence it is easily seen that $\varphi_\mu(0)$ is a monotonically decreasing function of the parameter μ and vanishes only for $\mu=\Lambda/2\sqrt{e}$ (curve l_3 in Fig. 1). It is also obvious from Eq. (A5) that $\varphi_\mu(\mu)>0$ holds for all values of the chemical potential. If $\mu<\Lambda/2\sqrt{e}$ holds, then we have $\varphi_\mu(0)>0$ and $\varphi_\mu(\Sigma)>0$ for all $\Sigma\in(0,\mu)$. Therefore, in this range of the parameter μ the function $f_\mu(\Sigma)$ is a monotonically increasing function of its argument Σ. (The curves *1* in Figs. 2 and 3 correspond to this case.) If the point (μ,M) lies above the curve l_3, then $\varphi_\mu(0)<0$. However, it is obvious from Eq. (A5) that in this case there exists a point $\Sigma_0\in(0,\mu)$ with $\varphi_\mu(\Sigma_0)=0$ and $f_\mu(\Sigma)$ possesses an absolute minimum (see curves *2* and *3* in Fig. 3).

[1] e-mail: alexandr@vvas.msk.ru
[2] e-mail: zhukovsk@th180.phys.msu.su
[3] e-mail: kklim@mx.ihep.su

---

[1] Y. Nambu and G. Jona-Lasinio, Phys. Rev. **122**, 345 () **124**, 2461 (1961).

[2] V. G. Vaks and A. I. Larkin, Zh. Eksp. Teor. Fiz. **40**, 282; 1392 (1961) [Sov. Phys. JETP **13**, 192; 979 (1961)]; B. A. Arbuzov, A. N. Tavkhelidze, and R. N. Faustov, Dokl. Akad. Nauk SSSR **139**, 345 (1961) [Sov. Phys. Dokl. **6**, 598 (1962)].

[3] E. M. Lifshitz and L. P. Pitaevskiĭ, *Statistical Physics*, Part 2, 3rd ed., Pergamon, Oxford, 1980.

[4] G. Semenoff and L. Wijewardhana, Phys. Rev. Lett. **63**, 2633 (1989); N. Dorey and N. Mavromatos, Phys. Lett. B **250**, 107 (1990); A. Kovner and R. Rosenstein, Phys. Rev. B **42**, 4748 (1990); M. Carena, T. E. Clarkaud, and C. E. M. Wagner, Nucl. Phys. B **356**, 117 (1991).

[5] A. A. Abrikosov, L. P. Gor'kov, and I. E. Dzyaloshinskiĭ, *Methods of Quantum Field Theory in Statistical Physics*, Prentice-Hall, Englewood Cliffs, N. J., 1963 [Russian original, GIFML, Moscow, 1962].

[6] Yu. A. Izyumov, Usp. Fiz. Nauk **161**, 1 (1991) [Sov. Phys. Usp. **34**, 935 (1991)].

[7] V. L. Ginzburg and K. G. Maksimov, SFKhT **5**, 1543 (1992).

[8] A. S. Aleksandrov and A. B. Krebs, Usp. Fiz. Nauk **162**, 1 (1992) [Sov. Phys. Usp. **35**, 345 (1992)].

[9] I. M. Lifshitz, *Selected Works. Electronic Theory of Metals. Physics of Polymers and Bipolymers* [in Russian], Nauka, Moscow, 1994.

[10] M. K. Volkov, Elem. Chast. Atom. Yad. **17**, 433 (1986) [Sov. J. Part. Nucl. **17**, 186 (1986); **24**, 81 (1993) [Phys. Part. Nuclei **24**, 35 (1993)].

[11] S. Weinberg, Phys. Rev. D **13**, 974 (1976).

[12] K. G. Klimenko, Teor. Mat. Fiz. **89**, 211 (1991); Z. Phys. C **54**, 323 (1992); A. S. Vshivtsev, K. G. Klimenko, and B. V. Magnitskiĭ, Teor. Mat. Fiz. **101**, 391 (1994); Yad. Fiz. **57**, 2260 (1994); [Phys. At. Nucl. **57**, 2171 (1994)]; JETP Lett. **62**, 283 (1995); Nuovo Cimento A **107**, 439 (1994).

[13] V. P. Gusynin, V. A. Miransky, and I. A. Shovkovy, Phys. Rev. Lett. **73**, 3499 (1994); R. R. Parwani, Phys. Lett. B **358**, 101 (1995).

[14] S. Kawati and H. Miyata, Phys. Rev. D **23**, 3010 (1981); J. Fuch, Z. Phys. C **22**, 83 (1984).

[15] V. Bernard, U. G. Meissner, and I. Zahel, Phys. Rev. D **36**, 829 (1987); Ch. V. Christov and K. Goeke, Acta Phys. Pol. B **22**, 187 (1991); D. Ebert, Yu. I. Kalinovsky, L. Munchow, and M. K. Volkov, Int. J. Mod. Phys. A **8**, 1295 (1993); T. Inagaki, T. Kuonto, and T. Muta, Int. J. Mod. Phys. A **10**, 2241 (1995).

[16] D. Ebret and M. K. Volkov, Yad. Fiz. **36**, 1265 (1982) [Sov. J. Nucl. Phys. **36**, 736 (1982); Z. Phys. C **16**, 205 (1983); D. Ebert and H. Reinhardt, Nucl. Phys. B **271**, 188 (1986); D. Ebert and M. K. Volkov, Phys. Lett. B **272**, 86 (1991); S. Klevansky and R. H. Lemmert, Phys. Rev. D **39**, 3478 (1991); M. Faber, A. N. Ivanov, M. Nagy, and N. I. Troitskaya, Mod. Phys. Lett. A **8**, 335 (1993).

[17] T. Inagaki, T. Muta, and S. D. Odinstov, Mod. Phys. Lett. A **8**, 2117 (1993); E. Elizalde, S. Lieseduarte, and S. D. Odintsov, Phys. Rev. D **49**, 5551 (1994); Phys. Lett. B **347**, 33 (1995); D. K. Kim and I. G. Koh, Phys. Rev. D **51**, 4573 (1995).

[18] I. P. Bazarov, É. V. Gevorkyan, and P. P. Nikolaev, *Thermodynamics and Statistical Physics* [in Russian], Moscow State University Press, Moscow, 1986.

[19] V. A. Osipov and V. K. Fedyanin, Teor. Mat. Fiz. **73**, 393 (1987); K. G. Klimenko, Teor. Mat. Fiz. **75**, 226 (1988).

[20] K. G. Klimenko, Z. Phys. C **37**, 457 (1988).

Translated by M. E. Alferieff

Nature of the Darwin term and $(Z\alpha)^4 m^3/M^2$ contribution to the Lamb shift for an arbitrary spin of the nucleus

I. B. Khriplovich[1)] and A. I. Milstein[2)]

Budker Institute of Nuclear Physics, 630090, Novosibirsk, Russia

R. A. Sen'kov

Novosibirk University, 630090, Novosibirsk, Russia

(Submitted 3 September 1996)
Zh. Éksp. Teor. Fiz. **111**, 1935–1941 (June 1997)

The contact Darwin term is demonstrated to be of the same origin as the spin-orbit interaction. The $(Z\alpha)^4 m^3/M^2$ correction for the Lamb shift, generated by the Darwin term is found for an arbitrary nonvanishing spin of the nucleus, both half-integer and integer. There is also a contribution of the same nature to the nuclear quadrupole moment. © *1997 American Institute of Physics.* [S1063-7761(97)00206-0]

1. The literature, including that of the pedagogical nature, abounds with assertions on the nature of the Darwin correction which are, in our view, at least doubtful. In particular, we cannot agree with the conclusion[1] that the Darwin term is absent for a particle with spin 1 (see also Ref. 2). This subject is of real interest now for the interpretation of high precision experiments in atomic spectroscopy.[3–5]

To study the problem we consider in this note the Born amplitude for scattering of a particle with an arbitrary spin in an external electromagnetic field. In the case of practical interest, that of an atom, this is the interaction of the nucleus with the electromagnetic field of an electron. We thus derive the general form of the Darwin term for an arbitrary nuclear spin and obtain the corresponding order $(Z\alpha)^4 m^3/M^2$ correction to the Lamb shift (here and below m is the electron mass, and Z and M are, respectively, the charge and mass of the nucleus).

2. The wave function of a particle with an arbitrary spin can be written as (see, for example, §31 of Ref. 6)

$$\Psi=\begin{pmatrix}\xi\\ \eta\end{pmatrix}. \tag{1}$$

The spinors

$$\xi=\{\xi^{\alpha_1\alpha_2\ldots\alpha_p}_{\dot\beta_1\dot\beta_2\ldots\dot\beta_q}\}$$

and

$$\eta=\{\eta^{\dot\beta_1\dot\beta_2\ldots\dot\beta_q}_{\alpha_1\alpha_2\ldots\alpha_p}\}$$

are symmetric in the dotted and undotted indices separately, and

$$p+q=2I,$$

where I is the particle spin. For a particle of half-integer spin one can choose

$$p=I+\frac{1}{2},\quad q=I-\frac{1}{2}.$$

In the case of integer spin it is convenient to use

$$p=q=I.$$

The spinors ξ and η are chosen in such a way that under reflection they go over into each other (up to a phase). At $p\neq q$ they are different objects which belong to different representations of the Lorentz group. If $p=q$, these two spinors coincide. Nevertheless, we will use the same expression (1) for the wave function of any spin; i.e., we will also introduce formally the object η for an integer spin, keeping in mind that it is expressed in terms of ξ. This will allow us to perform calculations in the same way for the integer and half-integer spins.

In the rest frame both ξ and η coincide with a nonrelativistic spinor ξ_0, which is symmetric in all indices; in the rest frame there is no difference between dotted and undotted indices. The Lorentz transformation of ξ_0, up to the terms $\sim(v/c)^2$ included, is

$$\begin{aligned}\xi&=\left(1+\frac{\mathbf{\Sigma v}}{2}+\frac{(\mathbf{\Sigma v})^2}{8}\right)\xi_0,\\ \eta&=\left(1-\frac{\mathbf{\Sigma v}}{2}+\frac{(\mathbf{\Sigma v})^2}{8}\right)\xi_0.\end{aligned} \tag{2}$$

Here

$$\mathbf{\Sigma}=\sum_{i=1}^{p}\boldsymbol{\sigma}_i-\sum_{i=p+1}^{p+q}\boldsymbol{\sigma}_i.$$

and $\boldsymbol{\sigma}_i$ acts on the ith index of the spinor ξ_0 as follows:

$$\boldsymbol{\sigma}_i\xi_0=(\boldsymbol{\sigma}_i)_{\alpha_i\beta_i}(\xi_0)_{\ldots\beta_i\ldots}. \tag{3}$$

In the Lorentz transformation (2) for ξ, after the action of the operator $\mathbf{\Sigma}$ on ξ_0 the first p indices are identified with the upper undotted indices and the next q indices we identified with the lower dotted indices. The inverse situation takes place for η.

We will use, however, Eq. (2) since it no longer distinguishes between the upper and lower or the dotted and undotted spinor indices. It allows us to introduce in a natural way the «standard» representation for the spinors, in close analogy with that for spin 1/2:

$$\phi=(\xi+\eta)/2,\quad \chi=(\xi-\eta)/2.$$

In it the wave function is written as

$$\Psi=\begin{pmatrix}[1+(\boldsymbol{\Sigma}\mathbf{v})^2/8] & \xi_0\\ (\boldsymbol{\Sigma}\mathbf{v}/2) & \xi_0\end{pmatrix}. \tag{4}$$

It is convenient to introduce the object

$$\overline{\Psi}=(\phi^*,-\chi^*).$$

Then

$$\overline{\Psi}\Psi=\phi^*\phi-\chi^*\chi=\xi_0^*\xi_0$$

is an invariant. We will use the common noncovariant normalization of the particle number density

$$\rho=\frac{E}{M}\,\overline{\psi}\psi=1, \tag{5}$$

where the wave function ψ is

$$\psi=\sqrt{\frac{M}{E}}\begin{pmatrix}[1+(\boldsymbol{\Sigma}\mathbf{v})^2/8] & \xi_0\\ (\boldsymbol{\Sigma}\mathbf{v}/2) & \xi_0\end{pmatrix}. \tag{6}$$

Here E is the particle energy.

3. Let us go over now to the scattering amplitude itself. The order $1/M^2$ terms in it arise only in the time component of the electromagnetic current. Restricting the analysis to the form factors of the lowest multipolarity, electric F_e and magnetic G_m, we can write this component for an arbitrary spin as follows:

$$j_0=F_e\,\frac{E+E'}{2M}\,\overline{\psi'}\psi+\frac{G_m}{2M}\,\psi'^*\boldsymbol{\Gamma}\mathbf{q}\psi, \tag{7}$$

where

$$\mathbf{q}=\mathbf{p}'-\mathbf{p}.$$

The matrix

$$\boldsymbol{\Gamma}=\begin{pmatrix}0 & \boldsymbol{\Sigma}\\ -\boldsymbol{\Sigma} & 0\end{pmatrix} \tag{8}$$

is a natural generalization of the corresponding expression for spin 1/2 (which is valid in the spinor and standard representations):

$$\gamma=\begin{pmatrix}0 & \boldsymbol{\sigma}\\ -\boldsymbol{\sigma} & 0\end{pmatrix}. \tag{9}$$

This generalization is fairly obvious in the spinor representation. According to (9), here $\boldsymbol{\sigma}$ connects a dotted index in the initial spinor ψ with the undotted one in $\overline{\psi}$, and $-\boldsymbol{\sigma}$ connects an undotted index from ψ with the dotted one in $\overline{\psi}$. This is exactly what $\boldsymbol{\Gamma}$ does. It is straightforward now to prove expression (8) for the standard representation. Let us mention also that Eq. (8) is confirmed by the final result, which reproduces correctly the spin-orbit interaction; the form of the latter is well known for an arbitrary spin (see, e.g., §41 of Ref. 6).

The term with G_m in the current density is

$$j_{0m}=\frac{G_m}{2M}\,\xi_0'^*(1,\boldsymbol{\Sigma}\mathbf{v}'/2)\begin{pmatrix}0 & \boldsymbol{\Sigma}\mathbf{q}\\ -\boldsymbol{\Sigma}\mathbf{q} & 0\end{pmatrix}\begin{pmatrix}1\\ \boldsymbol{\Sigma}\mathbf{v}/2\end{pmatrix}\xi_0$$

$$=\frac{G_m}{4M^2}\,\xi_0'^*(-\boldsymbol{\Sigma}\mathbf{q})^2+4\overline{i\mathbf{I}}[\mathbf{qp}])\xi_0. \tag{10}$$

The spin operator is

$$\mathbf{I}=\frac{1}{2}\sum_{i=1}^{2I}\boldsymbol{\sigma}_i.$$

The first term with F_e in Eq. (7) reduces to an analogous structure:

$$j_{0\,ch}=F_e\,\frac{E'+E}{2\sqrt{EE'}}\,\xi_0'^*\left(1+\frac{(\boldsymbol{\Sigma}\mathbf{v})^2}{8}+\frac{(\boldsymbol{\Sigma}\mathbf{v}')^2}{8}\right.$$

$$\left.-\frac{(\boldsymbol{\Sigma}\mathbf{v}')(\boldsymbol{\Sigma}\mathbf{v})}{4}\right)\xi_0=F_e\xi_0'^*\left(1+\frac{(\boldsymbol{\Sigma}\mathbf{q})^2}{8M^2}-i\,\frac{\mathbf{I}[\mathbf{qp}]}{2M^2}\right)\xi_0. \tag{11}$$

Thus the total charge density is

$$j_0=\xi_0'^*\left(F_e-(2G_m-F_e)\,\frac{(\boldsymbol{\Sigma}\mathbf{q})^2}{8M^2}+(2G_m\right.$$

$$\left.-F_e)i\,\frac{\mathbf{I}[\mathbf{qp}]}{2M^2}\right)\xi_0.$$

We disregard for now the charge radius of the nucleus, so that

$$F_e=F_e(0)=1.$$

The dependence of the spin-orbit interaction on the gyromagnetic ratio g is universal for any spin; this ratio enters through the factor $g-1$. Therefore, the magnetic form factor is normalized as follows:

$$G_m(0)=\frac{g}{2}.$$

Let us split now $(\boldsymbol{\Sigma}\mathbf{q})^2$ into the contact and quadrupole parts:

$$\Sigma_i\Sigma_j q_i q_j=\frac{\mathbf{q}^2}{3}\,\Sigma_i\Sigma_i+\left(q_iq_j-\frac{1}{3}\,\mathbf{q}^2\delta_{ij}\right)\Sigma_i\Sigma_j. \tag{12}$$

The contact term in (12) is

$$\boldsymbol{\Sigma}\boldsymbol{\Sigma}=\left(\sum_{i=1}^{p}\boldsymbol{\sigma}_i\right)^2-2\left(\sum_{i=1}^{p}\boldsymbol{\sigma}_i\right)\left(\sum_{i=p+1}^{p+q}\boldsymbol{\sigma}_i\right)+\left(\sum_{i=p+1}^{p+q}\boldsymbol{\sigma}_i\right)^2$$

$$=3(p+q)+2\left(\frac{p(p-1)}{2}+\frac{q(q-1)}{2}-pq\right)=4I(1+\zeta), \tag{13}$$

$$\zeta=\begin{cases}0, & \text{integer spin},\\ 1/4I, & \text{half-integer spin}.\end{cases}$$

In deriving Eq. (13) we used the symmetry in any pair of spinor indices, α_1, α_2 [see Eq. (3)]. This symmetry means that the corresponding spins 1 and 2 add up to the total spin $S=1$. Therefore,

$$(\boldsymbol{\sigma}_1\boldsymbol{\sigma}_2)\xi_0=\xi_0.$$

The interaction operator is proportional to the Fourier transform of the Born amplitude (see, e.g., §83 of Ref. 6). We thus obtain from (13) the following contact interaction between a nucleus of charge Z and the electron:

$$U(\mathbf{r})=\frac{2\pi}{3}\,\frac{Z\alpha}{M^2}\,(g-1)I(1+\zeta)\,\delta(\mathbf{r}). \tag{14}$$

The corresponding energy correction is

$$\Delta E_n=\frac{2}{3}\,\frac{m^3}{M^2}\,(g-1)I(1+\zeta)\,\frac{(Z\alpha)^4}{n^3}\,\delta_{0l}. \tag{15}$$

For the hydrogen atom ($I=1/2$) this correction was obtained long ago in Ref. 7.

Let us consider now the quadrupole part of (12). Using again the complete symmetry of ξ_0, one can easily calculate the corresponding quadrupole interaction:

$$U_2(\mathbf{r})=-\frac{1}{6}\,\nabla_i\nabla_j\,\frac{e}{r}\,\delta Q_{ij}. \tag{16}$$

Here

$$\delta Q_{ij}=-\frac{3}{4}\,\frac{Ze(g-1)}{M^2}\,\Lambda\left\{I_iI_j+I_jI_i-\frac{2}{3}\,\delta_{ij}I(I+1)\right\}, \tag{17}$$

$$\Lambda=\begin{cases}1/(2I-1), & \text{integer spin,}\\ 1/2I, & \text{half-integer spin.}\end{cases}$$

Expression (17) is a correction to the nuclear quadrupole moment. Its existence for $I=1$ was pointed out in Ref. 1.

This correction to the quadrupole moment can be estimated as

$$\delta Q\approx-0.22(g-1)\,\frac{ZI}{A^2}\,e\ \text{mbarn}.$$

For the deuteron ($Z=1$, $A=2$, $g=2\mu_d=1.714$. $Q=2.86\ e$ mbarn) it is $-0.04\ e$ mbarn.

4. Let us return now to the discussion of the contact term. There is some ambiguity in its definition related to the nuclear charge radius. The contribution of the latter produces a contact interaction and enters physical observables in a sum with the expression $(g-1)\mathbf{q}^2I(1+\zeta)/(6M^2)$. In particular, the elastic cross section of the electron-nucleus scattering at small $\mathbf{q}^2$, up to the terms $\mathbf{q}^2/M^2$ included, is

$$\frac{d\sigma}{d\Omega}=\frac{\alpha^2}{4\epsilon^2}\,\frac{\cos^2(\theta/2)}{\sin^4(\theta/2)}\,\frac{1}{1+2\sin^2(\theta/2)\epsilon/M}\times\left\{\left[1-\frac{1}{6}\,\langle r^2\rangle_F\mathbf{q}^2-(g-1)\,\frac{\mathbf{q}^2}{6M^2}\,I(1+\zeta)\right]^2+\frac{4}{3}\,G_m^2I(I+1)[2\tan^2(\theta/2)+1]\right\}, \tag{18}$$

where $\langle r^2\rangle_F$ is defined in terms of the expansion of the form factor F_e

$$F_e(q^2)\approx 1-\frac{1}{6}\,\langle r^2\rangle_F\mathbf{q}^2. \tag{19}$$

We note here that the expression in square brackets in Eq. (18) reduces for the proton ($I=1/2$) to

$$1-\frac{1}{6}\,\langle r^2\rangle_F\mathbf{q}^2-(g-1)\,\frac{\mathbf{q}^2}{8M^2}, \tag{20}$$

and for the deuteron ($I=1$) to

$$1-\frac{1}{6}\,\langle r^2\rangle_F\mathbf{q}^2-(g-1)\,\frac{\mathbf{q}^2}{6M^2}. \tag{21}$$

However, the proton charge radius is usually defined differently from that in Eq. (20), i.e., in terms of the expansion of the so-called Sachs form factor

$$G_e=F_e-\frac{\mathbf{q}^2}{4M^2}\,G_m.$$

Obviously, the charge radius defined in terms of the form factor G_e is

$$-\frac{1}{6}\,\langle r^2\rangle_G=\frac{\partial G_e}{\partial\mathbf{q}^2}=-\frac{1}{6}\,\langle r^2\rangle_F-\frac{g}{8M^2}.$$

Correspondingly, expression (20) is usually rewritten as

$$1-\frac{1}{6}\,\langle r^2\rangle_G\mathbf{q}^2+\frac{\mathbf{q}^2}{8M^2}, \tag{22}$$

and the Darwin correction for the proton is defined as

$$\frac{\mathbf{q}^2}{8M^2},$$

but not as

$$-\frac{(g-1)\mathbf{q}^2}{8M^2}.$$

We could redefine the electric form factor for the deuteron from F_e to G_e in such a way that here

$$-\frac{1}{6}\,\langle r^2\rangle_G=\frac{\partial G_e}{\partial\mathbf{q}^2}=-\frac{1}{6}\,\langle r^2\rangle_F-\frac{g}{6M^2},$$

so that the Darwin correction for the deuteron becomes

$$\frac{\mathbf{q}^2}{6M^2},$$

instead of

$$-\frac{(g-1)\mathbf{q}^2}{6M^2}.$$

However, for a deuteron the common definition of the charge radius is neither F_e nor G_e but

$$-\frac{1}{6}\,\langle r^2\rangle_D=-\frac{1}{6}\,\langle r^2\rangle_F-\frac{g-1}{6M^2}.$$

Of course, under this definition the whole Darwin term is swallowed up by $\langle r^2\rangle_D$. No wonder that the authors of Ref. 1, using $\langle r^2\rangle_D$ instead of $\langle r^2\rangle_E$ or $\langle r^2\rangle_G$, have concluded that for the deuteron, in contrast with the proton, the Darwin correction is absent.

Clearly, this distinction between the deuteron and the proton is based only on a rather arbitrary definition of the charge radius of the former; this distinction has no physical meaning and has nothing to do with the nature of the Darwin term.

5. In summary, the Darwin interaction exists for any nonvanishing spin and is of the same nature as the spin-orbit interaction. In particular, like the spin-orbit interaction, the Darwin term is not directly related to the so-called zitterbewegung. Of course, there is a certain difference between the spin-orbit and contact energy corrections. The former has a classical limit together with $\langle 1/r^3 \rangle$, while the latter, which is proportional to $|\psi(0)|^2$, does not. However, this fact has nothing to do with relativity and negative energies, and therefore is certainly unrelated to the zitterbewegung.

(1) E-mail address: khriplovich@inp.nsk.su
(2) E-mail address: milstein@inp.nsk.su

---

[1] K. Pachucki and S. G. Karshenboim, J. Phys. B **28**, L211 (1995).

[2] K. Pachucki, D. Leibfried, M. Weitz, A. Huber, W. König, and T. W. Hänsch, J. Phys. B **29**, 177 (1996).

[3] F. Schmidt-Kaler, D. Leibfried, M. Weitz, and T. W. Hänsch Phys. Rev. Lett. **70**, 2261 (1993).

[4] D. Shiner, R. Dixson, and V. Vedantham, to be published.

[5] M. Weitz *et al.*, Phys. Rev. A **52**, 2664 (1995).

[6] V. B. Berestetskii, E. M. Lifshitz, and L. P. Pitaevskii, *Quantum Electrodynamics*, Pergamon Press, Oxford (1982).

[7] W. A. Barker and F. N. Glover, Phys. Rev. **99**, 317 (1955).

Published in English in the original Russian journal. Reproduced here with stylistic changes by the Translation Editor.

Generation of phase-correlated light pulses in a hyper-Raman active medium

Yu. P. Malakyan and A. R. Mkhitaryan

Institute of Physics Research, Armenian National Academy of Sciences, 378410 Ashtarak-2, Republic of Armenia

(Submitted 13 August 1996)
Zh. Éksp. Teor. Fiz. **111**, 1942–1954 (June 1997)

We analysis within the quantum setting the effect of correlation of monochromatic fields of the Stokes and parametric radiation generated in a two-photon-absorbing medium by hyper-Raman scattering and four-wave mixing, respectively. Our results show that when there is destructive interference between the two processes, both modes are amplified in the medium in a correlated way, so that beyond a certain characteristic distance they propagate in a stationary manner with the same amplitudes and photon statistics but with opposite phases, while the medium becomes transparent to them. We calculate the intensity of the fields, the emission linewidths, and the diffusion of the total phase, the latter determining the extent of phase noise correlation at the output. We also show that no matter how the phases of the separate modes diffuse, the quantum fluctuations in the total phase may be completely squeezed.

[S1063-7761(97)00306-5]

1. INTRODUCTION

Generation of phase-correlated light beams in which quantum noise is quenched below the coherence threshold still remains an object of intensive studies,[1–11] with the aim of building high-precision laser interferometers. To raise the sensitivity of these devices, usually limited by vacuum fluctuations in passive interferometers and spontaneous-emission fluctuations in active interferometers, using squeezed states of light[1,2] and correlated-spontaneous-emission lasers[3–5] was proposed. Generation of double-mode squeezed light by four-wave mixing, whose theoretical aspects were studied in Refs. 6 and 7, was achieved by several researchers[8–10] with squeezing of up to 20%, and the quenching of spontaneous-emission fluctuations in the relative phase of a double-mode correlated-emission laser below the Schawlow–Townes limit was observed by Winters *et al.*[11] However, there are certain difficulties that limit the application of these schemes. As shown by Bondurant *et al.*,[6] mode losses impose a severe restriction on the extent of squeezing that can be obtained in four-wave mixing. More than that, the losses must be low not only in absolute value but also in comparison with the constant of the nonlinear interaction between the modes, which cannot always be achieved since strong nonlinearities in a resonant medium are accompanied by large losses (see the first paper in Ref. 7). The decisive factor in quenching spontaneous-emission fluctuations in correlated-emission lasers is the coherence of the atomic states, and before atoms are to be injected into the laser cavity they must be prepared in a coherent superposition of these states,[5] which requires employing lasing methods that do not involve population inversion.

In this paper we propose a new mechanism for generation of phase-correlated light. The method is based on destructive interference between hyper-Raman scattering and the accompanying four-wave mixing. The reader will recall that hyper-Raman scattering in a medium with two-photon absorption generates Stokes radiation that is in quasiresonance with the atomic $2\rightarrow 3$ transition (Fig. 1). This radiation and the pump field induce, via four-wave mixing, a parametric field on the frequency of the $3\rightarrow 1$ transition. These processes have been thoroughly studied both theoretically[12,13] and experimentally,[14–17] and the destructive interference effect has been predicted in our earlier work[12,18] and recently corroborated in experiments.[19,20] The physical meaning of this effect is discussed below.

We present the quantum theory of these processes and show that in destructive interference the quantum fluctuations in the total phase of two modes can be squeezed practically up to 100% and the fields can be amplified, even though, in contrast to the previous schemes of four-wave mixing, the parametric coupling constant of the fields is smaller than the linear losses in the modes. Actually we are speaking about generating bright squeezed light in experiments that can easily be realized. We also find that at large distances a stationary propagation regime sets in, which guarantees that the amplitudes of the fields are stabilized and the medium becomes transparent. An important feature of this scheme is that the frequencies of both modes can vary over a broad interval. For instance, in the case of hyper-Raman scattering on the $5S-5P_{3/2}$ transition through the intermediate state $5D_{5/2}$ in rubidium vapor the two frequencies are approximately the same, while in heavy-metal vapors the Stokes radiation is generated in the IR range and the parametric field in the vacuum UV range.[16,17]

In our previous paper[21] we used the semiclassical approach to show that the Stokes and parametric fields are correlated in amplitude and phase. Here we study the effect of quantum noise on the magnitude of this correlation. We calculate the spectral emission widths and the diffusion of the total phase, which determines the mode correlation at the exit. To this end we solve the stochastic differential equations for the complex-valued amplitudes of single-mode Stokes and parametric radiation. The starting point of our quantum analysis is the basic equation for the density matrix of the two modes. Section 2 is devoted to deriving this equation. There we also report on the discovery of a coherent

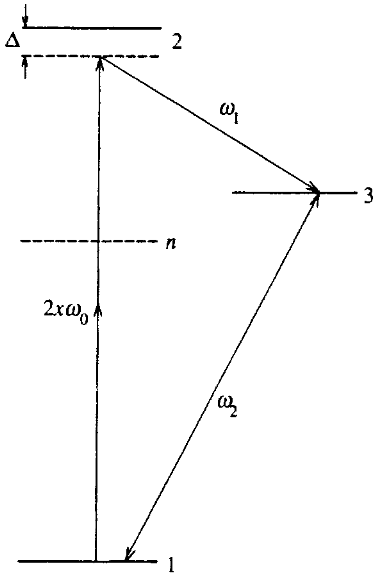


FIG. 1. Energy level diagram for a hyper-Raman active atom; ω_0, ω_1, and ω_2 are the frequencies of the pump, Stokes, and parametric fields, respectively.

field superposition propagating in the medium undisturbed. In Sec. 3 we find the general solutions for the semiclassical amplitudes, the average number of photons, and the standard deviation of the number of photons. These solutions describe the evolution of the correlation between modes as a function of the propagation length. In Sec. 4 we calculate the emission linewidths and the quantum fluctuations in the total phase of the two modes; we also discuss the extent of phase correlation at the exit from the medium. Concluding remarks are presented in Sec. 5.

2. THE BASIC EQUATIONS FOR THE DENSITY MATRIX OF TWO MODES

Let us discuss the interaction between a gaseous medium consisting of identical atoms with a level configuration depicted in Fig. 1 and a classical pump field E at a frequency ω_0 with a two-photon detuning $\Delta=\omega_{21}-2\omega_0$, and quantized single-mode fields of the Stokes and parametric radiation of frequencies $\omega_1=\omega_{21}-\Delta$ and $\omega_2\sim\omega_{31}$, with the quantized fields described by the annihilation (creation) operators a_1 (a_1^+) and a_2 (a_2^+), respectively. The system's interaction Hamiltonian, which can be obtained by adiabatically excluding the population amplitude of level 2, in the resonant approximation has the following form:

$$H=\frac{i\hbar}{\sqrt{V}}\sum_i (Ga_1^+-fa_2)\sigma_{31}^{(i)}+\text{H.c.},\tag{1}$$

where V is the quantization volume (assumed equal to the volume occupied by the medium), $\sigma_{31}^{(i)}$ is the operator of the atomic transition $1\rightarrow3$ for the ith atom, and the coefficients

$$G=\sqrt{\frac{2\pi\omega_2}{\hbar}}\,E^2\mu_{23}\sum_n\frac{\mu_{1n}\mu_{n2}}{\hbar^2\Delta(\omega_{n1}-\omega_0)},$$

$$f=\sqrt{\frac{2\pi\omega_3}{\hbar}}\,\mu_{31}\tag{2}$$

are the effective three-photon matrix element of the $1\rightarrow3$ transition with emission of a Stokes photon, and the parametric-field coupling constant. For simplicity we assume that G and f are real and positive. The two are linked to the gain α of the Stokes field and the absorption coefficient β of the parametric field by the following relationships:[21]

$$\alpha=\frac{2G^2N}{c\Gamma},\quad \beta=\frac{2f^2N}{c\Gamma},\tag{3}$$

where N is the atom number density, and Γ is the transverse relaxation of the polarization at the frequency $\omega_2\cong\omega_{31}$.

Destructive interference between hyper-Raman scattering and four-wave mixing occurs only if[18]

$$\alpha<\beta.\tag{4}$$

This condition can easily be met if one bears in mind that α depends on two free parameters, the intensity and two-photon detuning of the pump radiation, and that β is a constant proportional to the oscillator strength of the $1\rightarrow3$ transition.

The reader will recall that destructive interference occurs only in the forward direction (the direction of propagation of the pump pulse), since in this direction the phase detuning of the wave vectors of the three waves, $\Delta k=2k-k_1-k_2$, is small compared to α. In the backward direction the parametric processes are suppressed and there is only hyper-Raman scattering. We also note that since destructive interference is highly dependent on the relative phases of the three fields and is sensitive to collision and Doppler broadening, the detuning Δ is assumed to be much larger than the widths of these processes.

The equation for the density matrix ρ of the two modes can be derived in the standard way by using the Hamiltonian (1) and assuming that all atoms remain in the ground state 1:

$$\frac{1}{c}\frac{d\rho}{dt}=-\frac{\alpha}{2}\,[a_1a_1^+\rho-2a_1^+\rho a_1+\rho a_1^+a_1]-\frac{\beta}{2}\,[a_2^+a_2\rho-2a_2\rho a_2^++\rho a_2a_2^+]+\frac{\sqrt{\alpha\beta}}{2}\,[(a_1a_2\rho+\rho a_1a_2-2a_2\rho a_1)+\text{H.c.}].\tag{5}$$

Two remarks concerning Eq. (5) are in order. First, direct losses occur only in the ω_2-mode in the form of absorption on the $1\rightarrow3$ transition, which is described by the second term on the right-hand side of Eq. (5). There are no individual losses in the ω_1-mode, since we ignored population motion by assuming that level 3 is not occupied. Nevertheless, the Stokes field can be absorbed due to a parametric interaction with a coupling constant $\sqrt{\alpha\beta}/2$, which according to (4) is smaller than the linear losses in the ω_2-mode. All other losses are ignored. The first term on the right-hand side of Eq. (5) corresponds to the emission of a Stokes photon in the three-photon transition $1\rightarrow3$ of an atom. Second, since we are discussing the stationary interaction with the field, assuming that the pulse length is much larger than all the relaxation times of the medium, we can say that in the single-mode approximation Eq. (5) is equivalent to the propagation equation $z=ct$. Then, introducing the operator

$$M=\frac{1}{\sqrt{\beta-\alpha}}\,[\sqrt{\alpha}a_1^+-\sqrt{\beta}a_2] \tag{6}$$

with the commutation property $[M,M^+]=1$, we can write Eq. (5) as

$$\frac{d\rho}{dz}=-g[M^+M\rho+\rho M^+M-2M\rho M^+],\quad g=\frac{\beta-\alpha}{2}, \tag{7}$$

which can be solved if we introduce the boundary condition $\rho=\rho_1\rho_2$, where ρ_1 and ρ_2 are the initial density matrices of the Stokes and parametric modes (they will be specified later). Equation (7) possesses a remarkable feature: it contains only the linear combination M of the fields, while the second linear combination

$$L=\frac{1}{\sqrt{\beta-\alpha}}\,[-\sqrt{\beta}a_1^++\sqrt{\alpha}a_2] \tag{8}$$

is a constant of motion, since $[M,L]=[M^+,L]=0$. This leads to a situation in which from the very beginning the fields are amplified and propagate in the medium in a correlated manner so that the statistical properties of the composite mode L remain unchanged. The correlation becomes strongest in the region $z>g^{-1}$, where both fields take on their stationary values and the medium becomes transparent. The solution of Eq. (7) in this limit can be found from the fact that $M\rho=\rho M^+=0$, i.e., $\rho=|\phi\rangle\langle\phi|$, with $|\phi\rangle$ satisfying the following equation:[22]

$$(\sqrt{\alpha}a_1^+-\sqrt{\beta}a_2)|\phi\rangle=0. \tag{9}$$

This equation makes it possible to find all the asymptotic relationships between the amplitudes, the average number of photons, and the statistics of the two modes without solving Eq. (7). In particular, Eq. (9) implies that

$$\langle a_1(\infty)\rangle=\sqrt{\frac{\beta}{\alpha}}\,\langle a_2^+(\infty)\rangle,\quad \langle n_1(\infty)\rangle=\frac{\beta}{\alpha}\,\langle n_2(\infty)\rangle-1,$$

$$\langle\Delta n_1^2(\infty)\rangle=\frac{\beta(\beta^2-2g^2)}{\alpha(\alpha\beta-2g^2)}\,\langle\Delta n_2^2(\infty)\rangle, \tag{10}$$

where $\langle\Delta n^2\rangle=\langle n^2\rangle-\langle n\rangle^2$ is the variance of the number of photons. When the field gain is high, we can replace a_i with $E_i/\sqrt{\omega_i}$ $(i=1,2)$ and thus go over to classical amplitudes E_1 and E_2, which, as Eqs. (10) imply, are equal in magnitude (to within the factor $\sqrt{\beta\omega_1/\alpha\omega_2}$) and have opposite phases. Also, when $\alpha\leqslant\beta$ holds, the modes have practically the same statistics. What is more interesting, however, is that if we use Eqs. (10) and the fact that L is spatially invariant,

$$\langle(L^+)^mL^p\rangle|_{z>g^{-1}}=\langle(L^+)^mL^p\rangle|_{z=0}, \tag{11}$$

where m and p are integers, we can establish the correlation that exists between the asymptotic fields and the input pulses. Let us examine the case where the signal field exists only at the frequency ω_1 and the second field is generated in the medium. Then, using (10) and (11), we obtain

$$\langle a_1(\infty)\rangle=\frac{\beta}{2\alpha}\,\langle a_1(0)\rangle,$$

$$\langle n_1(\infty)\rangle=\frac{\beta^2}{4g^2}\,(\langle n_1(0)\rangle+1)-1, \tag{12}$$

and

$$\langle\Delta n_1^2(\infty)\rangle=\frac{\beta^4}{16g^4}\,[\langle\Delta n_1^2(0)\rangle+\langle n_1(0)\rangle]-\frac{\beta^2}{4g^2}\,\langle n_1(0)\rangle. \tag{13}$$

The last result, which links the statistics of the photons in the ω_1-mode at entrance to the medium and in the asymptotical region, is especially interesting. We see that in the limit $\alpha\ll\beta$, the statistics remains practically unchanged, but no field gain is achieved (see Eqs. (12)), and, according to (10), the ω_2-mode is not generated. But if $\alpha\leqslant\beta$ holds, there is considerable gain, while the statistics worsens to a greater extent. There is a certain range of optimum values of α/β in which for an initially coherent ω_1-field the asymptotic modes are coherent, too, but at the same time are considerably amplified. In Sec. 3 we discuss these aspects in greater detail.

In concluding this section we note that invariant field combinations of type L have also been discovered in the generation of correlated pulses in three-level Λ-systems in conditions of electromagnetically induced transparency.[23–25] Their simple properties provide the means for a compact description of the correlations between fields without changing the dynamics of the processes.

3. CORRELATED AMPLIFICATION OF FIELDS: THE MAGNITUDE OF CORRELATIONS OF THE NUMBER OF PHOTONS

In this section we discuss the laws of correlated field amplification and the strengthening of the correlation between the intensities of two modes as a function of z, assuming that initially the modes were in pure states and were uncorrelated. The quantum dynamics described by Eq. (7) leads to a strong mixing, or entanglement, of the fields, and for $z>0$ the density matrix ρ can no more be represented by the direct product of the state vectors of individual modes. We know of no analytical solution of Eq. (7) in general form, so that we find the amplitudes, the average number of photons, and the photon dispersion by solving equations derived from (7). For the average values of the complex-valued amplitudes and the number of photons these equations are

$$\frac{d\langle a_1\rangle}{dz}=\frac{\alpha}{2}\,\langle a_1\rangle-\frac{1}{2}\,\sqrt{\alpha\beta}\langle a_2^+\rangle,$$

$$\frac{d\langle a_2\rangle}{dz}=-\frac{\beta}{2}\,\langle a_2\rangle+\frac{1}{2}\,\sqrt{\alpha\beta}\langle a_1^+\rangle \tag{14}$$

and

$$\frac{d\langle n_1\rangle}{dz}=\alpha\langle n_1\rangle-D+\alpha,\qquad \frac{d\langle n_2\rangle}{dz}=-\beta\langle n_2\rangle+D,$$

$$D=\sqrt{\alpha\beta}\ \mathrm{Re}\ \eta,\qquad \eta=\langle a_1a_2\rangle, \tag{15}$$

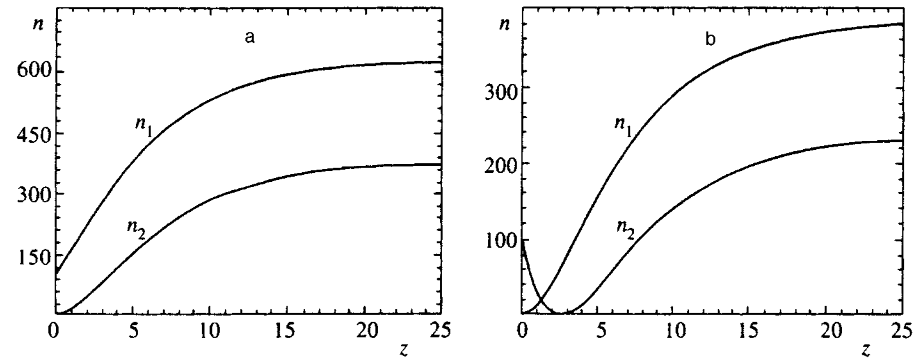


FIG. 2. The number of photons as a function of z (in units of g^{-1}) in the Stokes (n_1) and parametric (n_2) modes at $n_{10}=100$, $n_{20}=0$, and $\alpha=0.6\beta$ (a) and at $n_{10}=0$ and $n_{20}=100$ (b).

$$\frac{d\eta}{dz}=-g\eta+\tfrac{1}{2}\sqrt{\alpha\beta}\;(\langle n_1\rangle-\langle n_2\rangle+1).$$

Equations (14) and (15) have the following solutions:

$$\begin{aligned}\langle a_1(z)\rangle&=\frac{a_{10}}{2g}\;(\beta-\alpha e^{-gz})-\frac{\sqrt{\alpha\beta}}{2g}\;a_{20}^{*}(1-e^{-gz}),\\ \langle a_2(z)\rangle&=\frac{a_{20}}{2g}\;(\beta e^{-gz}-\alpha)+\frac{\sqrt{\alpha\beta}}{2g}\;a_{10}^{*}(1-e^{-gz})\end{aligned}\tag{16}$$

and

$$\begin{aligned}\langle n_1(z)\rangle&=\frac{n_{10}}{4g^2}\;(\beta-\alpha e^{-gz})^2+\frac{\alpha\beta}{4g^2}\;n_{20}(1-e^{-gz})^2+n_{1s},\\ \langle n_2(z)\rangle&=\frac{n_{20}}{4g^2}\;(\beta e^{-gz}-\alpha)^2+\frac{\alpha\beta}{4g^2}\;n_{10}(1-e^{-gz})^2+n_{2s},\end{aligned}\tag{17}$$

$$\eta(z)=\frac{\sqrt{\alpha\beta}}{4g^2}\;[n_{10}(\beta-\alpha e^{-gz})+n_{20}(\alpha-\beta e^{-gz})]\times(1-e^{-gz})+\eta_s,$$

where $n_{i0}=|a_{i0}|^2$, $i=1,2$, a_{i0} are the initial values of the amplitudes, and n_{is} and η_s stand for the spontaneous-photon contributions:

$$\begin{aligned}n_{1s}&=\frac{1}{4g^2}\;(\beta-\alpha e^{-gz})^2-1,\\ n_{2s}&=\frac{\alpha\beta}{4g^2}\;(1-e^{-gz})^2,\\ \eta_s&=\frac{\sqrt{\alpha\beta}}{4g^2}\;(\beta-\alpha e^{-gz})(1-e^{-gz}).\end{aligned}\tag{18}$$

In deriving Eqs. (17) we allowed for the fact that $\eta(0)=0$. Note that η is a real quantity.

Let us explain the effect of destructive interference between hyper-Raman scattering and four-wave mixing using the equation for $\langle n_1\rangle$ through an example (see the first line in Eq. (15)). The first term $\alpha\langle n_1\rangle$ on the right-hand side of the equation describes the increase in the number of Stokes photons caused by hyper-Raman scattering with gain α. The second term D corresponds to the contribution of four-wave mixing and is negative, since the two processes compete. However, the destructive interference between the two processes leads to their suppression only if D is comparable to $\alpha\langle n_1\rangle$ or is larger than $\alpha\langle n_1\rangle$. And this happens only if $\beta\geqslant\alpha$.

The solutions imply that the amplification of the fields strongly depends on the initial conditions. For instance, if only a Stokes signal enters the medium and the parametric field is generated in the medium, i.e., $a_{10}\neq 0$ and $a_{20}=0$, then, as Eqs. (16) imply, both field monotonically increase in z with a constant phase. In the opposite case, where we have $a_{20}\neq 0$ and $a_{10}=0$, the ω_2-field is first completely absorbed in the vicinity of $z_0=g^{-1}\ln(\beta/\alpha)$, and then is regenerated, but with a phase shifted by π in relation to the initial value. The physical meaning of this effect is that the ω_2-mode is absorbed due to the destructive interference (4). Here coherent polarization is created on the $3\rightarrow 1$ transition but, rather than relaxing, due to parametric coupling it emits radiation again into the ω_2-mode with a phase, however, that corresponds to four-wave mixing. The effect is clearly visible in Fig. 2, which illustrates the behavior of the z-dependence of $\langle n_{1,2}(z)\rangle$ for two initial values, with the spontaneous-photon contribution ignored.

A parameter that is more convenient for studying photon statistics than the variance of the number of photons is the second-order correlation function $G_i^{(2)}=[\langle n_i^2\rangle-\langle n_i\rangle]/\langle n_i\rangle^2$, whose values occupies a small interval. The reader will recall that for a coherent field $G^{(2)}=1$. To calculate $G^{(2)}$ we must know the solution for the $\langle n_i^2\rangle$. These in turn are related to new correlators: $\langle n_1 n_2\rangle$, $\langle a_1^{+2}a_2^{+2}\rangle$, etc. The equations for these quantities are quite lengthy and we do not give them here, the more so that analytical solutions for these equations have been found only for the case of initial coherent fields or when they are generated from a vacuum state. In both cases the solutions for $\langle n_i^2\rangle$ and $\langle n_1 n_2\rangle$ have the form (we have ignored the terms n_{is}^2 and η_s^2)

$$\langle n_i^2\rangle=\langle n_i\rangle^2+(2n_{is}+1)\langle n_i\rangle,\quad i=1,2,\tag{19}$$

$$\langle n_1 n_2\rangle=\langle n_1\rangle\langle n_2\rangle+2\eta\eta_s.\tag{20}$$

Correspondingly, for the functions $G_i^{(2)}$ we obtain

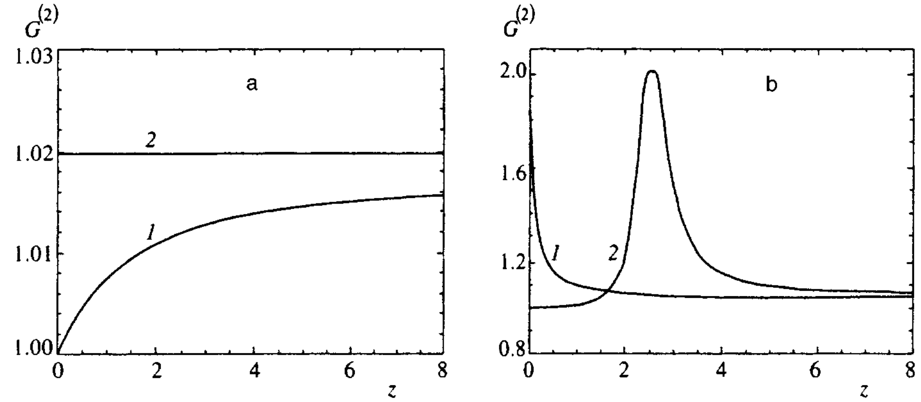


FIG. 3. Behavior of second-order correlation functions $G_1^{(2)}(z)$ (curve *1*) and $G_2^{(2)}$ (curve *2*) (in units of g^{-1}) with $\alpha=0.6\beta$ at (a) $n_{10}=100$ and $n_{20}=0$ and (b) $n_{10}=0$ and $n_{20}=100$.

$$G_i^{(2)}=1+\frac{2n_{is}}{\langle n_i\rangle}. \tag{21}$$

Figure 3 illustrates the behavior of $G_i^{(2)}$ for different initial conditions. We see that in the first case, where the initial Stokes field is coherent, $a_{10}\neq 0$ and $G_i^{(2)}=1$, and the parametric field is in a vacuum state, both modes are practically coherent from the start. In the opposite case the Stokes field is chaotic at small distances, but becomes coherent very fast and much earlier than $\langle n_1(z)\rangle$ reaches its stationary value. At the same time, the parametric mode becomes chaotic in the total-absorption region, but also rapidly restores the statistics of the initial coherent field. Thus, in both cases the quantum noise is considerably suppressed, which leads to a Poisson distribution in both modes.

Now let us discuss the size of the correlation of the photon-number fluctuations. It is determined by the cross-correlation function $C=\langle n_1n_2\rangle-\langle n_1\rangle\langle n_2\rangle$, which is zero for uncorrelated modes. The larger the value of C, the stronger the mixing of the fields and the greater the extent of correlation. Obviously, the only spontaneous photons in the modes that are correlated are those that are created in pairs. Indeed, Eq. (20) implies that

$$C=2\eta\eta_s, \tag{22}$$

i.e., $C\propto\eta_s$, which determines the process of pair emission of spontaneous photons at frequencies ω_1 and ω_2. Since $C\sim n_{i0}/g^4$, the correlation of photon-number fluctuations grows in proportion to the intensity of the signal field, and the closer the ratio β/α is to unity the faster the growth. Figure 4 shows that, all other things being equal, strong correlation exists for an initial coherent Stokes field. In any case, the positive value of C for large z suggests that there is correlated creation of spontaneous photons in the modes and the number of such photons increases with z, reaching its maximum in the stationary regime.

4. THE EMISSION SPECTRUM: CORRELATION OF PHASE FLUCTUATIONS

In linear theory the amplitude and phase fluctuations provide equal contributions to the emission-spectrum width. But in the high-gain region the contribution of amplitude fluctuations is suppressed due to stabilization of the amplitudes, with the result that the emission spectrum is determined solely by phase diffusion in each mode. We start with stochastic differential equations for the c-number complex-valued amplitudes A_i and A_i^+, $i=1,2$, whose stochastic averages correspond to the mathematical expectations of the operators $\langle a_i\rangle$ and $\langle a_i^+\rangle$. The equations can be obtained from Eq. (5) by expanding the density matrix ρ in the positive P-representation. They have the form

$$\begin{aligned}\frac{1}{c}\dot{A}_1&=\frac{\alpha}{2}A_1-\frac{\sqrt{\alpha\beta}}{2}A_2^++f_1,\\ \frac{1}{c}\dot{A}_2^+&=-\frac{\beta}{2}A_2^++\frac{\sqrt{\alpha\beta}}{2}A_1+f_2^+,\end{aligned} \tag{23}$$

where the fluctuation operators f_i and f_i^+ have zero mathematical expectations and delta-function correlation, which corresponds to the Markovian approximation in the corresponding decay processes. The values of the correlators are given in the Appendix.

Now we go over to the polar coordinates $A_j=r_j\exp(i\varphi_j)$, $j=1,2$, where the r_j are assumed constant and equal to $\sqrt{\langle n_j(\infty)\rangle}$. The asymptotic values of the number

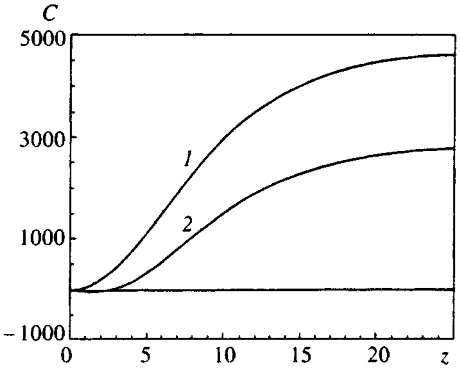


FIG. 4. Evolution of the cross-correlation C of two modes for two initial values of mode filling: curve *1*—$n_{10}=100$ and $n_{20}=0$, and curve *2*—$n_{10}=0$ and $n_{20}=100$; $\alpha=0.6\beta$.

of photons can be found from Eqs. (17) for $z \gg g^{-1}$. Then Eqs. (23) yield the following equations for the phases:

$$\begin{aligned} \frac{1}{c}\,\dot{\varphi}_1 &= \frac{1}{2}\,\sqrt{\alpha\beta}\,\frac{r_2}{r_1}\,\sin\,\phi + F_1(t), \\ \frac{1}{c}\,\dot{\varphi}_2 &= -\frac{1}{2}\,\sqrt{\alpha\beta}\,\frac{r_1}{r_2}\,\sin\,\phi + F_2(t), \end{aligned} \tag{24}$$

where $\phi = \varphi_1 + \varphi_2$, and the $F_i(t)$ are the phase noise operators with $\langle F_i(t)\rangle = 0$.

The diffusion coefficients of the $F_i(t)$ are also given in the Appendix. The emission spectrum can be calculated by the formula

$$\begin{aligned} S_j(\omega) &= \lim_{t\to\infty}\,\mathrm{Re}\int_{-\infty}^{\infty} d\tau\, e^{-i\omega\tau}\langle A_j^+(t+\tau)A_j(t)\rangle \\ &= \mathrm{Re}\,\int_{-\infty}^{\infty} d\tau\, e^{-i\omega\tau} S_j(\tau), \quad j = 1,\, 2, \end{aligned} \tag{25}$$

where, as it can easily be shown,

$$S_j(\tau) = \langle n_j(\infty)\rangle \exp[-\tfrac{1}{2}\langle \Delta\varphi_j^2(\tau)\rangle], \tag{26}$$

and the variances of the phases are expressed in terms of the diffusion coefficients as follows:

$$\langle \Delta\varphi_j^2(\tau)\rangle = 2D_{jj}\tau. \tag{27}$$

Plugging in the D_{jj}, we can use (25) to find the emission linewidths of the Stokes and parametric modes, respectively:

$$\gamma_1 = \frac{\alpha}{8\langle n_1(\infty)\rangle}, \quad \gamma_2 = \frac{\beta}{8\langle n_2(\infty)\rangle}. \tag{28}$$

These results coincide, to within a factor of $1/2$, with the Schawlow–Townes formula for the emission linewidth of a standard laser above the threshold with a linear-absorption mechanism. It is suppressed by a factor of two because earlier we ignored amplitude fluctuations. Using (10) to plug $\langle n_1(\infty)\rangle = \beta\langle n_2(\infty)\rangle/\alpha$ into (28), we arrive at the following relationship:

$$\frac{\gamma_2}{\gamma_1} = \left(\frac{\beta}{\alpha}\right)^2 > 1, \tag{29}$$

which shows that the parametric mode is broadened to a greater extent than the Stokes mode but that the Stokes mode contains more spontaneous photons than the parametric mode: $\langle n_{1s}\rangle > \langle n_{2s}\rangle$ (see Eqs. (18)). This is because the widths γ_1 and γ_2 are determined by different stochastic processes. Phase diffusion in the ω_1-mode is indeed caused by spontaneous-photon emission, and so $\gamma_1 \approx \alpha$. At the same time, the phase of the ω_2-mode diffuses because of photon emission with a probability $\beta > \alpha$, and this leads, on the one hand, to a decrease in the number of spontaneous photons in the second mode and, on the other, to an increase in emission linewidth. The result (29) for the ratio γ_2/γ_1 is a problem in its own right and can easily be verified in experiments.

Now let us discuss the extent to which phase fluctuations are correlated. Since, as shown earlier, the field amplitudes are correlated with opposite phases, the extent to which these fluctuations are correlated is determined by the diffusion of the total phase ϕ. An equation for this phase can be found from (24):

$$\frac{1}{c}\,\dot{\phi} = -g\,\sin\,\phi + F_1(t) + F_2(t), \tag{30}$$

which implies that the stationary value of ϕ is $\langle\phi(\infty)\rangle = \phi_0 = 0$. By linearizing the weak fluctuations near ϕ_0 in the form $\phi = \phi_0 + \delta\phi \cong \delta\phi$ we arrive at the following equation for the variance of the total phase:

$$\frac{1}{c}\,\frac{d\langle\delta\phi^2\rangle}{dt} = -2g\langle\delta\phi^2\rangle + 2\langle\phi[F_1(t) + F_2(t)]\rangle. \tag{31}$$

The second term on the right-hand side of Eq. (31) is calculated by the following formula (see Ref. 26):

$$\langle\phi F_j(t)\rangle = \sum_{i=1}^{2}\,\frac{d\phi}{d\varphi_i}\,D_{ij}(\varphi). \tag{32}$$

Using this, we can write the stationary solution of Eq. (31) in the form

$$\langle\delta\phi^2\rangle = \frac{(D_{11}+D_{22})(1-\varepsilon)}{g} = \frac{\beta-\alpha}{4\sqrt{\alpha\beta}\,r_1 r_2}, \tag{33}$$

where

$$\varepsilon = -\frac{D_{12}+D_{21}}{D_{11}+D_{22}} > 0.$$

We note, first, that the stationary solution for $\langle\delta\phi^2\rangle$ exists only if destructive interference is present, when $\beta > \alpha$. Second, even in this case the phase fluctuations are saturated to a constant value $(D_{11}+D_{22})g^{-1}$ if the correlation between the modes is weak $(D_{12} \ll D_{11})$ or is absent altogether. But thanks to the correlation of D_{12}, whose contribution becomes significant when the parametric coupling constant $\sqrt{\alpha\beta}$ becomes comparable to the losses β in the ω_2-mode, the fluctuations in the total phase may be completely (for all practical purposes) squeezed. And if we allow for the fact that the field amplitudes r_j are also large (see Eqs. (17)), the system in this case acts as a source of bright squeezed light.

Thus, when $\alpha < \beta$, the variance of the total phase is much smaller than the emission linewidths of the two modes. This means that no matter how the phases of the individual modes diffuse, these fluctuations cancel out in the total phase. Hence we can expect the distribution function P to have a sharp peak at $\langle\phi\rangle = 0$. The solution of the Fokker–Planck equation is outside the scope of the present paper. We only note here by using the results of Sec. 3 that the system becomes a classical mixture of correlated coherent states

$$|r_1 e^{i\varphi}\rangle_1 \left| r_1\,\sqrt{(\alpha/\beta)}\;e^{-i\varphi}\right\rangle_2$$

with an arbitrary phase φ, where the subscripts 1 and 2 refer to the ω_1- and ω_2-modes respectively. Some of these states are macroscopically separated in the phase space. Hence, if it is possible to prepare a system with a linear combination of these states, correlation between the fields allows generating a ''Schrödinger's cat'' state in one of the modes by performing a measurement in the other mode.

5. CONCLUSION

We have found that in conditions of destructive interference two events take place simultaneously: correlated amplification of the Stokes and parametric fields, and the squeezing of spontaneous noise in the total phase of the fields. In both cases the magnitude of the effect depends on the extent to which the modes are correlated, which can easily be monitored using its dependence on the intensity and two-photon detuning of the pump radiation. We also have found the conditions in which suppression of quantum fluctuations is practically complete. An important feature here is the fact that the effect is at its maximum when the medium is transparent, which occurs when the field intensities are constant and stable in the stationary propagation regime. These results and the classical solutions for the amplitudes and emission linewidths obtained in Secs. 3 and 4 can be verified in experiments, although there are some problems that have yet to be resolved. In particular, the results were obtained by ignoring pump depletion and cavity losses, two facts that can affect the correlation of phase fluctuations, especially if the individual losses in the modes are not controlled. These problems and the possibility of generating in each mode a superposition state of two macroscopically separated coherent states merit a separate investigation.

The present work was financed by the budget of the Republic of Armenia (Study No. 96-770).

APPENDIX

Here we give the diffusion coefficients of the noise operators f_i in Eqs. (23) calculated by Eq. (5):

$$\begin{aligned}
&\langle f_1^+(t) f_1(t')\rangle = \alpha\delta(t-t'),\\
&\langle f_1^+(t) f_2^+(t')\rangle = \langle f_2(t) f_1(t')\rangle = \sqrt{\alpha\beta}\,\delta(t-t'),\\
&\langle f_2(t) f_2^+(t')\rangle = \beta\delta(t-t'). \qquad \text{(A1)}
\end{aligned}$$

The other correlators are equal to zero.

Now let us calculate the diffusion coefficients of noise sources F_i in Eqs. (24). Allowing for the fact that

$$F_j = \frac{1}{2ir_j}\,[f_j^+ \exp(-i\varphi_j) - f_j \exp(i\varphi_j)],$$

and using Eqs. (A1), we find that

$$\langle F_l(t) F_m(t')\rangle = 2D_{lm}\delta(t-t'), \qquad \text{(A2)}$$

where, with allowance for $\langle\phi\rangle = 0$,

$$D_{11} = \frac{\alpha}{8r_1^2}, \quad D_{22} = \frac{\beta}{8r_2^2}, \quad D_{12} = D_{21} = -\frac{\sqrt{\alpha\beta}}{8r_1r_2}. \qquad \text{(A3)}$$

[1] C. M. Caves, Phys. Rev. D **26**, 1817 (1982).
[2] R. S. Bondurant and J. H. Shapiro, Phys. Rev. D **30**, 2548 (1984).
[3] W. Schleich and M. O. Scully, Phys. Rev. A **37**, 1261 (1988).
[4] J. Bergou, M. Orszag, and M. O. Scully, Phys. Rev. A **38**, 754 (1988); **38**, 768 (1988).
[5] M. O. Scully, K. Wodkiewicz, and M. S. Zubairy, J. Bergou, Ning Lu, and J. Meyer ter Vehn, Phys. Rev. Lett. **60**, 1832 (1988).
[6] R. S. Bondurant, P. Kumar, J. H. Shapiro *et al.*, Phys. Rev. A **30**, 343 (1984).
[7] P. Kumar and J. H. Shapiro, Phys. Rev. A **30**, 1568 (1984); M. D. Reid and D. F. Walls, Phys. Rev. A **31**, 1622 (1985); B. Yurke, Phys. Rev. A **32**, 300 (1985).
[8] R. E. Slusher, L. W. Hollberg, B. Yurke, J. C. Mertz, and J. F. Valley, Phys. Rev. Lett. **55**, 2409 (1985).
[9] R. M. Shelby, M. D. Levenson, S. H. Perlmutter, R. G. DeVoe, and D. F. Walls, Phys. Rev. Lett. **57**, 691 (1986).
[10] B. L. Schumaker, S. H. Perlmutter, R. M. Shelby, and M. D. Levenson, Phys. Rev. Lett. **58**, 357 (1987).
[11] M. Winters, J. L. Hall, and P. E. Toschek, Phys. Rev. Lett. **65**, 3116 (1990).
[12] Yu. P. Malakyan, Kvant. Elektron. (Moscow) **7**, 1365 (1985) [Sov. J. Quantum Electron. **15**, 905 (1985)]; **16**, 1879 (1989) [**19**, 1204 (1989)].
[13] Yu. P. Malakyan, J. Mod. Opt. **39**, 509 (1992).
[14] W. Hartig, Appl. Phys. **15**, 427 (1978).
[15] J. Heinrich and W. Behmenburg, Appl. Phys. **23**, 333 (1980).
[16] F. S. Tomkins and R. Manon, Opt. Lett. **6**, 179 (1981).
[17] P.-L. Zhang, Y.-C. Wang, and A. L. Schawlow, J. Opt. Soc. Am. B **1**, 9 (1984).
[18] Yu. P. Malakyan, Opt. Commun. **69**, 315 (1989).
[19] M. A. Moore, W. R. Garret, and M. G. Payne, Opt. Commun. **68**, 310 (1988).
[20] M.-H. Lu and Yu Mei Liu, Appl. Phys. B **57**, 167 (1993).
[21] Yu. P. Malakyan and A. R. Mkhitaryan, J. Mod. Opt. **43**, 537 (1996).
[22] Yu. P. Malakyan, Opt. Commun. **78**, 67 (1990).
[23] S. E. Harris, Phys. Rev. Lett. **72**, 52 (1994).
[24] J. H. Eberly, M. L. Pons, and H. R. Haq, Phys. Rev. Lett. **72**, 56 (1994).
[25] J. H. Eberly, Quantum Semiclassic. Opt. **7**, 373 (1995).
[26] M. Lax, *Fluctuation and Coherence Phenomena in Classical and Quantum Physics* [in Russian], Mir Publishers, Moscow (1976), p. 119 [English original edition: Gordon & Breach, New York (1968)].

Translated by Eugene Yankovsky

Polarization of light: fourth-order effects and polarization-squeezed states

D. M. Klyshko

M. V. Lomonosov Moscow State University, 119899 Moscow, Russia

(Submitted 1 November 1996)
Zh. Éksp. Teor. Fiz. **111**, 1955–1983 (June 1997)

The polarization properties of a monochromatic beam of light are ordinarily determined by three numbers, for example, the Stokes parameters. However, three numbers are no longer sufficient when intensity fluctuations in the polarized modes (or the correlation between them) are recorded. It is shown that in this case nine parameters, which can be arranged into 3×3 matrices, must be prescribed. The transformation properties of these matrices under polarization converters and the invariants of the matrices are analyzed. Specifically, the fourth-order polarization P_4 is introduced. Several examples are examined of light with ''hidden'' polarization—light which is not polarized in the ordinary sense ($P_2=0$) but is polarized in fourth order ($P_4\neq0$)—as well as ''polarization-squeezed'' light in which the quantum fluctuations of the Stokes parameters are suppressed.
[S1063-7761(97)00406-X]

1. INTRODUCTION

The polarization of a quasimonochromatic plane wave is ordinarily described by three numbers, for example, the degree of polarization P_2, the ratio of the axes, and the orientation of the polarization ellipse (see Ref. 1). Another possible set is the three Stokes parameters S_1, S_2, and S_3 normalized to the overall intensity S_0 of the wave. The polarization can also be fixed by two spherical coordinates Θ, Φ of a point on the Poincaré sphere and the degree of polarization P_2, or by the 2×2 polarization matrix $\mathbf{K}$. These equivalent sets completely determine the properties of the wave with respect to different polarization converters placed in front of a detector (see Fig. 1) which records the average intensity of the light, i.e., the second moment of the field. However, as is clear from the example in Fig. 2, three parameters are clearly inadequate in cases when not the average intensity but rather other statistical parameters of the wave are recorded. Specifically, the concept of polarized and unpolarized light needs to be generalized.

In the present paper sets of nine real parameters which are necessary to describe experiments measuring the fourth-order moments in the field amplitudes are examined. The transformations of these parameters under the action of polarization converters and the corresponding invariants are determined. In particular, the degree of fourth-order polarization P_4 is introduced. This parameter characterizes the degree of anisotropy of the fluctuations of the Stokes vector: $P_4=0$ in the case $\Delta S_1=\Delta S_2=\Delta S_3$ and $P_4=1$ in the case $\Delta S_X=0$ for some direction $\mathbf{X}$. An example of light with $P_4=0$ is light in a two-mode coherent state $|\alpha,\beta\rangle$. In this case we have $P_2=1$ and $\Delta S_k=S_0 (k=1,2,3)$—polarization noise is caused only by quantum (photon) fluctuations. States with a definite number of photons in two polarization modes possess the property $P_4=1$.

The development of a convenient formalism for describing fourth-order polarization effects is an important problem. It could be helpful for systematizing some experiments in quantum optics as well as in connection with the problem of increasing the accuracy of polarization interferometry, polarimetry, measurements of the relative phase of light beams, and so on. The accuracy of optical measurements is now limited by quantum fluctuations and the possibility of overcoming this limit involves squeezed states of light beams.[2,3] Quantum cryptography methods, which employ polarization modulation of light, pose new problems.[4]

The mathematical description of two polarization modes of one beam and their transformations is isomorphic to the description of two spatially separated polarized beams of light—in both cases the amplitudes a, b of the modes and their linear unitary transformations realize a representation of the group $SU(2)$ (in both the classical and quantum descriptions). As a result, the problems of increasing the sensitivity of Mach–Zender two-beam interferometers[2,3] and polarization interferometers have much in common. The close relationship between the group $SU(2)$ and the rotation group $SO(3)$ makes it possible to represent transparently the evolution of a two-mode field in three-dimensional space by means of the Stokes vector. In Ref. 3 an analog of the Stokes vector was essentially used to describe a Mach–Zender interferometer.

Recording the fourth or higher moments, even of light which is unpolarized in the ordinary sense, can exhibit hidden polarization. This concept was proposed in Ref. 5 to describe light beams which are not polarized in the ordinary sense ($P_2=0$) but exhibit a transverse structure when the fourth ($P_4\neq0$) or higher moments of the field are recorded. For example, a superposition of the fields of two independent lasers with the same frequencies, the same average intensities, and orthogonal linear polarizations gives light with hidden polarization: Even though we have $P_2=0$, the intensity fluctuations and correlation observed in the scheme in Fig. 1 depend on the parameters of the polarization converter.[5–8] Such an experiment was recently performed by Guzun and Penin.[8] We underscore that replacing lasers (which possess stable intensities) by ordinary light sources with Gaussian statistics gives $P_2=P_4=0$. Then the light is unpolarized in all orders, it has no distinguished polarization basis, and it is completely invariant under the action of any polarization converter.[5,6] This property of a plane wave can be termed

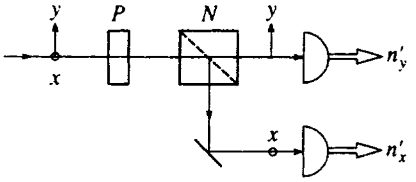


FIG. 1. Arrangement of an experiment on measuring the polarization properties of a beam of monochromatic light. The light passes through a polarization converter P with known parameters and then through a polarization beam splitter (Nicol prism) N, which separates two orthogonal linearly polarized components (x,y) of the field at the exit of P. Two detectors record the "instantaneous" intensities n'_x and n'_y. Moments of the form $\langle n'^k_x n'^l_y \rangle$ can be determined by repeating the procedure and performing a statistical analysis of the data.

"transverse isotropy." Another example of hidden polarization is a two-photon light beam consisting of a flux of pairs of photons with orthogonal polarization.[5,6] In this case, the amplitudes of the fields in the modes do not possess a definite phase difference (as in the case of independent lasers) and therefore all types of polarization are equivalent, $P_2=0$. Such light is obtained in practice by means of parametric scattering (type-II, i.e., with orthogonally polarized modes) in crystals with a quadratic nonlinearity. The corresponding state of the field is called a two-mode squeezed vacuum.

Radiation with type-II parametric scattering possesses another important feature: Since photons in modes are produced only in pairs, simultaneously, the fluctuations of the intensities N_x and N_y in the modes are completely correlated (this effect was confirmed experimentally in Ref. 9). This means that one of the Stokes parameters does not fluctuate and even quantum noise is absent: $\Delta S_1=0$.[7] Chirkin *et al.*[10] introduced the convenient term polarization-squeezed light for radiation in which quantum fluctuations of some components of the Stokes vector $\mathbf{S}_k (k=1,2,3)$ are diminished. In Refs. 10–13 methods for obtaining such light by anisotropic cubic nonlinearity of transparent materials were examined.

For completeness of exposition and continuity of the notations, in Secs. 2 and 3 below some well-known relations of polarization optics are repeated using the modern notation. In Sec. 2 the Stokes vector $\mathbf{S}$ and the corresponding operator $\mathbf{s}(\mathbf{S}\equiv\langle \mathbf{s}\rangle$ as well as the second-order polarization matrix $\mathbf{K}$ and its corresponding operator $\mathbf{k}(\mathbf{K}\equiv\langle \mathbf{k}\rangle$ are introduced. Section 3 is devoted to a description of the transformations of these operators in the Heisenberg representation $(\mathbf{s}'=\mathbf{Rs},\mathbf{k}'=\mathbf{Dk})$ with the aid of the Mueller $\mathbf{R}$ (H. Mueller, 1943) and Jones $\mathbf{D}$ (R. C. Jones, 1941) matrices from classical optics (see Ref. 1). For comparison, the action of polarization converters is also examined in the Schrödinger representation. The relations with the quantities which are observed in a typical experiment are described in Sec. 4. In Sec. 5, two parallel sets of fourth-order parameters arranged into two matrices $\mathbf{K}_4=\langle :\mathbf{k}\otimes\mathbf{k}:\rangle$ and $\mathbf{Q}\equiv\langle :\mathbf{s}\otimes\mathbf{s}:\rangle$ are introduced. Their transformation properties are analyzed in Sec. 6. Next, in Secs. 7–9 this formalism used to analyze several characteristic cases: coherent states, two lasers with independent phases, N-photon states, generalized coherent states of the $SU(2)$ group,[14–16] and a squeezed two-mode vacuum and the effect of additional coherent components (homodyne field) on it. The analysis is limited to the case of a quasimonochromatic stationary plane wave and its loss-free polarization conversions. Mainly the Heisenberg representation is used, which makes it possible to underscore the general features of the quantum and classical approaches to the description of polarization effects.

2. FOURTH-ORDER POLARIZATION PARAMETERS

A quasimonochromatic plane wave is described by the creation operators $a_x^+\equiv a^+$ and $a_y^+\equiv b^+$ and annihilation operators $a_x\equiv a$ and $a_y\equiv b$ for photons in two orthogonally polarized modes x, y. In accordance with the experimental procedure under study, which employs a polarization beam splitter (Fig. 1), we assume that the mode x possesses linear horizontal polarization and the mode y possesses linear vertical polarization. The quantum properties of the field are determined by the commutation relations of the Heisenberg–Weyl group $W(2)$ (see Ref. 14):

$$[a,a^+]=[b,b^+]=1,$$
$$[a,a]=[b,b]=[a,b]=[a,b^+]=0. \tag{2.1}$$

Analyzing the polarization transformations (just as other linear transformations, see Refs. 5 and 6) in the Heisenberg representation makes it possible to switch easily from quantum optics to classical statistical optics, based on the classical representation of the probabilities $P(a,b)$ for the ampli-

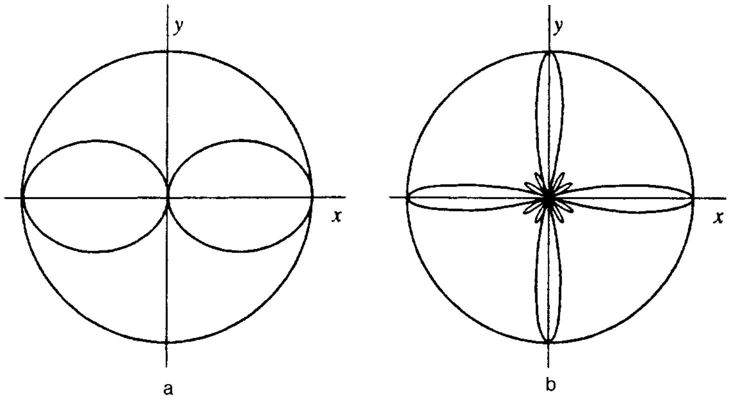


FIG. 2. Examples of the manifestation of the transverse structure of a plane wave in second and twelfth orders. a—Malus' law $\cos^2\chi$ in polar coordinates, i.e., the probability of detecting a photon in channel x in Fig. 1 as a function of the orientation of the Nicol prism (the incident light is polarized along the x direction). b—$[P_3(\cos 2\chi)]^2$, proportional to the probability of detecting three photons in each of the two exit modes of the Nicol prism in the case when three photons are present in the entrance modes ($P_3(x)$ is a Legendre polynomial).[5]

tudes of the field. In this case, a and b are dimensionless classical amplitudes of the field in the modes $(a^+\to a^*, b^+\to b^*)$.

The polarization properties of a monochromatic plane wave are ordinarily determined in terms of the following sets of second moments:[1]

$$\langle a^+a\rangle\equiv\langle n_x\rangle\equiv N_x,\quad \langle b^+b\rangle\equiv\langle n_y\rangle\equiv N_y,$$
$$\langle a^+b\rangle\equiv\langle s_+\rangle\equiv S_+,\quad \langle ab^+\rangle\equiv S_-=S_+^*. \tag{2.2}$$

The nine stationary moments of the form $\langle a^2\rangle$, $\langle ab\rangle$, and $\langle b^2\rangle$, as a rule, have no effect on the observable quantities and therefore are not considered (see Sec. 9 for an exception). The statistics of the field is assumed to be stationary and ergodic so that the time-averaged values of the observed quantities can be described by means of a quantum statistical ensemble, in which case $\langle\ldots\rangle=\mathrm{Tr}(\rho\ldots)$, where ρ is the field density operator. To switch to the classical description it is sufficient to interpret the operation $\langle\ldots\rangle$ as a classical average performed using $P(a,b)$ and to drop the quantum noise, i.e., only the normal-ordered moments are retained.

An alternative set of numbers is also employed in parallel with the moments (2.2)—the Stokes parameters S_0, S_1, S_2, and S_3. We shall define them as the average values $S_m\equiv\langle s_m\rangle$ of the following Stokes operators:

$$s_0\equiv a^+a+b^+b,\quad s_1\equiv a^+a-b^+b,$$
$$s_2\equiv a^+b+ab^+,\quad s_3\equiv -i(a^+b-ab^+). \tag{2.3}$$

From Eqs. (2.2) and (2.3) follow the relations

$$S_0=N_x+N_y,\quad S_1=N_x-N_y,\quad S_2=2\,\mathrm{Re}\,S_+,$$
$$S_3=2\,\mathrm{Im}\,S_+. \tag{2.4}$$

The Cauchy–Schwartz inequality gives the restriction $|S_k|\leqslant S_0(k=1,\,2,\,3)$. The relations inverse to Eq. (2.3) have the form

$$n_x=\frac{1}{2}(s_0+s_1),\quad s_+=\frac{1}{2}(s_2+is_3),$$
$$n_y=\frac{1}{2}(s_0-s_1),\quad s_-=\frac{1}{2}(s_2-is_3). \tag{2.5}$$

It follows from Eqs. (2.1) and (2.3) that

$$[s_1,s_2]=2is_3,\quad [s_2,s_3]=2is_1,\quad [s_3,s_1]=2is_2,$$
$$[s_0,s_m]=0,\quad s_1^2+s_2^2+s_3^2=s_0(s_0+2). \tag{2.6}$$

Therefore the operators s_m form a Lie algebra $su(2)$ $(m=0,1,2,3)$. In other words, the commutation relations for the operators $(1/2)s_k$ are the same as for the operators projecting the angular momenta j_k of the particles with spin $J=(1/2)S_0$. This gives additional possibilities for classifying the possible states of a two-mode field. According to Eq. (2.6), the second moments of the Stokes operators are related by

$$\langle s_1^2\rangle+\langle s_2^2\rangle+\langle s_3^2\rangle-\langle s_0^2\rangle=2S_0. \tag{2.7}$$

It is convenient to expand the moments N_x, N_y, S_+, and $S_-=S_+^*$ in the form of a hermitian polarization matrix $\mathbf{K}\equiv\langle\mathbf{k}\rangle$ (also called the coherence matrix[1]):

$$\mathbf{K}=\begin{pmatrix}N_x & S_+\\ S_- & N_y\end{pmatrix}=\begin{pmatrix}S_0+S_1 & S_2+iS_3\\ S_2-iS_3 & S_0-S_1\end{pmatrix}$$
$$=S_0\mathbf{I}+\sum_{k=1}^{3}S_k\sigma_k^*. \tag{2.8}$$

Here σ_k are the Pauli matrices which together with the unit matrix $\mathbf{I}$ serve as a convenient basis in the space of arbitrary 2×2 hermitian matrices.

The moments N_x, N_y, and S_+ given by (2.2) and the Stokes parameters S_m are uniquely related and carry identical information, but the parameters S_m possess simpler transformation properties and admit a transparent geometric representation. Thus, the statistical properties of a wave in second order in the field amplitudes are determined by four real parameters and a point in the space $\mathbf{R}^4$ can be associated to them. Under the action of a polarization converter this point moves along some orbit.

In the case of a loss-free converter the total intensity $s_0=n_x+n_y$ of the wave is conserved and therefore only three parameters need be taken into account, for example, the three components S_k $(k=1,2,3)$. It is convenient to represent them graphically as a vector $\mathbf{S}$ in a three-dimensional Euclidean space $\mathbf{R}^3$, which we shall call the Stokes–Poincaré space. Let the vectors $\mathbf{e}_k$ form an orthonormalized basis of this space, i.e., $\mathbf{e}_k\cdot\mathbf{e}_l=\delta_{kl}$. Then we have the representation

$$\mathbf{S}=\sum_{k=1}^{3}S_k\mathbf{e}_k,$$

where $S_k=\mathbf{e}_k\cdot\mathbf{S}$. The norm of the Stokes vector

$$S=\sqrt{\sum S_k^2}=\sqrt{(\mathrm{Tr}\,\mathbf{K})^2-4\det\mathbf{K}}\equiv P_2S_0 \tag{2.9}$$

together with S_0 determines the ordinary (second-order) degree of polarization P_2. The normalized Stokes vectors $\mathbf{S}/S$ map onto a Poincaré sphere, each point of which corresponds to a definite state of polarization. For example, horizontal polarization corresponds to the point $\mathbf{e}_1=(1,0,0)$, linear polarization at an angle of 45° to the x axis corresponds to $\mathbf{e}_2=(0,1,0)$, and right-hand circular polarization corresponds to $\mathbf{e}_3=(0,0,1)$.

The vector Stokes operator is defined analogously:

$$\mathbf{s}\equiv\sum_{k=1}^{3}s_k\mathbf{e}_k.$$

Its square is also an operator:

$$\mathbf{s}^2=(\mathbf{ss})=\sum s_k^2=s_0^2+2s_0=:s_0^2:+3s_0 \tag{2.10}$$

(the double dots indicate normal ordering with respect to the operators a^+, a and b^+, b).

3. POLARIZATION CONVERSION AS A ROTATION

If the phase factor common to both modes is neglected, then the effect of a loss-free converter can be represented in the Heisenberg representation in the form

$$a'=t^*a+r^*b,\quad b'=-ra+tb. \tag{3.1}$$

Here t are r are the amplitude transmission and reflection coefficients of the converter. For example,

$$t=\cos\,\delta+i\,\sin\,\delta\cos 2\chi,\quad r=i\,\sin\delta\sin 2\chi \tag{3.2}$$

corresponds to a linear phase plate with optical thickness δ and orientation χ relative to the horizontal direction x (a $\lambda/4$ plate gives $\delta=\pi/4$). In vector form we have $\mathbf{a}'=\mathbf{D}^*\mathbf{a}$, where $\mathbf{a}$ is a vector with components a and b and

$$\mathbf{D}\equiv\begin{pmatrix} t & r \\ -r^* & t^* \end{pmatrix}. \tag{3.3}$$

The matrices $\mathbf{D}$ were introduced into classical optics by Jones (see Ref. 11). In the absence of losses they are unitary and unimodal

$$\mathbf{D}\mathbf{D}^+=\mathbf{D}^+\mathbf{D}=I,\quad \det\mathbf{D}=|t|^2+|r|^2=1, \tag{3.4}$$

and therefore they realize the unitary representation of the $SU(2)$ group.

We shall adopt the following parameterization of the matrices $\mathbf{D}$:[17,18]

$$t(\phi,\theta,\psi)\equiv\cos(\theta/2)\exp[i(\phi+\psi)/2],$$

$$r(\phi,\theta,\psi)\equiv-\sin(\theta/2)\exp[i(\phi-\psi)/2]. \tag{3.5}$$

Here it can be assumed that $0\leqslant\theta<\pi$, $0\leqslant\psi<2\pi$, and $0\leqslant\phi<4\pi$. The inverse relations have the form

$$\phi=\arg\,t+\arg\,r+\pi,$$

$$\psi=\arg\,t-\arg\,r-\pi,\quad \theta=2\,\arctan|r/t|. \tag{3.6}$$

Using Eqs. (3.1), we find the action of a converter on the operators n_x, n_y, and $s_+=a^+b$:

$$n_x'=Tn_x+Rn_y+tr^*s_++t^*rs_+^*,$$

$$n_y'=Rn_x+Tn_y-tr^*s_+-t^*rs_+^*,$$

$$s_+'=tr(n_y-n_x)+t^2s_+-r^2s_+^*. \tag{3.7}$$

Here $T\equiv|t|^2$ and $R\equiv|r|^2$. Averaging these expressions we find the transformed polarization matrix

$$\mathbf{K}'=\mathbf{D}\mathbf{K}\mathbf{D}^+. \tag{3.8}$$

The parameters

$$\det\mathbf{K}=N_xN_y-|S_+|^2,\quad \mathrm{Tr}\,\mathbf{K}=N_x+N_y\equiv S_0,$$

$$S\equiv(S_1^2+S_2^2+S_3^2)^{1/2},\quad P_2=S/S_0$$

are invariants of the converter.

It follows from Eqs. (3.7) that in the case of measurements which are performed according to the scheme in Fig. 1 the two transformations $\mathbf{D}$ and $-\mathbf{D}$ give the same observed quantities—these transformations differ only by the identical half-wave phase change in both channels. Therefore the maximum value of ϕ can be limited by 2π (instead of 4π). The remaining set of matrices $\mathbf{D}$ now represents only the rotation group $SO(3)$, which is a subgroup of $SU(2)$. (We note that this narrowing of the space of the representation leads to a loss of information about the general phase of the wave, which can be observed in some experimental arrangements; see Ref. 19.)

The action of all possible converters on three-dimensional real Stokes vectors $\mathbf{S}$ realizes the irreducible representation of the group $SO(3)$ in the space $\mathbf{R}^3$: A definite rotation of the vector $\mathbf{S}$ in the Stokes–Poincaré space can be associated to each converter. We now substitute the expression (3.1) into the definition (2.3). The result can be represented in the form $\mathbf{s}'=\mathbf{R}\mathbf{s}$, where

$$\mathbf{R}(t,r)\equiv\begin{pmatrix} |t|^2-|r|^2 & 2\,\mathrm{Re}(tr^*) & -2\,\mathrm{Im}(tr^*) \\ -2\,\mathrm{Re}(tr) & \mathrm{Re}(t^2-r^2) & -\mathrm{Im}(t^2+r^2) \\ -2\,\mathrm{Im}(tr) & \mathrm{Im}(t^2-r^2) & \mathrm{Re}(t^2+r^2) \end{pmatrix}. \tag{3.9}$$

In classical optics the Mueller matrices (see Ref. 1) correspond to the matrices $\mathbf{R}$. In the presence of dissipation the Mueller matrices are 4×4 matrices and also describe the transformation of the component S_0. Substituting here the expressions (3.5), we obtain a different parameterization

$$\mathbf{R}(\phi,\theta,\psi)$$

$$=\begin{pmatrix} c_\theta & -s_\theta c_\psi & s_\theta s_\psi \\ c_\phi s_\theta & -s_\phi s_\psi+c_\phi c_\theta c_\psi & -s_\phi c_\psi-c_\phi c_\theta s_\psi \\ s_\phi s_\theta & c_\phi s_\psi+s_\phi c_\theta c_\psi & c_\phi c_\psi-s_\phi c_\theta s_\psi \end{pmatrix}. \tag{3.10}$$

Here $c_x\equiv\cos x$, $s_x\equiv\sin x$. We have expressed $\mathbf{R}$ in terms of the Euler angles ϕ, θ, and ψ, describing the rotation of a solid with a fixed point (the polar axis is the axis 1). The equivalent rotation of the coordinate axes in the Stokes–Poincaré space (''passive point of view'') is performed by the inverse matrix $\mathbf{R}^{-1}$ (here $(R^{-1})_{mn}=R_{nm}$, since $\mathbf{R}$ is orthogonal). It is easy to verify that

$$\mathbf{R}^{-1}(\phi,\theta,\psi)=\mathbf{R}(-\psi,-\theta,-\phi)=\mathbf{R}(\pi-\psi,\theta,\pi-\phi).$$

Any rotation in $\mathbf{R}^3$ can also be specified by indicating the direction of the rotation axis $\mathbf{n}$ and the rotation angle $\alpha(\alpha=0-2\pi)$, i.e., a different parameterization of the converter is possible:

$$\mathbf{R}=\mathbf{R}(\mathbf{n},\alpha)\equiv\mathbf{R}(\alpha\mathbf{n}).$$

It is assumed that $\alpha>0$ corresponds to a clockwise rotation, looking along the direction $\mathbf{n}$. Then any converter corresponds to a vector $\alpha\,\mathbf{n}$ in the Stokes–Poincaré space and a point $\mathbf{n}$ on the Poincaré sphere. It follows from the definition of the rotation axis that $\mathbf{n}$ is a characteristic vector of $\mathbf{R}$ with eigenvalue $+1$. According to Ref. 18

$$\cos\,\alpha=\frac{1}{2}\,[\mathrm{Tr}\,\mathbf{R}-1]$$

$$=\mathrm{Re}\,t^2-|r|^2=\cos(\psi+\phi)\cos^2(\theta/2)-\sin^2(\theta/2) \tag{3.11}$$

or, in a different form,

$$\cos(\alpha/2)=\pm\cos(\theta/2)\cos[(\psi+\phi)/2]. \tag{3.12}$$

The vector $\mathbf{n}$ has the following components:[18]

$$n_1=C(R_{32}-R_{23})=2C\,\,\mathrm{Im}\,t^2=C(1+\cos\,\theta)\sin(\psi+\theta).$$

$$n_2=C(R_{13}-R_{31})=4C\,\,\mathrm{Re}\,t\,\,\mathrm{Im}\,r$$

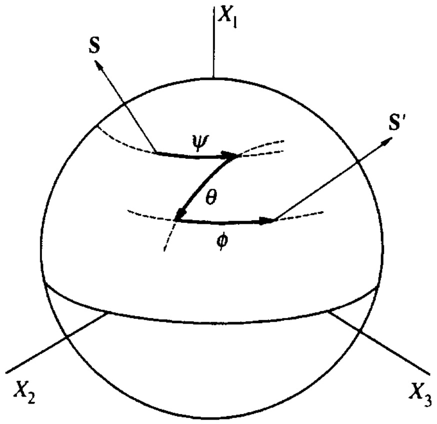


FIG. 3. Representation of the transformation of the polarization of a plane wave as a result of three successive rotations of the Stokes vector **S** in the Stokes–Poincaré space by an angle ψ about the axis *1*, by an angle θ about the axis *3*, and by an angle ϕ about the axis *1*.

$$=C\sin\theta(\sin\psi-\sin\phi),$$

$$n_3=C(R_{21}-R_{12})=-4C\,\mathrm{Re}\,t\,\mathrm{Re}\,r$$

$$=C\sin\theta(\cos\psi+\cos\phi). \tag{3.13}$$

Here $C\equiv 1/2\sin\alpha$ is a normalization constant. The functions $\alpha(t,r)n_k(t,r)$ (or $\alpha(\psi,\theta,\psi)n_k(\psi,\theta,\psi)$), defined in Eqs. (3.11) and (3.13), define a map of the set of parameters (t,r) (or (ψ,θ,ψ)) of the converter into the Stokes–Poincaré space; this makes it possible to represent conveniently rotations of the Stokes vector under the action of the converter. The direction of the rotation axis in the spherical coordinates is determined from Eq. (3.13) and the relations $n_1=\cos\Theta$, $n_2=\sin\Theta\cos\Phi$, and $n_3=\sin\Theta\sin\Phi$. For example, for a $\lambda/4$ phase plate we obtain from Eq. (3.2) $\alpha=2\delta$, and the rotation axis **n** lies in the *(1,2)* plane at an angle 2χ with respect to the axis *1*, i.e., $\Theta=2\chi$, $\Phi=0$. Therefore the effect of a plate with the parameters (δ,χ) is to rotate the vector **S** by an angle 2δ about the direction $(2\chi,0)$.

An arbitrary rotation **R** can be represented in the form of three successive rotations around the basis axes $\mathbf{e}_1$ and $\mathbf{e}_3$ (Fig. 3):

$$\mathbf{R}(\phi,\theta,\psi)=\mathbf{R}(\phi,0,0)\mathbf{R}(0,\theta,0)\mathbf{R}(0,0,\psi)$$

$$=\mathbf{R}_1(\phi)\mathbf{R}_3(\theta)\mathbf{R}_1(\psi), \tag{3.14}$$

where $\mathbf{R}_n(\alpha)\equiv\mathbf{R}(\alpha\mathbf{e}_n)$. The unit vector $\mathbf{e}_1$ transforms under the action of **R** into $\mathbf{e}_1'=\mathbf{R}_1(\phi)\mathbf{R}_3(\theta)\mathbf{e}_1$. According to Eq. (3.10),

$$(\mathbf{e}_1')_k=R_{k1}=(\cos\theta,\sin\theta\cos\phi,\sin\theta\sin\phi),$$

i.e., the spherical coordinates of the rotated vector $\mathbf{R}(\phi,\theta,\psi)\mathbf{e}_1$ are identical to the Euler angles of the converter: $\Theta=\theta$, $\Phi=\phi$. The parameter ψ has no effect, since $\mathbf{e}_1$ is a characteristic vector for $\mathbf{R}_1(\psi)$. On the other hand, the quantity

$$\mathbf{s}_1'=\sum R_{1k}\mathbf{s}_k=\mathbf{s}_1\cos\theta-\mathbf{s}_2\sin\theta\cos\psi$$

$$+\mathbf{s}_3\sin\theta\sin\psi \tag{3.15}$$

observed in the experiment in Fig. 1 does not depend on ϕ—the common phase factor of t and r.

A representation analogous to (3.14) is also possible for the Jones matrices **D**:

$$\mathbf{D}(\phi,\theta,\psi)=\mathbf{D}(\phi,0,0)\mathbf{D}(0,\theta,0)\mathbf{D}(0,0,\psi)$$

$$=\begin{pmatrix} e^{i\phi/2} & 0 \\ 0 & e^{-i\phi/2}\end{pmatrix}\begin{pmatrix}\cos(\theta/2) & -\sin(\theta/2)\\ \sin(\theta/2) & \cos(\theta/2)\end{pmatrix}$$

$$\times\begin{pmatrix} e^{i\psi/2} & 0 \\ 0 & e^{-i\psi/2}\end{pmatrix}. \tag{3.16}$$

We shall now give a brief description of the effect of a converter in the Schrödinger representation, when the converter does not change the field operators but rather the wave function of the two-mode field: $|\psi\rangle\rightarrow|\psi\rangle'$. It is easy to show that

$$|\psi\rangle'=U|\psi\rangle,$$

$$U(\phi,\theta,\psi)\equiv\exp(-i\phi j_1)\exp(-i\theta j_3)\exp(-i\psi j_1). \tag{3.17}$$

(This follows from the equivalence of the transformations $\mathbf{s}'=\mathbf{R}\mathbf{s}$ and $\mathbf{s}'=U^+\mathbf{s}U$.) Here the operators $j_k=(1/2)s_k$ play the role of group generators, i.e., $j_1=i[dU/d\phi]_0$ and so on.

The wave function $|\psi\rangle'$ can also be expressed directly in terms of the parameters t, r of the Jones matrix **D**. Let the initial state at the converter entrance have the form

$$|\psi\rangle=f(a^+,a,b^+,b)|\mathrm{vac}\rangle.$$

Then it is easy to show that at the exit

$$|\psi\rangle'=f'(a^+,a,b^+,b)|\mathrm{vac}\rangle,$$

$$f'\equiv f(t^*a^+-rb^+,ta-r^*b,r^*a^++tb^+,ra+t^*b). \tag{3.18}$$

Here the function f' is formed from f by the substitution $\mathbf{a}\rightarrow(\mathbf{D}^*)^{-1}\mathbf{a}$, $\mathbf{a}^+\rightarrow\mathbf{D}^{-1}\mathbf{a}^+$.

We shall examine several possible bases of the space of states of a two-dimensional oscillator **H**.

1. The effect of a converter is most simply described in a continuous basis $|\alpha,\beta\rangle\equiv|\alpha\rangle_x|\beta\rangle_y$ formed by the coherent states of the group $W(2)$—the characteristic vectors of the operators $a_x\equiv a$ and $a_y\equiv b$.[14] Then **H** maps into the space $\mathbf{C}^2$ or $\mathbf{R}^4$. Substituting into Eq. (3.18) the displacement operator $f(a^+,a,b^+,b)=\exp(\alpha a^+-\alpha^*a+\beta b^+-\beta^*b)$, we obtain

$$|\alpha,\beta\rangle'=|t^*\alpha+r^*\beta,-r\alpha+t\beta\rangle. \tag{3.19}$$

The set of states (3.19) comprises the "energy" surface in **H** or the orbit of the element $|\alpha,\beta\rangle$ on which it moves under the action of different unitary transformations. In the case of a coherent state a unit polarization vector $\mathbf{e}(\alpha,\beta)$ analogous to the classical polarization vector can be defined:

$$\mathbf{e}\equiv(e_x,e_y),\quad e_x\equiv\frac{\alpha}{\sqrt{N}},\quad e_y=\frac{\beta}{\sqrt{N}},\quad N\equiv|\alpha|^2+|\beta|^2.$$

Here, just as in classical optics, $\mathbf{e}'=\mathbf{D}^*\mathbf{e}$.

2. A discrete (Fock) basis $|N_x,N_y\rangle\equiv|N_x\rangle_x|N_y\rangle_y$ of the space **H** is generated by the common eigenvectors of the

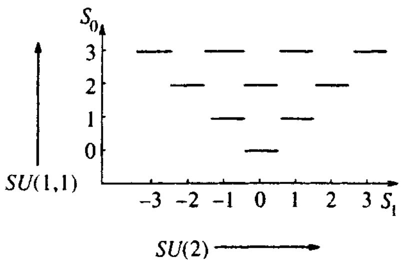


FIG. 4. Classification of the states of a plane wave with a definite number of photons N_x and N_y in the polarization modes. The difference $S_1=N_x-N_y$ is plotted along the abscissa and the sum $S_0=N_x+N_y$ is plotted along the ordinate. Loss-free polarization converters give rise to ''horizontal'' transitions inside the polarization multiplet ($S_0=$const) (in this case, generalized coherent states of the group $SU(2)$ are formed). ''Vertical'' transitions with a change in energy S_0 and with conservation of S_1 are produced by parametric converters and are associated with the group $SU(1,1)$.

operators n_x, n_y and is parameterized by two nonnegative integers N_x, N_y—the numbers of photons in the polarization modes.

3. A different parameterization of the space **H** is also possible—parameterization by means of the numbers $N=S_0=N_x+N_y$ and $S_1=N_x-N_y$. N photons can be distributed between two modes in $N+1$ ways. Therefore the stationary state of a two-dimensional oscillator with energy N is $(N+1)$-fold degenerate, i.e., it is a multiplet (see Fig. 4). The general eigenvectors of the energy operator $s_0=n=n_x+n_y$ and the energy difference operator between the two modes $s_1=n_x-n_y$ form a basis in **H**, parameterized by two integers—the eigenvalues of these operators $S_0=0, 1, 2,\ldots,$ and $S_1=-N, -N+2,\ldots,N$.

To switch to the standard notation of the theory of angular momenta, we set $S_0\equiv 2J$ and $S_1\equiv 2M$. Then the basis vectors assume the form $|J,M\rangle$, where $M=-J,-J+1,\ldots,J$. In the old notation $|J,M\rangle=|N_x\rangle_x|N_y\rangle_y$, where $N_x=J+M$, $N_y=J-M$, $J=N/2$. For example, $|J,-J\rangle=|0\rangle_x|N\rangle_y$ and $|J,J\rangle=|N\rangle_x|0\rangle_y$. The vectors $|J,M\rangle$ form a so-called canonical basis in **H**:

$$\exp(-i\phi j_1)|J,M\rangle=\exp(-iM\phi)|J,M\rangle.$$

Let us fix N and study an arbitrary state $|\psi\rangle_J \in \mathbf{H}_j$ with $N=2J$ photons. In the basis $|J,M\rangle\equiv|M\rangle$ it has the form

$$|\psi\rangle_J=\sum_{M=-J}^{J} e_M|M\rangle,$$

so that the polarization properties of the field are completely described by the unit vector $\mathbf{e}^{(J)}\equiv\mathbf{e}\equiv\{e_M\}$ with $N+1$ components (the vector $\mathbf{e}^{(1/2)}$ transforms in the same manner as the standard classical polarization vector). The corresponding projective space is the sphere $\mathbf{S}^{2N}$. The state space $\mathbf{H}_J$ with a fixed total number of photons $N=2J$ is invariant relative to arbitrary loss-free converters. Within the space $\mathbf{H}_J$ the operators of the observables $j_k=(1/2)s_k$ (like the operators $U=U^{(J)}$, according to Eq. (3.17)) are $(N+1)\times(N+1)$ matrices. For example, in the case $N=1$ we have $j_k\to(1/2)\sigma_k$. The effect of the converter on the basis vectors $|M\rangle$ of a multiplet is described by the well-known matrices $\mathbf{D}^{(J)}(\phi,\theta,\psi)$ (Wigner functions), which give the unitary irreducible representations of the group $SU(2)$ in the space $\mathbf{H}_J=\{|M\rangle\}$:

$$|M\rangle' = U|M\rangle = \sum_{M_1} D^{(J)}_{M_1 M}|M_1\rangle. \tag{3.20}$$

(The explicit form of the functions $\mathbf{D}^{(J)}(t,r)$ is given in, for example, Refs. 5, 6, 17, and 18). Specifically, the matrices $\mathbf{D}^{(1/2)}\equiv\mathbf{D}$ are identical to the Jones matrices (3.3). The matrices $\mathbf{D}^{(1)}$ realize the three-dimensional unitary representation of the group $SU(2)$ in the space of two-photon states H_1 (while the real Mueller matrices **R** realize the representation of the group $SO(3)$ in $\mathbf{R}^3$).

Converters with basis vectors $|M\rangle$ can also be made to act on the polarization vector **e**. According to Eqs. (3.17) and (3.20)

$$|\psi\rangle_J'=U|\psi\rangle_J=\sum e_M'|M\rangle,$$

where $\mathbf{e}'=\mathbf{D}^{(J)}\mathbf{e}$. The matrices $\mathbf{D}^{(J)}$ are determined by three parameters, so that for $N=2J\geqslant 2$ the set $\mathbf{S}^{2N}$ is not a homogeneous space relative to the polarization transformations, i.e., these transformations cannot be used to form an arbitrary polarization state in $\mathbf{H}_J$. (A realistic method for preparing two-photon states with arbitrary polarization, i.e., arbitrary vectors in $\mathbf{H}_1$ by means of parametric scattering is examined in Ref. 20.)

4. Let the state $|M=-J\rangle$ be prepared at the entrance of the optical system. Under the action of all possible transformations it is possible to obtain at the exit a set of states $|\phi,\theta,\psi;-J\rangle\equiv U(\phi,\theta,\psi)|-J\rangle$ (orbit of the element $|-J\rangle$). This set $|\phi,\theta,\psi;-J\rangle$ forms a homogeneous subspace relative to the transformations parameterized by three Euler angles (if the phase of the state is neglected, then two angles are sufficient; see Sec. 8). The states $|\phi,\theta,\psi;-J\rangle$ are called generalized coherent states for the $SU(2)$ group.[14–16] They form a continuous basis in $\mathbf{H}_J$—analogously to the Glauber coherence states $|\alpha\rangle$ for the group $W(1)$. Therefore any state $|\psi\rangle_J$ can be represented as a continuous superposition of coherent states $|\phi,\theta,\psi;-J\rangle$.

4. OBSERVABLES

Let us examine the relation between the quantities introduced above and experiment. The operators n_x and n_y can be observed with the aid of a Nicol prism and two photodetectors which measure the intensities of the fields in the modes x and y (Fig. 1). Our linear basis (x,y) is tied to the optic axes of the Nicol prism, but in the general case the indices x and y can refer to waves with right- and left-hand circular polarization or to any other pair of orthogonal waves. The sum and difference of the detector currents in the absence of converters is proportional to the parameters s_0 and s_1 of the initial light. Under the action of the converter the total intensity s_0 is redistributed between the two output modes. For example, let a $\lambda/4$ plate be placed in front of the Nicol prism so that the axis of the plate makes an angle $\chi=45°$ with the direction x. Then, according to Eqs. (3.2), (3.9), (3.11), and (3.13), we have $\mathbf{R}=\mathbf{R}_2(\pi/2)$, i.e., the initial Stokes vector

s rotates by 90° around the axis *2*. Now the difference of the indications of the detectors is proportional to the parameter s_3: $s_1'=s_3$. A phase plate with $\chi=0$ gives a rotation $\mathbf{R}_1(\pi/2)$. The combined action of two plates is described by the matrix $\mathbf{R}=\mathbf{R}_1(\pi/2)\mathbf{R}_2(\pi/2)$. In this case, $s_1'=s_2$ is observed (the same result is obtained by rotating the Nicol prism together with the detectors by 45° around the incident light beam). Therefore the operators s_0, s_1, s_2, and s_3 correspond to actual observable quantities. In the general case of arbitrary converters the observed dependence of the photocurrents is described by the operator

$$s_1'\equiv\sum_{k=1}^{3} R_{1k}s_k\,.$$

Therefore, by synthesizing a definite converter, it is possible to observe the projection

$$s_{\mathbf{x}}=(\mathbf{s}\mathbf{X})=\sum s_kX_k$$

of the operator Stokes vector $\mathbf{s}$ on any prescribed direction $\mathbf{X}=\Sigma X_k\mathbf{e}_k$ in the Stokes–Poincaré space (here $\Sigma X_k^2=1$).

The average intensity in one output channel has the form, according to Eqs. (3.5) and (3.7),

$$N_x'=tt^*N_x+rr^*N_y+t^*rS_++r^*tS_-=N_x\cos^2(\theta/2)+N_y\sin^2(\theta/2)-|S_+|\sin\theta\cos\psi', \tag{4.1}$$

where $\psi'\equiv\psi-\arg S_+$. Under a continuous variation of some parameter of the converter, for example, a variation produced by rotating a $\lambda/4$ phase plate, the observed intensity varies periodically: $N_x'\propto 1\pm V\cos 2\chi$, $0\leqslant V\leqslant 1$. The parameter V can be called the visibility factor of the polarization interference. There arises the question: What is the maximum possible value of V for a given light source? We shall interpret Eq. (4.1) as a quadratic function with matrix $\mathbf{K}$ of a complex vector (t,r) with unit norm, $|t|^2+|r|^2=1$. As is well known, the extremal values of a hermitian quadratic function are determined by the minimum $\lambda_{\min}$ and maximum $\lambda_{\max}$ eigenvalues of its matrix, i.e., by the solutions of the characteristic equation

$$f(\lambda)\equiv\det(\mathbf{K}-\lambda\mathbf{I})=\lambda^2-\mathrm{Tr}\,\mathbf{K}\lambda+\det\mathbf{K}=0. \tag{4.2}$$

Hence we find the extremal values

$$(N_x')_{\max,\min}=\lambda_{\max,\min}=\frac{1}{2}\{\mathrm{Tr}\,\mathbf{K}\pm[(\mathrm{Tr}\,\mathbf{K})^2-4\det\mathbf{K}]^{1/2}\} \tag{4.3}$$

(the same values can also be found directly by differentiating the expression (4.1) with respect to ψ and θ). The maximum possible interference visibility assumes the form

$$V_{\max}=\frac{\lambda_{\max}-\lambda_{\min}}{\lambda_{\max}+\lambda_{\min}}=\sqrt{1-\frac{4\det\mathbf{K}}{(\mathrm{Tr}\,\mathbf{K})^2}}=\frac{S}{S_0}. \tag{4.4}$$

Therefore $V_{\max}$ is equal to the degree of polarization P_2, so that the latter quantity has a direct operational meaning. It is easy to show that the correlation coefficient $\gamma\equiv S_+/(N_xN_y)^{1/2}$ does not exceed P_2 in magnitude. This gives another formal definition of the degree of polarization as the maximum value $|\gamma|_{\max}$ of the correlation coefficient of the mode amplitudes.

The invariant (''isotropic'' or ''unpolarized'') part can be separated from the matrix $\mathbf{K}$:

$$\mathbf{K}=\lambda_{\min}\mathbf{I}+\mathbf{K}_0\,.$$

Then we have $\det\mathbf{K}_0=0$, i.e., the matrix $\mathbf{K}_0$ is degenerate—its rank is less than 2 and it has only one eigenvalue $\lambda_{\max}$. The condition of total polarization $P_2=1$ (or $\lambda_{\min}=0$, $\mathbf{K}=\mathbf{K}_0$) means, from the graphite classical point of view, that the tip of the field vector $\mathbf{E}_0$ describes in the transverse plane of the wave an ellipse with constant parameters. Therefore there exists a definite basis in which one of the polarization components equals zero. Then $S_+'=0$ and the Stokes vector $\mathbf{S}'$ is parallel to the axis *1* (this follows formally from the fact that the hermitian matrix $\mathbf{K}$ can be put into a diagonal form by a canonical transformation). In the quantum theory, in this basis, one of the modes is in the vacuum state: $|\psi\rangle=|\psi_1\rangle_1|\mathrm{vac}\rangle_2$. From the operational standpoint, the condition $P_2=1$ means that there exists a converter such that the average photocurrent of one of the detectors vanishes:

$$N_x'=N,\quad N_y'=0.$$

Let us now examine experiments described by the fourth moments of the mode amplitudes. Three types of such quantities can be recorded in the arrangement shown in Fig. 1: (1) fluctuations of the indications of the detectors, (2) the correlation between the indications of two detectors, and (3) fluctuations of the difference of the indications of the detectors. These quantities are described by the following parameters of the field at the exit of the Nicol prism:

$$\Delta N_x'^2\equiv\langle\Delta n_x'^2\rangle=G_{xx}'-N_x'^2+N_x'\,.$$

$$\Delta N_y'^2\equiv\langle\Delta n_y'^2\rangle=G_{yy}'-N_y'^2+N_y'\,.$$

$$G_{xy}'\equiv\langle n_x'n_y'\rangle,\quad \Delta S_1'^2\equiv\langle\Delta s_1'^2\rangle\equiv\langle s_1'^2\rangle-S_1'^2 \tag{4.5}$$

(here $G_{\alpha\beta}\equiv\langle:n_\alpha n_\beta:\rangle=\langle a_\alpha^+a_\beta^+a_\alpha a_\beta\rangle$).

Let us examine the fluctuations $\Delta S_{\mathbf{X}}$ of the vector $\mathbf{S}$ along a definite direction $\mathbf{X}$ with the coordinates (Θ,Φ) in the Stokes–Poincaré space. Let $\Delta\mathbf{s}\equiv\mathbf{s}-\mathbf{S}$. Then

$$\Delta s_{\mathbf{x}}\equiv(\mathbf{X}\Delta\mathbf{s})=\sum_{k=1}^{3}\Delta s_kX_k\,.$$

Hence

$$(\Delta S_{\mathbf{x}})^2=\sum_{k,l=1}^{3}\langle\Delta s_k\Delta s_l\rangle X_kX_l\equiv\sum_{k,l=1}^{3}\Delta Q_{kl}X_kX_l\,. \tag{4.6}$$

(Here we have introduced the matrix $\Delta\mathbf{Q}\equiv\langle\Delta\mathbf{s}\otimes\Delta\mathbf{s}\rangle$). To measure $\Delta S_{\mathbf{X}}$ it is necessary to use a converter with Mueller matrix $\mathbf{R}$, which rotates $\mathbf{X}$ in the direction of the axis *1*: $\mathbf{R}\mathbf{X}=\mathbf{e}_1$, i.e., $\mathbf{X}_k=(\mathbf{R}^{-1})_{k1}=R_{1k}$. From Eq. (3.10) we find the required converter parameters: $\theta=-\Theta$, $\psi=-\Phi$ (see Fig. 3). Therefore the fluctuations of the Stokes vector $\mathbf{s}$ in any direction $\mathbf{X}$ of the Stokes–Poincaré space can be measured by adjusting the parameters of the converter placed in front of the Nicol prism.

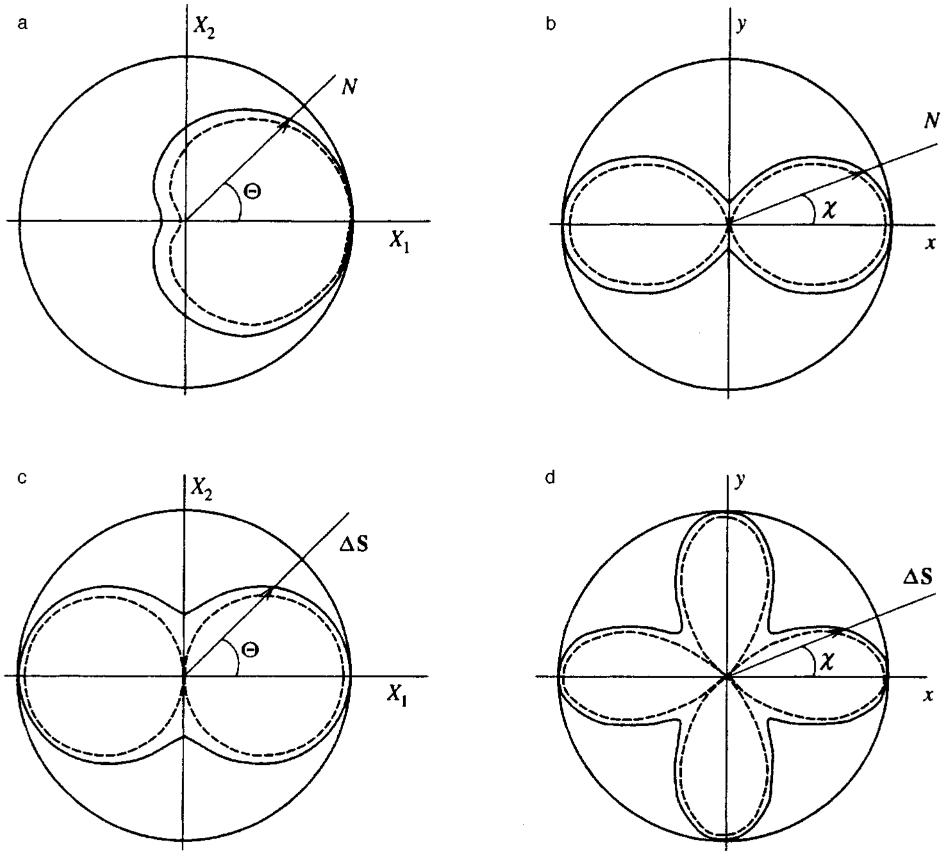


FIG. 5. Intensities $N\equiv N_x$ (Figs. a,b) and uncertainties $\Delta\mathbf{S}$ of the Stokes vector (c,d) versus the angle Θ in the Stokes-Poincaré space (a,c) and orientation of the Nicol prism $\chi=\Theta/2$ in the laboratory coordinate system (b,d). The visibility V equals 0.75 (solid lines) and 1 (dashed lines).

Typical observed dependences are shown in Fig. 5c. The form of the curve $\Delta S(\Theta)\propto\sqrt{1+V\cos 2\Theta}$ is far from an ellipse, but in standard parlance one talks about ''ellipses'' or ''ellipsoids'' of uncertainty (ellipsoids are engendered in the Stokes–Poincaré space by the condition $\Delta S(\Theta,\Phi)=\text{const}$). According to the definition given below, a field is completely polarized in fourth order ($P_4=1$) if the fluctuations $\Delta S_{\mathbf{X}}$ vanish in at least one direction $\mathbf{X}$ (dotted curve in Fig. 5c).

Let the entrance modes be excited by two independent lasers with orthogonal polarizations and the same average intensities ($N_x=N_y=(1/2)N$). Then $G'_{xy}\propto 1+(1/3)\cos 2\Theta$ and $\Delta N_x'^2=\Delta N_y'^2\propto 1-V\cos 2\Theta$, where $V=N/(N+8)$ (see Sec. 7). The corresponding plots, illustrating the ''antiphase'' character of these dependences and the invariance of the sum $\Delta N_x^2+\Delta N_y^2+2G_{xy}$ (see Eq. (6.1) below), are presented in Fig. 6.

5. FOURTH-ORDER POLARIZATION PARAMETERS

Let us examine the stationary fourth moments, i.e., the moments with the same number of positive- and negative-frequency amplitudes. An ordered set of such moments can be obtained using the direct product $\mathbf{k}\otimes\mathbf{k}$. After normal-ordering, averaging, and crossing out the redundant rows and columns, we obtain the following matrix of fourth moments:

$$\mathbf{K}_4\equiv\begin{pmatrix} A & D & E \\ D^* & C & F \\ E^* & F^* & B \end{pmatrix}. \tag{5.1}$$

Here

$$A\equiv\langle\overline{a^2}a^2\rangle,\quad B\equiv\langle\overline{b^2}b^2\rangle,\quad C\equiv\langle\overline{a}\,\overline{b}ab\rangle,$$

$$D\equiv\langle\overline{a^2}ab\rangle,\quad E\equiv\langle\overline{a^2}b^2\rangle,\quad F\equiv\langle\overline{a}\,\overline{b}b^2\rangle. \tag{5.2}$$

The diagonal components $A=G_{xx}$, $B=G_{yy}$, and $C=G_{xy}$ characterize the classical parts of the fluctuations and the correlation of the intensities in the polarization modes. The total number of independent real parameters determining the hermitian matrix $\mathbf{K}_4$ is equal to nine.

It is convenient to define also the real symmetric matrix $\mathbf{Q}\equiv\langle:\mathbf{s}\otimes\mathbf{s}:\rangle$ consisting of the normal-ordered second moments of the Stokes operators:

$$Q_{mn}\equiv\langle:s_m s_n:\rangle=Q_{nm}\quad(m,n=0,1,2,3). \tag{5.3}$$

We find the following relations ($k=1,2,3$) with the aid of Eqs. (2.1) and (2.3):

$$\langle s_m^2\rangle=Q_{mm}+S_0,$$

$$\langle s_0 s_m\rangle=\langle s_m s_0\rangle=Q_{0m}+S_m,$$

$$\langle s_k s_{k'}\rangle=Q_{kk'}+i\varepsilon_{kk'k''}S_{k''}\quad(k\neq k'),$$

$$\langle s_k s_{k'}+s_k s_{k'}\rangle=2Q_{kk'},$$

$$Q_{00}=Q_{11}+Q_{22}+Q_{33}. \tag{5.4}$$

In accordance with Eq. (2.10), $\langle s^2\rangle=Q_{00}+3S_0$. The linear terms here—the components of the Stokes vector

S_m—appear because the field operators do not commute, and this is the only formal difference between the quantum and the classical description. The matrix $\mathbf{Q}$ can be expressed in terms of the components of the polarization matrix $\mathbf{K}_4$:

$$\mathbf{Q}=2\begin{bmatrix}(A+B)/2+C & (A-B)/2 & \mathrm{Re}(D+F) & \mathrm{Im}(D+F)\\ (A-B)/2 & (A+B)/2-C & \mathrm{Re}(D-F) & \mathrm{Im}(D-F)\\ \mathrm{Re}(D+F) & \mathrm{Re}(D-F) & C+\mathrm{Re}\,E & \mathrm{Im}\,E\\ \mathrm{Im}(D+F) & \mathrm{Im}(D-F) & \mathrm{Im}\,E & C-\mathrm{Re}\,E\end{bmatrix}. \tag{5.5}$$

The inverse transformations have the form

$$\begin{aligned}&4A=Q_{00}+Q_{11}+2Q_{01},\\&4B=Q_{00}+Q_{11}-2Q_{01},\quad 4C=Q_{00}-Q_{11},\\&4D=Q_{02}+Q_{12}+i(Q_{03}+Q_{12}),\\&4F=Q_{02}-Q_{12}+i(Q_{03}-Q_{12}),\\&4E=Q_{22}-Q_{33}+2iQ_{23}.\end{aligned} \tag{5.6}$$

Hence, the polarization properties of the waves in fourth order are determined in terms of nine real parameters—the components of the normal-ordered matrices $\mathbf{K}_4$ or $\mathbf{Q}$ (the quantum part of the fluctuations is determined according to Eq. (5.4) by the total energy S_0). These parameters can be associated with a corresponding point in $\mathbf{R}^9$. Under the action of converters the point moves along a trajectory in this space.

It is convenient to determine also the matrix

$$\Delta\mathbf{Q}\equiv\mathbf{Q}-\mathbf{S}\otimes\mathbf{S}+\mathbf{S}_0\mathbf{I},$$

in terms of which the variances of the Stokes parameters are simply expressed ($m=0,1,2,3$):

$$\begin{aligned}&\langle\Delta s_m\Delta s_n\rangle=\Delta Q_{mn}=Q_{mn}-S_mS_n+S_0\delta_{mn},\\&\Delta S_m^2=\Delta Q_{mm}=Q_{mm}-S_m^2+S_0.\end{aligned} \tag{5.7}$$

The term $Q_{mm}-S_m^2$ describes the classical (''excess'') noise, for example, thermal noise, and S_0 describes the quantum noise associated with the noncommutativity of the algebra of observables. According to Eq. (2.6), the variances are related by

$$\sum_{k=1}^{3}\Delta S_k^2=\Delta S_0^2+2S_0+S_0^2(1-P_2^2). \tag{5.8}$$

There are a number of restrictions on polarization squeezing. For example, it follows from Eq. (5.8) that

$$\sum_{k=1}^{3}\Delta S_k^2\geqslant 2S_0. \tag{5.9}$$

Here equality obtains only for some subset of N-photon states giving $\Delta S_0=0$ and $P_2=1$. These states are called generalized coherent states for the group $SU(2)$ (see Sec. 8). Furthermore, there are the uncertainty relations

$$\Delta S_1\Delta S_2\geqslant|S_3|,\quad \Delta S_2\Delta S_3\geqslant|S_1|,\quad \Delta S_3\Delta S_1\geqslant|S_2|. \tag{5.10}$$

Hence $\Delta S_1\Delta S_2\Delta S_3\geqslant|S_1S_2S_3|^{1/2}$.

6. TRANSFORMATION OF THE FOURTH-ORDER PARAMETERS

Let us examine the effect of converters on the fourth moments. Recall that converters conserve total energy $s_0=n_x+n_y$. Therefore all moments $\langle s_0^p\rangle$ are invariants. Specifically, the quantities

$$\begin{aligned}&\langle :s_0^2:\rangle\equiv Q_{00}=A+B+2C.\\&\Delta S_0^2=\Delta N_x^2+\Delta N_y^2+2\langle\Delta n_x\Delta n_y\rangle=Q_{00}+S_0-S_0^2\end{aligned} \tag{6.1}$$

are conserved. The invariance of ΔS_0 can be called the law of conservation of the sum of the intensity fluctuations and correlations.

The effect of converters on the polarization matrix $\mathbf{K}_4$ can be found directly with the aid of Eq. (3.1). For example, we obtain from Eqs. (3.1) and (5.2)

$$\begin{aligned}&A'=T^2A+R^2B+4TRC+2\,\mathrm{Re}[tr^*(2TD+2RF\\&\qquad+tr^*E)],\\&B'=R^2A+T^2B+4TRC-2\,\mathrm{Re}[tr^*(2RD+2TF\\&\qquad-tr^*E)],\\&C'=TR(A+B)+(T-R)^2C+2\,\mathrm{Re}\{tr^*[(T-R)(F\\&\qquad-D)-tr^*E]\}.\end{aligned} \tag{6.2}$$

The component $A'\equiv G'_{xx}\equiv\langle :n_x'^2:\rangle$ determines the variance of the intensity fluctuations in the exit channel x:

$$\Delta N_x'^2=A'+N_x'-N_x'^2. \tag{6.3}$$

Here, according to Eq. (3.7),

$$N_x'=TN_x+RN_y+2\,\mathrm{Re}(tr^*S_+).$$

The quantities $\Delta N_y'^2$ and B' are related similarly.

Let $D=E=F=0$. Then the matrices K_4 and $\mathbf{Q}$ are diagonal in the initial basis:

$$\begin{aligned}&\mathbf{K}_4=\mathrm{diag}(A,C,B),\\&\mathbf{Q}=\mathrm{diag}(A+B+2C,A+B-2C,2C,2C).\end{aligned} \tag{6.4}$$

According to Eqs. (3.5) and (6.2)

$$A'=\frac{1}{8}\,[3A+3B+4C+4(A-B)\cos\,\theta+(A+B$$

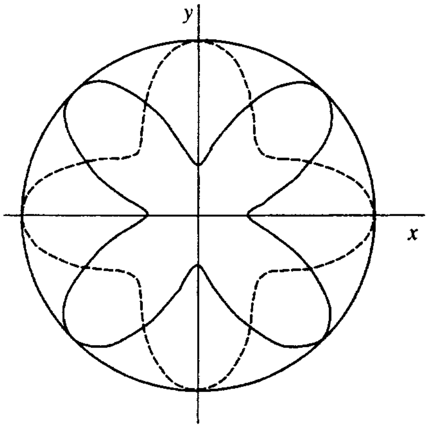


FIG. 6. Intensity fluctuations ΔN_x^2 (solid line) and correlation of intensities $G_{xy}\equiv\langle n_x n_y\rangle$ (dashed curve) at the exit of a Nicol prism as a function of the orientation of the prism in the case when the entrance modes are excited by two independent lasers with orthogonal polarizations and identical intensities, equal to 10 photons per mode (the visibility is $V_x=5/9$).

$$-4C)\cos 2\theta],$$

$$B'=\frac{1}{8}[3A+3B+4C-4(A-B)\cos\theta+(A+B-4C)\cos 2\theta],$$

$$C'=\frac{1}{8}(A+B+4C)[1-V_{xy}\cos 2\theta]. \tag{6.5}$$

Here we have introduced the visibility

$$V_{xy}\equiv\frac{A+B-4C}{A+B+4C}. \tag{6.6}$$

Therefore when the parameters of the converters are modulated, for example, by rotating the phase plate, the observed correlations C' and variances $\Delta N_x'^2$, $\Delta N_y'^2$, and ΔS_1^2 vary periodically (see Figs. 5 and 6). This effect can be defined as fourth-order polarization interference; this is an analog of intensity interference.

Switching to the classical theory, the normal-ordered moment

$$Q_{11}\equiv\langle :s_1^2:\rangle=A+B-2C$$

is the moment $\langle\langle s_1^2\rangle\rangle$ (here $\langle\langle\ldots\rangle\rangle$ denotes a classical average; see Refs. 21 and 22 for details). The latter moment cannot assume negative values. Therefore $(A+B-2C)_{\text{clas}}\geqslant 0$ and similarly $(A+B-2|E|)_{\text{clas}}\geqslant 0$. As a result, in the classical theory limits on the visibility arise, which can be violated in experiments with ''nonclassical'' light.

The matrix $\mathbf{Q}$ has simpler transformation properties than $\mathbf{K}_4$, since s_0 is an invariant and the action of the converters on the three-dimensional operator Stokes vector $\mathbf{s}$ reduces to a rotation of the vector, $\mathbf{s}'=\mathbf{R}\mathbf{s}$. As a result,

$$Q_{00}'=Q_{00},\quad Q_{k0}'=\sum_{p=1}^{3}R_{kp}Q_{p0}, \tag{6.7}$$

$$Q_{kl}'=\sum_{p,q=1}^{3}R_{kp}R_{lq}Q_{pq}\quad (k,l=1,2,3). \tag{6.8}$$

Therefore the 3×3 matrix Q_{kl} changes according to the tensor representation $\mathbf{R}\otimes\mathbf{R}$ of the group $SO(3)$.

The matrix $\Delta\mathbf{Q}'\equiv\mathbf{Q}'-\mathbf{S}'\otimes\mathbf{S}'+S_0\mathbf{I}$ determines the change in the uncertainties ΔS_k of the Stokes parameters. For example, the fluctuations of the quantity s_1' which are observed in the experiment in Fig. 1 depend on the converter as follows (compare Eq. (4.6)):

$$\Delta S_1'^2=\sum_{k,l=1}^{3}R_{1k}R_{1l}\Delta Q_{kl}. \tag{6.9}$$

This expression is a quadratic function (fixed by the initial matrix $\Delta\mathbf{Q}$) of the vector $\mathbf{X}$ with components from the first row of the matrix $\mathbf{R}$: $X_k\equiv R_{1k}$. Measuring $\Delta S_1'$ with different converter polarizations, it is possible to construct the uncertainty region $\Delta S(\Theta,\Phi)$, clearly representing the anisotropy of the fluctuations of the Stokes vector in the Stokes–Poincaré space (see Fig. 7). The dimensions of the region along the principal axes equal $\sqrt{\mu_k}$, where μ_k are the eigenvalues of the matrix $\Delta\mathbf{Q}$. They are invariants of the polarization conversion which are determined by the properties of the light source (we note that in accordance with Eq. (5.10) $\mu_1\mu_2\geqslant S_3^2$ and so on).

As the parameters of the converter are varied, the value of $\Delta S_1'^2$ remains within the interval between $\mu_{\min}$ and $\mu_{\max}$—the minimum and maximum numbers from the set

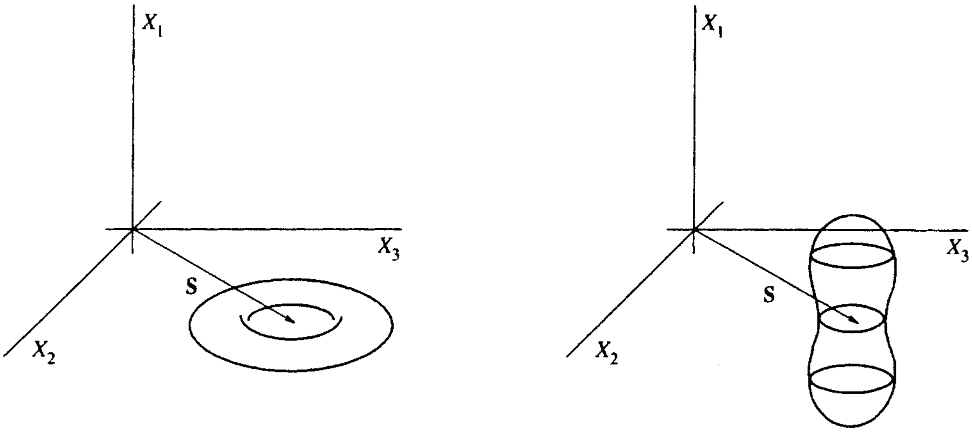


FIG. 7. Squeezing of fluctuations of the Stokes vector $\mathbf{S}$ by means of parametric conversion. On the left side the phase of the homodyne field is chosen so that the region of uncertainty is squeezed along the vertical and extended along the horizontal. On the right-hand side the phase of the homodyne field is additionally shifted by $\pi/2$, so that the region of uncertainty is squeezed in the horizontal direction and extended in the vertical direction.

$\{\mu_k\}$. We shall define the degree of fourth-order polarization P_4 by analogy with P_2—according to the maximum possible visibility of the fourth-order interference, observed according to the fluctuations of the Stokes parameter,

$$P_4 \equiv \frac{\mu_{\max} - \mu_{\min}}{\mu_{\max} + \mu_{\min}}. \tag{6.10}$$

The extremal values of other observable parameters can also be determined in terms of the eigenvalues of the matrices $\Delta\mathbf{Q}$ or $\mathbf{Q}$. For example, $G'_{xy} = C' = (1/4)(Q_{00} - Q'_{11})$, i.e., C' is also a quadratic form with the matrix $(1/4) \times (Q_{00}\mathbf{I} - \mathbf{Q})$, so that the extremal values of the correlation C' and visibility V_{xy} are determined by the eigenvalues η_k of the matrix $\mathbf{Q}$ (if $\mathbf{S}$ and $\mathbf{Q}$ are diagonal in the same basis, then $\mu_k = \eta_k - S_k^2 + S_0$). In the basis in which $\mathbf{Q}$ is diagonal, we have $D = F$ and $E = E^*$, so that the eigenvalues of the matrix $\mathbf{Q}$ have the form $\eta_k = A + B - 2C,\ 2(C \pm E)$.

The squeezed states in this case are defined in terms of the corresponding uncertainty relations (see Ref. 15). According to Eq. (5.10), this squeezing condition for the component s_1 has the form

$$\Delta S_1^2 < |S_3|. \tag{6.11}$$

However, this criterion is inconvenient in the case at hand, since it is non-invariant and becomes meaningless for $|S_3| = 0$. For this reason, we shall adopt a somewhat different definition: We shall term a state polarization-squeezed if at least for one direction of $\mathbf{X}$ in the Stokes–Poincaré space the following (invariant) condition is satisfied:

$$(\Delta S_{\mathbf{X}}^2)_{\min} < S_0. \tag{6.12}$$

Since $|S_k| \leqslant S_0$, this restriction is less severe than Eq. (6.11); it encompasses a larger set of states. It follows from the condition (6.12) that at least for one pair of orthogonally polarized modes (1,2), the variance of the energy difference between the modes $\langle[\Delta(n_1 - n_2)]^2\rangle$ is less than the Poisson value $\langle n_1\rangle + \langle n_2\rangle = S_0$. This signifies clearly that the fluctuations of the intensities in these modes—both quantum and classical—are correlated. The condition of complete squeezing $\mu_{\min} = 0$ gives $P_4 = 1$; then $\det(\Delta\mathbf{Q}) = 0$ (i.e., the rank of the matrix $\Delta\mathbf{Q}$ is less than 3) and the uncertainty ellipsoid for $\Delta S_{\mathbf{X}}^2$ degenerates into a section of a plane or straight line. Correspondingly, the operational meaning of the condition $P_4 = 1$ is that there exists a converter such that the observed fluctations of the difference of the photocurrents in the two detectors in Fig. 1 vanish, $\Delta S_1'^2 = 0$. We recall that the condition $P_2 = 1$ means that there exists a converter such that the observed photocurrent in one detector vanishes, $\langle N'_x\rangle = 0$. Let $\Delta S_1 = 0$, $\Delta S_2 \neq 0$, $\Delta S_3 \neq 0$. Then, according to the uncertainty relations (5.10), $S_2 = S_3 = 0$, i.e., the Stokes vector is directed along the axis *1* (or the magnitude of this vector equals zero; see Fig. 7) and fluctuates only in the transverse direction.

The definition (6.12) of polarization-squeezed states is actually identical to a criterion of nonclassical states for a two-mode light, which are based on a comparison of the quantum normal-ordered moments with classical moments; see Refs. 21 and 22. It follows from the inequality (6.12) that the matrix $\langle : s_m s_n : \rangle - S_m S_m$, describing the classical (excess) noise, is not positive-definite; one or two of its eigenvalues are negative—in contrast to the classical case. Therefore polarization-squeezed states belong to the set of nonclassical states. For clarity, squeezing can be regarded as resulting from anticorrelation of classical and quantum noise.[23] We emphasize in this connection that the separation of ΔS_k into classical and quantum components has a clear operational meaning—when the spectrum of fluctuations of the photocurrent is observed, the quantum fluctuations give frequency-independent ''white'' noise, while the excess-noise spectrum has an upper limit due to the dynamical properties of the light source or frequency filters in the optical channel (see Refs. 23 for details).

7. COHERENT STATE

The coherent states of the $W(2)$ group have the form $|\alpha,\beta\rangle \equiv |\alpha\rangle_x |\beta\rangle_y$, where α and β are arbitrary complex numbers. Here $\langle a\rangle = a$, $\langle\beta\rangle = b$ and the normal-ordered moments factorize:

$$N_x = |\alpha|^2, \quad N_y = |\beta|^2, \quad S_+ = \alpha^*\beta, \quad S_0 = |\alpha|^2 + |\beta|^2,$$
$$A = N_x^2, \quad B = N_y^2, \quad C = N_x N_y,$$
$$D = |\alpha|^2 \alpha^*\beta, \quad E = (\alpha^*\beta)^2, \quad F = |\beta|^2 \alpha^*\beta. \tag{7.1}$$

Hence $\mathbf{Q} = \mathbf{S} \otimes \mathbf{S}$ and $\Delta\mathbf{Q} = S_0\mathbf{I}$ hold, so that we have $\Delta S_k^2 = \mu_k = S_0$—the fluctuations of the vector $\mathbf{s}$ are isotropic in the Stokes–Poincaré space, the uncertainty region is a sphere. Then $P_2 = 1$ and $P_4 = 0$—the light is completely polarized in second order and unpolarized in fourth order.

We introduce the notation

$$\alpha \equiv \sqrt{N}\,\cos(\Theta/2)\exp(i\phi_x),$$
$$\beta \equiv \sqrt{N}\,\sin(\Theta/2)\exp(i\phi_y),$$
$$|\alpha|^2 + |\beta|^2 \equiv N, \quad \Phi \equiv \phi_y - \phi_x. \tag{7.2}$$

Then the Stokes parameters assume the form

$$S_1 = |\alpha|^2 - |\beta|^2 = N\cos\Theta,$$
$$S_2 = 2\,\mathrm{Re}(\alpha^*\beta) = N\sin\Theta\cos\Phi,$$
$$S_3 = 2\,\mathrm{Im}(\alpha^*\beta) = N\sin\Theta\sin\Phi. \tag{7.3}$$

Therefore S_2 and S_3 are proportional to the cosine and sine of the difference of the phases of the mode amplitudes. The angles Θ and Φ determine the spherical coordinates of the Stokes vector $\mathbf{S}$ in the Stokes–Poincaré space, if the axis *1* is taken as the polar axis.

The transition to the diagonal basis with $\mathbf{S}' = (N,0,0)$ is made by a transformation with the parameters determined from the condition

$$\beta' = -r\alpha + t\beta = 0.$$

We find from Eq. (3.6) the parameters of the converter (Euler angles) $\psi = \pi - \Phi$, $\theta = \Theta$, $\phi = 0$. This corresponds to a rotation of the vector $\mathbf{S}$ in the direction of the axis *1*; see Fig. 3.

We now consider the case when the modes x and y are discussed by two ideal lasers with the same frequencies and independently fluctuating phases ϕ_x and ϕ_y[5–8] (in the experiment of Ref. 8, one laser and a delay in one of the modes, much greater than the phase coherence time of the laser, were used). Then $P_2=0$ holds: the radiation is unpolarized in the ordinary sense. However, the photocurrent fluctuations in the scheme displayed in Fig. 1 will be periodic functions of the orientation χ of the Nicol prism:

$$\Delta N_x'^2 \propto 1-V_x \cos 4\chi.$$

For $\chi=0$ each detector ''sees'' one laser and the noise is of a Poisson character with variance N_x or N_y, and for $\chi=\pi/4$ additional noise is added on account of the transformation, by the Nicol prism, of the fluctuations of the phase difference into excess intensity fluctuations; see Fig. 6. The correlation $\langle n_x n_y\rangle$ of the photocurrents exhibits anisotropy similar to that observed in Ref. 8.

For a quantitative description of the effect, we assume that both lasers are in coherent states $|\alpha\rangle_x$ and $|\beta\rangle_y$, whose phases $\phi_x=\arg\alpha$ and $\phi_y=\arg\beta$ are classical random parameters with uniform independent distributions (i.e., the state of the field is mixed). Now $S_+=\langle\langle\alpha^*\beta\rangle\rangle=0$ ($\langle\langle\ldots\rangle\rangle$ denotes a classical average with respect to ϕ_x and ϕ_y). Then the parameters of the field have the form

$$S_1=N\cos\Theta,\quad S_2=S_3=0,\quad P_2=|\cos\Theta|,$$

$$Q_{10}=N_x^2-N_y^2,\quad A=N_x^2,\quad B=N_y^2,$$

$$C=N_xN_y=\frac{1}{4}N^2\sin^2\Theta,\quad \eta_0=N^2,$$

$$\eta_1=N^2\cos^2\Theta,\quad \eta_2=\eta_3=\frac{1}{2}N^2\sin^2\Theta,$$

$$\mu_0=\mu_1=N,\quad \mu_2=\mu_3=\eta_2+N. \tag{7.4}$$

Here μ_m and η_m are the eigenvalues of the matrices $\mathbf{\Delta Q}$ and $\mathbf{Q}$, respectively, and the parameter Θ is determined by the ratio of the laser intensities (see Eq. (7.2)): $\tan^2(\Theta/2)=N_y/N_x$. Using the relation (6.10) we find the degree of fourth-order polarization

$$P_4=\frac{\mu_2-\mu_1}{\mu_2+\mu_1}=\frac{N_xN_y}{N_x+N_y+N_xN_y}=\frac{N\sin^2\Theta}{4+N\sin^2\Theta}. \tag{7.5}$$

For equal laser intensities it is found that $P_2=0$ and $P_4=N/(N+4)$, i.e., the polarization is hidden; for $N\gg 4$ the contribution of the quantum noise can be neglected, so that $P_4=1$.

According to Eqs. (6.5) and (7.4), the intensity correlation depends on the converter parameter θ as

$$C'=\frac{1}{8}N^2\left(1+\frac{1}{2}\sin^2\Theta\right)[1+V_{xy}\cos 2\theta],$$

$$V_{xy}=\frac{3\sin^2\Theta-2}{\sin^2\Theta+2}. \tag{7.6}$$

Here the correlation of the intensity fluctuations is negative:

$$\langle\Delta n_x'\Delta n_y'\rangle=C'-N_x'N_y'=-\frac{1}{8}N^2\sin^2\Theta\sin^2\theta. \tag{7.6a}$$

Let $\sin^2\Theta=2/3$ ($N_y/N_x=\tan^2(\Theta/2)=0.27$ and $P_2=1/\sqrt{3}=0.577$). Then $V_{xy}=0$, i.e., $C'=C$—the intensity correlation is invariant. If, however, the intensities in the modes are the same ($\Theta=\pi/2$, $P_2=0$), then the visibility equals 1/3:

$$C'=\frac{3}{16}N^2\left[1+\frac{1}{3}\cos 2\theta\right].$$

For $\Theta\ll 1$ we obtain $V_{xy}=-1$.

According to Eqs. (3.7), 6.2), and (6.3), the intensity fluctuations at the converter exit have the form

$$\Delta N_x'^2=TN_x+RN_y+2TRN_xN_y. \tag{7.7}$$

The last term describes classical noise arising as a result of the conversion of the phase fluctuations into (anticorrelated) intensity fluctuations by a beam splitter, i.e., the scheme with the Nicol prism in Fig. 1 serves as a phase detector. The variance $\Delta N_y'^2$ is determined from Eq. (7.7) by interchanging the indices $x\leftrightarrow y$. The expression (7.7) can be represented in the form

$$\Delta N_x'^2=\frac{1}{2}N\{\cos^2[(\Theta+\theta)/2]+\cos^2[(\Theta-\theta)/2]\}$$

$$+\frac{1}{8}N^2\sin^2\Theta\sin^2\theta. \tag{7.8}$$

The quantum noise (proportional to N) is a periodic function of θ with period 2π, and the excess noise (proportional to N^2) is a periodic function of θ with period π. For equal intensities

$$\Delta N_x'^2=\frac{1}{2}N\left(1+\frac{1}{8}N\right)[1-V_x\cos 2\theta],\quad V_x=\frac{N}{N+8}. \tag{7.9}$$

The expressions (7.6a) and (7.7) satisfy the conservation law (6.1), according to which the sum of the detector noises $\Delta N_x'^2=\Delta N_y'^2$ and their cross correlation $\langle\Delta n_x'\Delta n_y'\rangle$ equals the total energy N. Therefore the beats of the detector noise and their correlation are in antiphase; see Fig. 6.

8. STATIONARY STATES

Let us find the polarization parameters of an N-photon state

$$|N_x\rangle_x|N_y\rangle=|JM\rangle\equiv|M\rangle$$

with a definite difference

$$S_1=N_x-N_y=2M$$

of the number of photons in the modes. Here

$$N_x=J+M,\quad N_y=J-M,\quad N=S_0=2J.$$

The operator $s_+=a^+b$ ''transfers photons'' from the y mode into the x mode, and $s_-=ab^+$ does the opposite. It follows from Eq. (2.1) that

$$s_\pm|M\rangle=[J(J+1)-M(M\pm 1)]^{1/2}|M\pm 1\rangle.$$

Hence $S_\pm=0$. The Stokes vector and the degree of second-order polarization have the form

$$S=(2J,2M,0,0),\quad P_2=|M|/J.$$

Therefore states with $M=\pm J$ (including the one-photon states $|\pm 1/2\rangle$) are completely polarized in second order; states with an even number of photons and $M=0$ are unpolarized; and all states with an odd number of photons are at least partially polarized.

The fourth-order polarization matrix (5.1) is diagonal $(D=E=F=0)$ and its elements are

$$\begin{aligned} &A=(J+M)(J+M-1)=N_x(N_x-1),\\ &B=(J-M)(J-M-1)=N_y(N_y-1),\\ &C=J^2-M^2=N_xN_y. \end{aligned} \tag{8.1}$$

Using Eq. (5.5) we find the following nonzero components of the matrix **Q**:

$$\begin{aligned} &Q_{00}=2J(2J-1)=N(N-1),\\ &Q_{10}=(J+M)(J+M-1)-(J-M)(J-M-1)\\ &\quad=S_1(S_0-1),\\ &Q_{11}=4M^2-2J=S_1^2-S_0.\\ &Q_{22}=Q_{33}=2(J^2-M^2)=2N_xN_y. \end{aligned} \tag{8.2}$$

According to Eq. (5.7), the variances $\Delta S_m^2=\mu_m$ of the Stokes parameters have the form

$$\mu_0=\mu_1=0, \quad \mu_2=\mu_3=2(J^2-M^2+J)=2N_xN_y+N. \tag{8.3}$$

Therefore for any M light is completely polarized in fourth order, $P_4=1$. The uncertainty $\Delta S'$ as a function of direction Θ varies from 0 to $\Delta S_2=\sqrt{\mu_2}$ (primes in Fig. 5c). In the case of the one-photon state we have $\mathbf{S}=(1,0,0)$ and $\Delta S_2=1$.

The states $|M\rangle$ give equality in the uncertainty relation: $\Delta S_1\Delta S_2=S_3=0$. Such states are said to be ''intelligent''; see Ref. 15. Furthermore, since both sides of the uncertainty relation take on their minimum values, the states $|M\rangle$ also refer to a class of states with minimum uncertainty. For N even and $M=0$ we have an example of hidden polarization with $P_2=0$ and $P_4=1$. We note that equality obtains in Eq. (5.9) for $M=\pm J$.

In the states $|M\rangle$ the quantum part S_0 of the fluctuations of the parameter S_1 is completely compensated by the excess noise $Q_{11}-S_1^2=-S_0$. In practice, such compensation can occur only in a frequency range of intensity fluctuations which has an upper limit; see Ref. 23.

Intnsity interference according to Eqs. (6.5) and (8.1) is described by the expressions

$$C'\propto 1+V_{xy}\cos 2\theta, \quad V_{xy}=\frac{J(J+1)-3M^2}{J(3J-1)-M^2}. \tag{8.4}$$

For example, for $N=2$ and $M=0$, i.e., in the case of a two-photon symmetric state $|1\rangle_x|1\rangle_y$, we obtain $V_{xy}=1$. For $T=R=1/2$, the correlation is completely suppressed, $C'=0$, i.e., both photons are observed only in the same exit channel. This is a photon anticorrelation effect,[5,6,24–27] which makes it possible to measure ultrashort group-delay times.

Now let $M=-J$, i.e., the x mode is in a vacuum state and the y mode is in a N-photon state, so that the field is completely polarized along y. Different transformers convert the field into a set of states of the form $U(\phi,\theta,\psi)|-J\rangle$, which are called generalized coherence states for the $SU(2)$ group (see Sec. 3). They can be parameterized by two Euler angles ϕ, θ and mapped onto the Poincaré sphere, setting $\theta=\Theta$, $\phi=\Phi$. Then, according to Eqs. (3.10) and (8.2),

$$\begin{aligned} &S_0=-S_1=N, \quad Q_{00}=-Q_{10}=Q_{11}=N(N-1),\\ &S_1'=-N\cos\Theta, \quad S_2'=-N\sin\Theta\cos\Phi,\\ &S_3'=-N\sin\Theta\sin\Phi. \end{aligned} \tag{8.5}$$

We have from Eq. (5.7)

$$\mu_1=0, \quad \mu_2=\mu_3=N$$

and from Eq. (6.9)

$$\Delta S_k'^2=N(1-R_{k1}^2).$$

Taking account of Eq. (3.10), we find the variances in the $SU(2)$-coherent states $U(\Phi,\Theta,\psi)|-J\rangle$:

$$\begin{aligned} &\Delta S_1'^2=N\sin^2\Theta.\\ &\Delta S_2'^2=N(1-\sin^2\Theta\cos^2\Phi),\\ &\Delta S_3'^2=N(1-\sin^2\Theta\sin^2\Phi). \end{aligned} \tag{8.6}$$

Equivalent results were obtained in Ref. 15 in the Schrödinger representation. We note that the sum of the three variances in Eq. (8.6) reaches its minimum possible value of $2N$ (see Eq. (5.9)); this is characteristic for coherent states.[14,16]

In all $SU(2)$ coherent states we have $\mu_1=\Delta S_1^2=0$, i.e., according to our definition, total polarization squeezing, $P_4=1$, obtains. At the same time, the standard condition for squeezed states of the type $\Delta S_2'^2\leqslant|S_1'|$,[15] which assumes, according to Eq. (8.6), the form

$$1-\sin^2\Theta\cos^2\Phi<|\cos\Theta|,$$

holds only for some bounded region on the Poincaré sphere, and therefore it is not a polarization invariant.

9. SQUEEZED VACUUM

In spontaneous type-II parametric scattering, photons are emitted only in pairs, one photon in each pair having x polarization and the other y polarization. We are interested in the degenerate case, when the modes have the same frequencies, equal to half the frequency of the pump field. The corresponding state of the field is said to be a two-mode squeezed vacuum. It is engendered from the vacuum by a Hamiltonian of the form

$$H\propto a^+b^++ab.$$

This operator, which belongs to the $su(1.1)$ algebra,[15,16] commutes with the Stokes operator $s_1=n_x-n_y$. As a result, the observable s_1 does not fluctuate, $\Delta S_1=0$, and the intensities of the field in each mode (which themselves undergo chaotic fluctuations) are completely correlated, $n_x=n_y$. Therefore the state is completely polarization-squeezed and $P_4=1$.

The Hamiltonian H creates from the vacuum a state with Gaussian statistics, which is completely determined by the second moments:

$$\langle a^+a\rangle \equiv N_x = N_y = \sinh^2\Gamma, \quad S_+ = \langle a^+b\rangle = 0,$$

$$\begin{aligned}\langle ab\rangle \equiv L &= [N_x(N_x+1)]^{1/2}\exp(i\phi_0)\\ &= \frac{1}{2}\sinh 2\Gamma\,\exp(i\phi_0)\end{aligned} \tag{9.1}$$

(Γ is the exponent of the parametric amplification, ϕ_0 is the pump phase plus $\pi/2$). Therefore $S_1=S_2=S_3=0$—the field is unpolarized in second order, $P_2=0$.

The matrices $\mathbf{K}_4$, $\mathbf{Q}$, and $\Delta\mathbf{Q}$ are diagonal and their components can be expressed in terms of the total energy $N\equiv 2\sinh^2\Gamma$ as follows:

$$A=B=\frac{1}{2}N^2, \quad C=N_x^2+|L|^2=\frac{1}{2}N(N+1).$$

$$\eta_0=N(2N+1), \quad \eta_1=-N, \quad \eta_2=\eta_3=N(N+1),$$

$$\mu_1=0, \quad \mu_0=\mu_2=\mu_3=N(N+2)=\sinh^2 2\Gamma. \tag{9.2}$$

This is an another example of a field with hidden polarization: $P_2=0$, $P_4=1$. The region of uncertainty for ΔS has the diameter $\sqrt{\mu_2}=\sinh 2\Gamma$ in the (*2,3*) plane and zero thickness along the *1* axis (dashes in Fig. 5c). It follows from Eqs. (6.9) and (3.10) that

$$\Delta S_1'^2=\frac{1}{2}N(N+2)[1-\cos 2\theta]=\sinh^2 2\Gamma\,\sin^2\theta. \tag{9.3}$$

From Eqs. (6.3) and (6.5) we find the form of the interference relations for the fluctuations of the intensity in one channel and for the correlation of the intensities:

$$\Delta N_x'^2=\frac{3}{8}N(N+2)\left[1-\frac{1}{3}\cos 2\theta\right],$$

$$C'=\frac{1}{8}N(3N+2)[1+V_{xy}\cos 2\theta],$$

$$V_{xy}=\frac{N+2}{3N+2}. \tag{9.4}$$

Now let a coherent component with complex parameters $\langle a\rangle=\alpha$ and $\langle b\rangle=\beta$, which plays the role of a homodyne field, be added to the squeezed vacuum. In practice this can be done with beam splitters with transmission close to 1. The field is then described by the following normal-ordered characteristic function (see Refs. 21 and 22):

$$\begin{aligned}\chi(x,x^*,y,y^*)=\exp[&-N_x(xx^*+yy^*)+L^*xy+Lx^*y^*\\ &+\alpha^*x-\alpha x^*+\beta^*y-\beta y^*].\end{aligned} \tag{9.5}$$

Here the quantity L is defined in Eq. (9.1) and $N_x=\sinh^2\Gamma\equiv N/2$ refers only to the intensity of the incoherent part of the field. From Eq. (9.5) we find by differentiating the second and fourth moments

$$\langle a^+a\rangle=N_x+|\alpha|^2, \quad \langle b^+b\rangle=N_x+|\beta|^2,$$

$$\langle a^+b\rangle\equiv S_+=\alpha^*\beta, \quad \langle ab\rangle=\alpha\beta+L,$$

$$A=\frac{1}{2}N^2+2N|\alpha|^2+|\alpha|^4, \quad B=\frac{1}{2}N^2+2N|\beta|^2+|\beta|^4,$$

$$C=\frac{1}{2}N(1+N+N_c)+|\alpha\beta|^2+2\,\mathrm{Re}(L^*\alpha\beta),$$

$$D=\alpha^{*2}\alpha\beta+N\alpha^*\beta+L\alpha^{*2}, \quad E=(\alpha^*\beta)^2,$$

$$F=\alpha^*\beta^*\beta^2+N\alpha^*\beta+L^*\beta^2. \tag{9.6}$$

Here $N_c\equiv|\alpha|^2+|\beta|^2$, $N\equiv 2N_x=2\sinh^2\Gamma$, so that $S_0=N+N_c$. The Stokes vector is determined only by the coherent part of the field, so that $P_2=N_c/S_0$.

The variances of the components of the Stokes vector, according to Eqs. (5.5), (5.7), and (7.3), have the form

$$\begin{aligned}\Delta S_1^2&=N_c(N+1)-4\,\mathrm{Re}(L^*\alpha\beta)\\ &=N_c[\cosh 2\Gamma-\sinh 2\Gamma\,\sin\Theta\,\cos\phi],\end{aligned}$$

$$\begin{aligned}\Delta S_k^2&=N(N+2)+N_c(N+1)+4\,\mathrm{Re}(L^*\alpha\beta)=\sinh^2 2\Gamma\\ &\quad+N_c[\cosh 2\Gamma+\sinh 2\Gamma\,\sin\Theta\,\cos\phi]\end{aligned}$$

$$(k=0,2,3). \tag{9.7}$$

Here $\phi\equiv\phi_x+\phi_y-\phi_0$ is the relative phase of the homodyne field and $\tan\Theta/2\equiv|\beta/\alpha|^2$. We shall examine two limiting cases.

1. Let $\phi=0$ and $\Theta=\pi/2$. Then ΔS_1 undergoes maximum squeezing:

$$\Delta S_1^2=N_c e^{-2\Gamma} \quad (\phi=0,\ \Theta=\pi/2). \tag{9.8}$$

Here two types of suppression of fluctuations are manifested simultaneously.[28]

(a) The Gaussian fluctuations of the parametric-scattering field are completely suppressed as a result of the total correlation of the intensities in the two modes ($n_x=n_y$), irrespective of the value of N_c.

(b) The Poisson fluctuations of the homodyne field are suppressed on account of the presence of an ‘‘anomalous’’ moment $L=\langle ab\rangle$—similarly to the case of one-mode squeezed light. In classical terms, this is evidence of phase anticorrelation of the modes, $\phi_x+\phi_y=\phi_0$. The necessary conditions for appreciable squeezing of this type are $N_c\gg N\gg 1$.

According to Eq. (9.7)

$$\Delta S_k^2=\sinh^2 2\Gamma+N_c e^{2\Gamma} \quad (k=0,2,3,\ \phi=0,\ \Theta=\pi/2). \tag{9.9}$$

Here the Gaussian fluctuations are fully manifested and the Poisson fluctuations are increased by a factor of e^{Γ}. Therefore the region of uncertainty is flattened along the axis *1* and extended in the directions *2* and *3* (Fig. 7a). By adjusting the difference $\phi_x-\phi_y$ (leaving the phase $\phi=\phi_x+\phi_y-\phi_0=0$ unchanged) the center of the region, placed at the point $\mathbf{S}$, can be displaced in the plane (*2,3*), obtaining, for example, squeezed linear (at angle 45°) or squeezed circular polarization.

2. Let $\Theta=\pi/2$ and $\phi=\pi$. Then the uncertainty region, according to Eq. (9.7), is squeezed in the directions *2* and *3* and extended in the direction *1* (Fig. 7b):

$$\Delta S_1^2 = N_c e^{2\Gamma},$$

$$\Delta S_k^2 = \sinh^2 2\Gamma + N_c e^{-2\Gamma} \quad (k=0,2,3). \tag{9.10}$$

Now the contribution of Gaussian fluctuations to the uncertainties ΔS_0, ΔS_2, and ΔS_3 is not suppressed, and the necessary conditions for large relative squeezing are $N_c \gg N \gg 1$.

For unsymmetric displacement, $\Theta \neq \pi/2$, $|\alpha| \neq |\beta|$ and the vector **S** does not lie in the (*2,3*) plane and squeezing diminishes.

10. CONCLUSIONS

Thus, the polarization properties of a monochromatic plane wave in fourth order are all determined by nine parameters, for example, the components of the 4×4 symmetric matrices $\Delta Q_{mn} = \langle \Delta s_m \Delta s_n \rangle$. The fluctuations $\Delta S_{\mathbf{X}}$ of the Stokes vector in different directions can be determined by six parameters—three eigenvalues μ_k of the 3×3 submatrix ΔQ_{kl}, which determine the dimensions of the uncertainty ellipsoid $\Delta S_{\mathbf{X}} = \text{const}$, and three angles, which determine the orientation of the ellipsoid in the Stokes–Poincaré space. The effect of polarization converters is graphically represented as a rotation of the ellipsoid (whose center can be placed at the tip of the Stokes vector). In the case of classical light we have for all directions $\Delta S_{\mathbf{X}} \geqslant N$, where N is the average number of photons in two polarization modes, which determines the quantum part of the fluctuations of the Stokes vector. In polarization-squeezed light,[10] we have $\Delta S_{\mathbf{X}} \leqslant N$ for some directions.

This formalism can also be used to describe two modes which differ not by polarization but rather by the propagation direction. The analog of the Stokes vector here makes it possible to visualize different states of two modes and their conversion by beam splitters and phase shifters.[3] The uncertainty ellipsoid for the Stokes vector (Fig. 7) gives a clear three-dimensional representation of two-mode squeezing, similar to the well-known two-dimensional diagrams with ellipses for one-mode squeezing.

In conclusion, we emphasize again that the polarization symmetry of a free light field is not specific to quantum theory. This becomes obvious in the Heisenberg representation, in which the transformations of the polarization are described by phenomenological Jones or Mueller matrices which are well-known in classical optics.

Here we examined only fourth-order polarization effects in the case of a single monochromatic plane wave. The directions of further generalizations are obvious: analysis of higher orders, taking account of many longitudinal modes (i.e., spectral-polarization analysis), and the case of two or more transverse modes. We note especially the case of two-photon states for two beams, which in recent years has been attracting extraordinary attention in connection with methodological problems of quantum mechanics (see Refs. 27 and 29–30). The polarization matrix of the density of such states was studied in Ref. 31. The polarization of n beams and nonplanar waves have also been investigated.[32–34] Some characteristic features of higher-order interference, specifically, its nonharmonic character, have been examined in Refs. 5, 6, and 35.

I am grateful to A. S. Chirkin for helpful discussions and for familarizing me with Refs. 12 and 13 prior to publication.

This work was supported by the Russian Fund for Fundamental Research under Grant No. 96-02-16334-a.

[1] M. Born and E. Wolf, *Principles of Optics*, Pergamon Press, N. Y., 1969, 4th edition [Russian translation, Nauka, Moscow, 1970]; A. Gerrad and J. M. Burch, *Introduction to Matrix Methods in Optics*, Dover, N. Y., 1994 [Russian translation, Mir, Moscow, 1978]; W. A. Shurcliff, *Polarized Light: Production and Use*, Harvard University Press, Cambridge, MA, 1962 [Russian translation, Mir, Moscow, 1965].
[2] C. M. Caves, Phys. Rev. D **23**, 1693 (1981).
[3] B. Yurke, S. L. McCall, and J. H. Klauder, Phys. Rev. A **33**, 4033 (1986).
[4] C. H. Bennet, F. Bessette, G. Brassard, L. Salvail, and J. Smolin, J. Crypto. **5**, 3 (1992); C. H. Bennett, Phys. Rev. Lett. **68**, 3121 (1992); Special Issue on Quantum Communication: J. Mod. Opt. **41**, No. 12 (1994).
[5] D. N. Klyshko, Phys. Lett. A **163**, 349 (1992).
[6] A. V. Belinsky and D. N. Klyshko, Laser Phys. **2**, 113 (1992).
[7] V. P. Karasev and A. N. Masalov, Opt. Spektrosk. **74**, 928 (1993) [Opt. Spectrosc. J. **74**, 551 (1993)].
[8] D. I. Guzun and A. N. Penin, Proc. SPIE **2799**, 249 (1996).
[9] A. Heidman, R. J. Horowicz, S. Reynaud, E. Giacobino, C. Fabre, and G. Camy, Phys. Rev. Lett. **59**, 2555 (1987).
[10] A. S. Chirkin, A. A. Orlov, and D. Yu. Parashchuk, Kvant. Élektron. **20**, 999 (1993) [Quan. Electron. **23**, 870 (1993)].
[11] A. P. Alodzhants, S. M. Arakelyan, and A. S. Chirkin, Zh. Éksp. Teor. Fiz. **108**, 63 (1995) [JETP **81**, 34 (1995)].
[12] N. V. Korolkova and A. S. Chirkin, J. Mod. Opt. **43**, 869 (1996).
[13] V. N. Beskrovnyĭ and A. S. Chirkin, Kvant. Élektron. **23**, 843 (1996).
[14] A. M. Perelomov, *Generalized Coherent States and Their Applications*, Springer-Verlag, N.Y., 1986 [Russian original, Nauka, Moscow, 1987].
[15] K. Wodkievicz and J. H. Eberly, J. Opt. Soc. Am. B **2**, 458 (1985).
[16] Ya. A. Smorodinskiĭ, A. L. Shelepin, and L. A. Shelepin, Usp. Fiz. Nauk. **162**, 1 (1992) [Sov. Phys. Usp. **35**, 1005 (1992)].
[17] N. Ya. Vilenkin, *Special Functions and the Theory of Group Representations*, American Mathermatical Soc., Providence, RI, 1968.
[18] G. Korn and T. Korn, *Mathematical Handbook*, McGraw-Hill, N. Y., 1961 [Russian translation, Nauka, Moscow, 1970].
[19] D. N. Klyshko, Usp. Fiz. Nauk **163**, 1 (1993).
[20] D. N. Klyshko, Zh. Éksp. Teor. Fiz. **105**, 1574 (1994) [JETP **78**, 848 (1994)].
[21] D. N. Klyshko, Phys. Lett. A **213**, 7 (1996).
[22] D. N. Klyshko, Usp. Fiz. Nauk **166**, 613 (1996).
[23] A. V. Masalov and D. N. Klyshko, Usp. Fiz. Nauk **165**, 1249 (1995).
[24] C. K. Hong, Z. Y. Ou, and L. Mandel, Phys. Rev. Lett. **59**, 2044 (1987).
[25] A. M. Steinberg, P. G. Kwiat, and R. Y. Chiao, Phys. Rev. Lett. **68**, 2421 (1992).
[26] Y. H. Shih and A. V. Sergienko, Phys. Lett. A **186**, 29 (1994).
[27] D. N. Klyshko, Usp. Fiz. Nauk **164**, 1187 (1994).
[28] D. N. Klyshko, Phys. Lett. A **146**, 471 (1990).
[29] T. B. Pittman, Y. H. Shih, A. V. Sergienko, and M. H. Rubin, Phys. Rev. A **51**, 3495 (1995).
[30] J. R. Torgersin, D. Branning, C. H. Mouken, and L. Mandel, Phys. Rev. A **51**, 4400 (1995); Phys. Lett. A **204**, 323 (1995).
[31] V. E. Mkrtchian and V. O. Chaltykian, Opt. Commun. **63**, 239 (1987).
[32] P. Roman, Nuovo Cimento **13**, 974 (1959).
[33] K. S. Vul'fson, Usp. Fiz. Nauk **152**, 667 (1987) [Sov. Phys. Usp. **30**, 724 (1987)].
[34] R. Barakat, J. Opt. Soc. Am. **53**, 317 (1963).
[35] A. V. Belinskiĭ and D. N. Klyshko, Zh. Éksp. Teor. Fiz. **102**, 1116 (1992) [Sov. Phys. JETP **75**, 606 (1992)].

Translated by M. E. Alferieff

The polarization-angular structure and elliptical dichroism of the cross sections for three-photon bound–bound transitions in atoms

N. L. Manakov*) and A. V. Merem'yanin

Voronezh State University, 394693 Voronezh, Russia

(Submitted 5 December 1996)
Zh. Éksp. Teor. Fiz. **111**, 1984–2000 (June 1997)

Using the electric dipole approximation, we present, in invariant form, the cross section of an arbitrary three-photon transition between the discrete states of an atom with total angular momenta J_i and J_f. The cross section contains scalar and mixed products of the photon polarization vectors, and invariant atomic parameters dependent only on the photon frequencies. We determine the number of independent atomic parameters at fixed values of J_i and J_f and obtain their explicit expressions in terms of the reduced composite dipole matrix elements. The polarization dependence of the cross sections is expressed in terms of the degrees l and ξ of linear and circular photon polarizations. We analyze the phenomenon of dissipation-induced circular dichroism in three-photon processes, i.e., the difference Δ of the cross sections for opposite signs of the degree of circular polarization of all the photons. We study in detail the case of two identical photons and the phenomenon of elliptical dichroism, when $\Delta \sim l\xi$ holds and dichroism occurs only when the photons are elliptically polarized, with $0<|\xi|<1$. Finally, we discuss the dissipation-induced effects of atom polarization in three-photon processes involving linearly polarized or unpolarized photons. © *1997 American Institute of Physics.* [S1063-7761(97)00506-4]

1. INTRODUCTION

In the early experiments on the interaction of laser light with atoms and molecules the light that was used was usually linearly polarized, so that the polarization dependence of the cross sections was not studied. At the same time, it is well-known that the polarization of the photon beam strongly influences the way in which multiphoton processes proceed. In particular, the dependence of cross sections of typical multiphoton processes on the absolute value of the ellipticity of the light field has been thoroughly studied (see, e.g., Refs. 1 and 2). A more interesting polarization effect, however, is the dependence of cross sections on the signs of the degree of circular polarization of the photons, i.e., on the sense of rotation of the electric field vector in the light wave, which is generally assumed elliptically polarized. Here the most interesting case is where the cross sections corresponding to different signs of the degrees of circular polarization of all photons participating in the process (incident photons and photons emitted as result of the interaction) are different. This specific effect, known as circular dichroism, is obvious for media with an inversion center (say, in chiral molecules), but under certain conditions it is present in atomic photoprocesses. In processes involving randomly oriented atomic particles, circular dichroism is determined by the interference of the real and imaginary parts of the partial amplitudes of the process and contains information about the nature of the interaction of the atomic particles and light that cannot be extracted from experiments with linearly polarized radiation.

The phenomenon of circular dichroism has been thoroughly studied (both theoretically and experimentally) only for the simplest photoprocess, the photoionization of atoms and molecules. Here it is present only in the ionization of pre-oriented (polarized) atoms or for a fixed orientation of the photoelectron spin. This fact is obvious from general symmetry considerations: since the degree ξ of circular photon polarization is a pseudoscalar quantity, the terms in the photoeffect cross section responsible for circular dichroism can contain ξ only in products of the type $\xi\mathbf{J}$, where $\mathbf{J}$ is the total angular momentum of the atom or the photoelectron spin, both of which are pseudovectors. The latest results in this field of research can be found in Refs. 3 and 4. They show that the difference in cross sections for clockwise- and counterclockwise-polarized photons may become quite large and make it possible to extract vital information concerning, in particular, the magnitude of the partial dipole matrix elements of a transition and the phases of electron scattering by a residual ion.

A more specific effect is circular dichroism in processes where photons interact with unpolarized atomic particles. Only recently has this effect been studied. For instance, Berakdar *et al.*[5] and Kabachnik and Schmidt[6] (see also Ref. 7) studied circular dichroism in the double photoeffect (the knock-out of two electrons by a single photon) and in photo-induced Auger decay. The conditions needed for dichroism to emerge in bremsstrahlung and inverse bremsstrahlung processes and in the scattering of electrons by atoms in the presence of a light wave have been discussed in Refs. 8 and 9. A detailed study of circular dichroism in Rayleigh and Raman scattering of light in gases can be found in Ref. 10, while the features of circular dichroism in the resonant two-photon excitation of atoms are discussed in Ref. 11. As shown in Ref. 10, circular dichroism emerges in two-photon bound–bound transitions only if one allows for the nondipole corrections in the atom–photon interaction and is essential in the range of frequencies in resonance with a dipole-forbidden transition in the atom, when the smallness of the nondipole effects in the cross section is compensated by the smallness of the resonance denominator.

In this paper we analyze the polarization effects in three-photon transitions between discrete atomic levels (three-photon excitation, hyper-Raman scattering, frequency mixing, etc.). In contrast to Placzek's well-known theory of two-photon scattering,[12] which is described in full by three invariant atomic parameters, separation of kinematic (i.e., depending on polarizations and directions of the wave vectors of the photons) and dynamic (atomic) factors in three-photon scattering cross sections is more complicated. Strizhevskiĭ and Klimenko[13] developed a phenomenological theory of nonresonant three-photon scattering in gases, but their approach did not clarify the relationship between the scattering parameters and the microscopic atomic parameters, and using the transparent-medium approximation excludes circular dichroism effects. The structure of the cross sections of three-photon processes in atoms was studied in Ref. 14, but the general results proved to be extremely complicated, since the angular part is expressed in terms of tensor products of six vectors (a structure that is difficult to analyze), and the atomic factors are expressed in terms of complicated combinations of reduced matrix elements incorporating Wigner $3nj$-symbols.

Employing in Sec. 2 a special technique for calculating tensor products of vectors (see Appendix A) and a convenient parametrization scheme for the photon polarization vectors in the general case of arbitrary (and, in particular, partial) polarization, we separate the geometric and dynamic factors in the cross section of an arbitrary three-photon transition between the bound states $|i\rangle$ and $|f\rangle$ of opposite parity, a transition allowed by the selection rules for electric dipole radiation. Generally, the cross section contains 15 different terms, four of which describe circular dichroism, which appears in three-photon processes even in the electric dipole approximation. As in two-photon transitions,[10] circular dichroism is present only if the partial transition amplitudes contain a skew-Hermitian part (''dissipation-induced dichroism'').

Section 3 analyzes the interesting (from the experimenter's viewpoint) situation when two of the three photons are identical, i.e., belong to the same laser pump beam. Here, for pump radiation that is arbitrarily (elliptically) polarized, an ''exhaustive experiment'' makes it possible to determine the six independent atomic parameters, while for linearly polarized radiation the cross section is described by only two different parameters. For identical photons, the dichroic term in the cross section contains the product of the degrees of linear and circular polarizations of the pump radiation, which means that circular dichroism is present in experiments involving two identical photons only if the pump radiation is elliptically polarized (elliptical dichroism).

Above we assumed that the target atoms are freely oriented in space, so that the cross sections are averaged and summed over the projections of the atomic angular momenta in the initial and final states. If we allow for dissipation effects caused by the skew-Hermitian part of the amplitude of three-photon processes, we can observe, in addition to circular dichroism, the specific effects of the orientation of atoms in their interaction with unpolarized or linearly polarized photons. In Sec. 4 we discuss these effects using the simple example of a three-photon transition between states whose total angular momentum J is $\frac{1}{2}$.

The paper uses the atomic system of units.

2. POLARIZATION-ANGULAR STRUCTURE OF THREE-PHOTON TRANSITION CROSS SECTIONS

2.1. The phenomenological approach

Let $\mathbf{e}_i$ be the unit complex-valued polarization vector of the ith photon with a frequency ω_i and a wave vector $\mathbf{k}_i=\boldsymbol{\kappa}_i\omega_i/c$, where $\mathbf{e}_i\mathbf{e}_i^*=1$ and $\mathbf{e}_i\boldsymbol{\kappa}_i=0$. In the electric dipole approximation, the cross section of an arbitrary three-photon transition between states $|i\rangle\equiv|\gamma_iJ_iM_i\rangle$ and $|f\rangle\equiv|\gamma_fJ_fM_f\rangle$ (J and M are the total angular momentum and its projection on the quantization axis, and γ are the other quantum numbers) of a freely orienting system contains, after averaging over M_i and summing over M_f, combinations of the six vectors $\mathbf{e}_i$ and $\mathbf{e}_i^*$ with $i=1, 2, 3$ and is a linear function of each of these vectors. Hence the general structure of the polarization–angular dependence of the cross section can be established phenomenologically by counting the number of linearly independent combinations of the vectors. Clearly, such combinations can be written in terms of pairs of scalar products, seven of which are real,

$$1,\ |\mathbf{e}_1\mathbf{e}_2|^2,\ |\mathbf{e}_1\mathbf{e}_2^*|^2,\ |\mathbf{e}_1\mathbf{e}_3|^2,\ |\mathbf{e}_1\mathbf{e}_3^*|^2,\ |\mathbf{e}_2\mathbf{e}_3|^2,\ |\mathbf{e}_2\mathbf{e}_3^*|^2, \tag{1}$$

and four are complex-valued,

$$A_1=(\mathbf{e}_1\mathbf{e}_2)(\mathbf{e}_2^*\mathbf{e}_3^*)(\mathbf{e}_1^*\mathbf{e}_3),\quad A_2=(\mathbf{e}_1^*\mathbf{e}_2^*)(\mathbf{e}_2\mathbf{e}_3^*)(\mathbf{e}_1\mathbf{e}_3),$$
$$A_3=(\mathbf{e}_1\mathbf{e}_2^*)(\mathbf{e}_2\mathbf{e}_3)(\mathbf{e}_1^*\mathbf{e}_3^*),\quad A_4=(\mathbf{e}_1^*\mathbf{e}_2)(\mathbf{e}_2^*\mathbf{e}_3)(\mathbf{e}_1\mathbf{e}_3^*). \tag{2}$$

On can easily see that the combinations A_1, A_2, A_3, and A_4 can be obtained from each other through complex conjugation and replacement of one pair of vectors $\mathbf{e}_i,\mathbf{e}_i^*$ by $\mathbf{e}_i^*,\mathbf{e}_i$. Thus, the cross section of an arbitrary three-photon process contains 15 terms, four of which (with $\mathrm{Im}\,A_i$), as we show below, describe circular dichroism.

In the general case of elliptically polarized photons the structure of (2) is extremely complicated since the vectors $\mathbf{e}_i$ are complex-valued. A convenient approach in analyzing the polarization effects in multiphoton processes is to employ the following invariant (with respect to the choice of coordinate system) parametrization of the vector $\mathbf{e}$ for the case of elliptically polarized photons with an ellipticity (the ratio of the semiaxes of the polarization ellipse) γ:

$$\mathbf{e}=\frac{\boldsymbol{\epsilon}+i\gamma[\boldsymbol{\kappa}\boldsymbol{\epsilon}]}{\sqrt{1+\gamma^2}},\quad -1\leqslant\gamma\leqslant 1, \tag{3}$$

where $\boldsymbol{\epsilon}$ is the unit vector along the major axis of the polarization ellipse, the vector that in the case of linear polarization (with $\gamma=0$) coincides with $\mathbf{e}$. In the adopted notation the electric field strength in the light field with amplitude F is described by the vector

$$\mathbf{F}(\mathbf{r},t)=F\ \mathrm{Re}\{\mathbf{e}\ \exp[-i\omega t+i(\mathbf{kr})]\}$$

$$=\frac{F\{\boldsymbol{\epsilon}\cos[\omega t-(\mathbf{kr})]+\gamma[\boldsymbol{\kappa}\boldsymbol{\epsilon}]\sin[\omega t-(\mathbf{kr})]\}}{\sqrt{1+\gamma^2}}. \tag{4}$$

Instead of ellipticity γ it is convenient to use the degrees of linear (l) and circular (ξ) polarizations,

$$l=\frac{1-\gamma^2}{1+\gamma^2}=(\mathbf{ee})=(\mathbf{e}^*\mathbf{e}^*),\quad \xi=\frac{2\gamma}{1+\gamma^2}=i(\boldsymbol{\kappa}[\mathbf{e}^*\mathbf{e}]), \tag{5}$$

related to the standard Stokes parameters.[12] For totally polarized radiation $l^2+\xi^2=1$.

Combining (3), and (5), we arrive at the following useful relationships:

$$|(\mathbf{ea})|^2=l(\boldsymbol{\epsilon}\mathbf{a})^2+\frac{1-l}{2}[\boldsymbol{\kappa}\mathbf{a}]^2,$$

$$(\mathbf{ea})(\mathbf{e}^*\mathbf{b})=\mathrm{Re}\{(\mathbf{ea})(\mathbf{e}^*\mathbf{b})\}-\frac{i}{2}\xi(\boldsymbol{\kappa}[\mathbf{ab}]),$$

$$2\,\mathrm{Re}\{(\mathbf{ea})(\mathbf{e}^*\mathbf{b})\}=2l(\mathbf{a}\boldsymbol{\epsilon})(\boldsymbol{\epsilon}\mathbf{b})+(l-1)([\mathbf{a}\boldsymbol{\kappa}][\boldsymbol{\kappa}\mathbf{b}]), \tag{6}$$

which are valid for real vectors **a** and **b**. These formulas make it possible to write the combinations in (1) containing two polarization vectors in terms of the vectors $\boldsymbol{\epsilon}$, $\boldsymbol{\kappa}$, $\boldsymbol{\epsilon}'$, and $\boldsymbol{\kappa}'$:

$$\begin{aligned}2|\mathbf{ee}'|^2={}&1-ll'(1-2(\boldsymbol{\epsilon}\boldsymbol{\epsilon}')^2)-\xi\xi'(\boldsymbol{\kappa}\boldsymbol{\kappa}')+l(l'-1)\\&\times(\boldsymbol{\epsilon}\boldsymbol{\kappa}')^2+l'(l-1)(\boldsymbol{\kappa}\boldsymbol{\epsilon}')^2-\tfrac{1}{2}(l-1)(l'-1)\\&\times[\boldsymbol{\kappa}\boldsymbol{\kappa}']^2.\end{aligned} \tag{7}$$

The vector combination A_i in (2) can also be written in a similar manner, but the result is more cumbersome. Below we give the expression for the imaginary part of A_1:

$$\begin{aligned}2\,\mathrm{Im}\,A_1\equiv{}&2\ \mathrm{Im}\{(\mathbf{e}_1\mathbf{e}_2)(\mathbf{e}_2^*\mathbf{e}_3^*)(\mathbf{e}_3\mathbf{e}_1^*)\}\\={}&\tfrac{1}{4}\xi_1\xi_2\xi_3(\boldsymbol{\kappa}_1[\boldsymbol{\kappa}_2\boldsymbol{\kappa}_3])-\xi_1P_{123}-\xi_2P_{213}\\&-\xi_3P_{312},\end{aligned} \tag{8a}$$

$$\begin{aligned}P_{ijk}={}&l_jl_k(\boldsymbol{\epsilon}_j\boldsymbol{\epsilon}_k)(\boldsymbol{\kappa}_i[\boldsymbol{\epsilon}_j\boldsymbol{\epsilon}_k])+\tfrac{1}{2}l_k(l_j-1)(\boldsymbol{\kappa}_j\boldsymbol{\epsilon}_k)(\boldsymbol{\kappa}_i[\boldsymbol{\kappa}_j\boldsymbol{\epsilon}_k])\\&+\tfrac{1}{2}l_j(l_k-1)(\boldsymbol{\epsilon}_j\boldsymbol{\kappa}_k)(\boldsymbol{\kappa}_i[\boldsymbol{\epsilon}_j\boldsymbol{\kappa}_k])+\tfrac{1}{4}(l_j-1)(l_k-1)\\&\times(\boldsymbol{\kappa}_j\boldsymbol{\kappa}_k)(\boldsymbol{\kappa}_i[\boldsymbol{\kappa}_j\boldsymbol{\kappa}_k]).\end{aligned} \tag{8b}$$

As noted earlier, the imaginary parts of A_2, A_3, and A_4 can also be expressed in the form (8) with the signs of two of the three parameters ξ_i reversed in (8a); for instance, $\mathrm{Im}\,A_2=\mathrm{Im}\,A_1(\xi_{1,2}\to-\xi_{1,2})$. Thus, by using the parametrization scheme (3) and (5) we can write the kinematic factors in the cross sections in a form convenient for analysis, i.e., in terms of the angles between the wave vectors of the photons and the unit vectors $\boldsymbol{\epsilon}_i$ that specify the directions of the major axes of the polarization ellipses.

Equations (8a) and (8b) show that the $\mathrm{Im}\,A_i$ change sign when ξ_i is replaced by $-\xi_i$, i.e., the corresponding terms in the three-photon transition cross sections describe circular dichroism. Clearly, these terms are of an interference nature and appear because of interference between the real and imaginary parts of the partial transition amplitudes. Indeed, the $\mathrm{Im}\,A_i$, being true scalars (the pseudoscalar mature of ξ_i is balanced by the presence of vector products of the polar vectors $\boldsymbol{\epsilon}_i$ and $\boldsymbol{\kappa}_i$ in each of the terms in (8a) and (8b)), are T-odd, since all the terms in (8a) and (8b) contain an odd number of vectors $\boldsymbol{\kappa}_i$, which change sign under time reversal. Hence the atomic factors with which the $\mathrm{Im}\,A_i$ enter into the cross sections must contain products of the real and imaginary parts of the partial amplitudes of the three-photon process being investigated, since only the skew-Hermitian (imaginary) part of the amplitude is T-odd. As the unitary property of the S-matrix implies,[12] the skew-Hermitian part of the amplitude of a specific process is always related to the amplitudes of other (in relation to the given process) physical processes that are realizable for the given initial states of the quantum system consisting of the atoms and the photons. These process, competing in a way with the given one, we term (in a generalized sense) dissipative processes, since they weaken the intensity of the beam of incident photons and, due to the irreversible nature of dissipation phenomena, introduce T-odd parameters into the problem; and the presence of these parameters is the cause for circular dichroism. In the case of three-photon bound–bound transitions under consideration, there are two dissipation channels: real population of the intermediate resonant level (here the dissipation parameter is the width of the resonant level, which is a T-odd quantity), and ionization of the atom if the energy of one or two incident photons is sufficient for ionizing the the atom from the initial state. Thus, circular dichroism in three-photon transitions is dissipation-induced, just as it is in two-photon transitions. The difference is that in two-photon transitions between bound states, circular dichroism emerges only if we allow for nondipole effects in the atom–photon interaction, since a T-odd scalar cannot be constructed from four vectors **e**, $\mathbf{e}^*$, $\mathbf{e}'$, and $\mathbf{e}'^*$. Allowing for this nondipole nature introduces two additional vectors $\boldsymbol{\kappa}$ and $\boldsymbol{\kappa}'$, so that the combinations of vectors in the cross sections of arbitrary two-photon transitions responsible for dichroism have the form[10]

$$E_1=(\mathbf{ee}'^*)(\mathbf{e}^*\boldsymbol{\kappa}')(\mathbf{e}'\boldsymbol{\kappa}),\quad E_2=(\mathbf{ee}')(\mathbf{e}^*\boldsymbol{\kappa}')(\mathbf{e}'^*\boldsymbol{\kappa}).$$

The adopted parametrization of **e** makes it possible to write these expressions in terms of the angles between real vectors:

$$\mathrm{Im}(E_1+E_2)=\xi l'(\boldsymbol{\epsilon}'\boldsymbol{\kappa})(\boldsymbol{\epsilon}'[\boldsymbol{\kappa}'\boldsymbol{\kappa}]),$$

$$\mathrm{Im}(E_1-E_2)=l\xi'(\boldsymbol{\epsilon}\boldsymbol{\kappa}')(\epsilon[\boldsymbol{\kappa}\boldsymbol{\kappa}']),$$

$$\begin{aligned}\mathrm{Re}(E_1+E_2)={}&2ll'(\boldsymbol{\epsilon}\boldsymbol{\epsilon}')(\boldsymbol{\epsilon}\boldsymbol{\kappa}')(\boldsymbol{\epsilon}'\boldsymbol{\kappa})+(\boldsymbol{\kappa}\boldsymbol{\kappa}')\{l(l'-1)\\&\times(\boldsymbol{\epsilon}\boldsymbol{\kappa}')^2+l'(l-1)(\boldsymbol{\epsilon}'\boldsymbol{\kappa})^2-\tfrac{1}{2}(l-1)\\&\times(l'-1)(1-(\boldsymbol{\kappa}\boldsymbol{\kappa}')^2)\},\end{aligned}$$

$$\mathrm{Re}(E_1-R_2)=\tfrac{1}{2}\xi\xi'((\boldsymbol{\kappa}\boldsymbol{\kappa}')^2-1).$$

2.2. Quantum mechanical formulas for the dynamic atomic factors

In the general case of photons with different frequencies, the amplitude A_{fi} of an arbitrary three-photon transition is determined by the sum of three composite matrix elements calculated in third-order perturbation theory, the matrix elements corresponding to the different combinations of photon emission and absorption acts. For instance, if, to be definite, we take the hyper-Raman scattering of two photons, $\mathbf{e}_1,\ \omega_1$ and $\mathbf{e}_2,\omega_2$, accompanied by emission of a scattered photon $\mathbf{e}_3,\omega_3$ $(E_i+\omega_1+\omega_2=E_f+\omega_3)$, we find that

$$\begin{aligned}
&A_{fi}(\mathbf{e}_1,\omega_1;\mathbf{e}_2,\omega_2;\mathbf{e}_3^*,-\omega_3)\\
&\quad=\langle\gamma_f J_f M_f|\{(\mathbf{e}_3^*\mathbf{d})G_{E_i+\omega_1+\omega_2}(\mathbf{e}_2\mathbf{d})G_{E_i+\omega_1}(\mathbf{e}_1\mathbf{d})\\
&\qquad+(\mathbf{e}_2\mathbf{d})G_{E_i+\omega_1-\omega_3}(\mathbf{e}_3^*\mathbf{d})G_{E_i+\omega_1}(\mathbf{e}_1\mathbf{d})\\
&\qquad+(\mathbf{e}_2\mathbf{d})G_{E_i+\omega_1-\omega_3}(\mathbf{e}_1\mathbf{d})G_{E_i-\omega_3}(\mathbf{e}_3^*\mathbf{d})\\
&\qquad+(\mathbf{e}_1\mathbf{d})G_{E_i+\omega_2-\omega_3}(\mathbf{e}_2\mathbf{d})G_{E_i-\omega_3}(\mathbf{e}_3^*\mathbf{d})\\
&\qquad+(\mathbf{e}_1\mathbf{d})G_{E_i+\omega_2-\omega_3}(\mathbf{e}_3^*\mathbf{d})G_{E_i+\omega_2}(\mathbf{e}_2\mathbf{d})\\
&\qquad+(\mathbf{e}_3^*\mathbf{d})G_{E_i+\omega_1+\omega_2}(\mathbf{e}_1\mathbf{d})G_{E_i+\omega_2}(\mathbf{e}_2\mathbf{d})\}|\gamma_i J_i M_i\rangle. \qquad (9)
\end{aligned}$$

Here $\mathbf{d}$ is the atomic dipole-moment operator,

$$G_E=\sum_{\gamma JM}\frac{|\gamma JM\rangle\langle\gamma JM|}{E_{\gamma J}-E+i0}=\sum_{JM}G_E^J|JM\langle JM|$$

is the Green's function of the atom, and G_E^J is the part of G_E corresponding to a total angular momentum J, and $|JM\rangle$ is the spin–angular part of the atomic wave function with a total angular momentum J determined by the angular-momentum coupling scheme in the particular atom. The amplitudes of the other processes involving three photons are of the same type, with the signs of the frequencies reversed and the corresponding polarization vectors $\mathbf{e}$ replaced by complex-conjugate vectors.

Writing the scalar product $(\mathbf{ed})$ in (9) in a spherical base and using the technique of irreducible tensor operators, we can separate the dependence of A_{fi} on the projections M_i and M_f from the dependence of A_{fi} on the polarization vectors $\mathbf{e}_i$ explicitly (see, e.g., Ref. 14):

$$\begin{aligned}
A_{fi}=&\sum_{x,\xi}(-1)^{x-\xi}C^{J_iM_i}_{J_fM_f\,x-\xi}\sum_{y=0,1,2}Q_{xy}(-\omega_3;\omega_1,\omega_2)\\
&\times\{\mathbf{e}_3^*\otimes\{\mathbf{e}_1\otimes\mathbf{e}_2\}_y\}_{x,-\xi}, \qquad (10)
\end{aligned}$$

where the atomic parameters

$$\begin{aligned}
&Q_{xy}(-\omega_3;\omega_1,\omega_2)=\sqrt{\frac{(2x+1)(2y+1)}{2J_i+1}}\\
&\quad\times\sum_{J_1,J_2}(-1)^{y+J_i-J_1}\left[\begin{Bmatrix}1&1&y\\J_i&J_1&J_2\end{Bmatrix}\right.\\
&\quad\times\begin{Bmatrix}1&x&y\\J_i&J_1&J_f\end{Bmatrix}T^y_{J_1J_2}(\omega_1+\omega_2,\omega_2;\omega_1+\omega_2,\omega_1)\\
&\quad+(-1)^{J_1+J_2+1-J_i-J_f+x}\begin{Bmatrix}1&1&y\\J_f&J_2&J_1\end{Bmatrix}\\
&\quad\times\begin{Bmatrix}1&x&y\\J_f&J_2&J_i\end{Bmatrix}T^y_{J_1J_2}(\omega_1-\omega_3,-\omega_3;\omega_2-\omega_3,-\omega_3)\\
&\quad+\begin{Bmatrix}J_f&1&J_1\\J_i&1&J_2\\x&y&1\end{Bmatrix}T^y_{J_1J_2}(\omega_2-\omega_3,\omega_2;\omega_1-\omega_3,\omega_1)\left.\vphantom{\begin{Bmatrix}1\\1\\1\end{Bmatrix}}\right] \qquad (11)
\end{aligned}$$

are expressed in terms of combinations of the reduced composite matrix elements

$$\begin{aligned}
T^y_{J_1J_2}(\alpha,\beta;\gamma,\delta)=&\langle\gamma_fJ_f\|\mathbf{d}\{G^{J_1}_{E_i+\alpha}\mathbf{d}G^{J_2}_{E_i+\beta}\\
&+(-1)^yG^{J_1}_{E_i+\gamma}\mathbf{d}G^{J_2}_{E_i+\delta}\}\mathbf{d}\|\gamma_iJ_i\rangle. \qquad (12)
\end{aligned}$$

Here we have used the standard notation of the angular-momentum technique[15] for Clebsch–Gordan coefficients, Wigner $3nj$-symbols, and tensor products.

Equation (10) implies that the polarization dependence of the cross sections can be expressed, to within factors determined by the type of process (the density of the final states, etc.), in terms of tensor products of the polarization vectors:

$$\begin{aligned}
M_{J_iJ_f}=&\frac{1}{2J_i+1}\sum_{M_i,M_f}|A_{fi}|^2\\
=&\sum_{x=0}^{3}\frac{1}{2x+1}\sum_{\xi=-x}^{x}\left|\sum_{y=0}^{2}Q_{xy}\{\mathbf{e}_3^*\otimes\{\mathbf{e}_1\otimes\mathbf{e}_2\}_y\}_{x,-\xi}\right|^2. \qquad (13)
\end{aligned}$$

Here the Q_{xy} with fixed x and y act as the partial transition amplitudes. As Eqs. (10) and (11) imply, generally (for arbitrary values of J_i and J_f) seven parameters Q_{xy} are nonzero, with the result that (13) consists of 15 different terms containing tensor products of six vectors. Using the rules governing the change of the coupling scheme in tensor products, we can write (13) in terms of tensor products of the photon polarization tensors $\{\mathbf{e}\otimes\mathbf{e}^*\}_{pm}$, but as a result the atomic coefficients become extremely cumbersome.[14] The method of calculating tensor products of vectors discussed in Appendix A allows writing the $M_{J_iJ_f}$ in terms of ordinary scalar products of vectors:

$$\begin{aligned}
M_{J_iJ_f}=&f_0+f_1|\mathbf{e}_1\mathbf{e}_2|^2+f_2|\mathbf{e}_1\mathbf{e}_2^*|^2+f_3|\mathbf{e}_1\mathbf{e}_3|^2+f_4|\mathbf{e}_1\mathbf{e}_3^*|^2\\
&+f_5|\mathbf{e}_2\mathbf{e}_3|^2+f_6|\mathbf{e}_2\mathbf{e}_3^*|^2+\sum_{i=1}^{4}(\mathrm{Re}\,g_i\,\mathrm{Re}\,A_i\\
&-\mathrm{Im}\,g_i\,\mathrm{Im}\,A_i), \qquad (14)
\end{aligned}$$

in conformity with the phenomenological ideas expressed in Sec. 2.1. The explicit expressions for the coefficients f_i and g_i in terms of bilinear combinations of Q_{xy} are given in Appendix B.

Thus, in the most general case of an exhaustive experiment, polarization measurements can provide the 15 independent atomic parameters f_i and g_i that describe the cross section of a three-photon process. In the absence of dissipative processes (nonresonant transitions with a total energy of the incident photons insufficient for ionizing the atom), the Q_{xy} are real-valued, with the result that $\mathrm{Im}\, g_i = 0$ and the number of independent parameters is reduced to 11.

Equation (14) gives the polarization–angular of the cross sections in the most general case of fixed polarizations $\mathbf{e}_i$ and directions of propagation $\boldsymbol{\kappa}_i$ of all three photons. If, for instance, in the above case of hyper-Raman scattering there is no scattered photon, averaging in (14) over the polarizations $\mathbf{e}_3$ and integrating over the directions $\boldsymbol{\kappa}_3$ (which amounts, as shown in Ref. 12, to replacing $(\mathbf{e}_3)_i(\mathbf{e}_3^*)_j$ by $(4\pi/3)\delta_{ij}$) yield the following expression for $M_{J_iJ_f}$:

$$M_{J_iJ_f} = \tfrac{1}{9}\alpha^s|\mathbf{e}_1\mathbf{e}_2|^2 + \tfrac{1}{18}\alpha^a(1-|\mathbf{e}_1\mathbf{e}_2^*|^2) + \tfrac{1}{30}\alpha^t(1+|\mathbf{e}_1\mathbf{e}_2^*|^2 - \tfrac{2}{3}|\mathbf{e}_1\mathbf{e}_2|^2). \tag{15}$$

In this case circular-dichroism effects vanish and the dependence of the cross section on the polarizations $\mathbf{e}_1$ and $\mathbf{e}_2$ of the incident photons is, obviously, the same as in the case of two-photon excitation in Placzek's theory. In our case the following combinations of the parameters Q_{xy} act as the scalar, skew-symmetric, and tensor parts of the cross-section of the two-photon process:

$$\alpha^s = |Q_{10}|^2, \quad \alpha^a = \sum_{x=0}^{2} |Q_{x1}|^2, \quad \alpha^t = \sum_{x=1}^{3} |Q_{x2}|^2.$$

If the only polarization that is not measured is that of the scattered photon, the dependence of the cross section on the vector $\boldsymbol{\kappa}_3$ is retained $((\mathbf{e}_3)_i(\mathbf{e}_3^*)_j \rightarrow (1/2)[\delta_{ij} - (\boldsymbol{\kappa}_3)_i(\boldsymbol{\kappa}_3)_j])$:

$$M_{J_iJ_f} = p_0 + p_1|\mathbf{e}_1\mathbf{e}_2|^2 + p_2|\mathbf{e}_1\mathbf{e}_2^*|^2 + p_3|\mathbf{e}_1\boldsymbol{\kappa}_3|^2 + p_4|\mathbf{e}_2\boldsymbol{\kappa}_3|^2 + \sum_{i=1}^{2} (\mathrm{Re}\, q_i\ \mathrm{Re}\, A_i' - \mathrm{Im}\, q_i\ \mathrm{Im}\, A_i'). \tag{16}$$

Here there are ten nonzero atomic parameters, two of which ($\mathrm{Im}\, q_{1,2}$) describe circular dichroism. The explicit expressions for p_i and q_i in terns of Q_{xy} are given in Appendix B. The real and imaginary parts of the vector combinations

$$A_1' = (\mathbf{e}_1\mathbf{e}_2)(\mathbf{e}_2^*\boldsymbol{\kappa}_3)(\mathbf{e}_1^*\boldsymbol{\kappa}_3), \quad A_2' = (\mathbf{e}_1\mathbf{e}_2^*)(\mathbf{e}_2\boldsymbol{\kappa}_3)(\mathbf{e}_1^*\boldsymbol{\kappa}_3)$$

can be written in terms of the real-valued vectors $\boldsymbol{\epsilon}_i$ and $\boldsymbol{\kappa}_i$. For instance,

$$\begin{aligned} 2\,\mathrm{Im}\, A_1' = {} & \xi_1\{l_2(\boldsymbol{\epsilon}_2\boldsymbol{\kappa}_3)(\boldsymbol{\epsilon}_2[\boldsymbol{\kappa}_1\boldsymbol{\kappa}_3]) - \tfrac{1}{2}(l_2-1)(\boldsymbol{\kappa}_2\boldsymbol{\kappa}_3) \\ & \times(\boldsymbol{\kappa}_3[\boldsymbol{\kappa}_1\boldsymbol{\kappa}_2])\} + \xi_2\{l_1(\boldsymbol{\epsilon}_1\boldsymbol{\kappa}_3)(\boldsymbol{\epsilon}_1[\boldsymbol{\kappa}_2\boldsymbol{\kappa}_3]) \\ & - \tfrac{1}{2}(l_1-1)(\boldsymbol{\kappa}_1\boldsymbol{\kappa}_3)(\boldsymbol{\kappa}_3[\boldsymbol{\kappa}_1\boldsymbol{\kappa}_2])\}. \end{aligned}$$

Equation (16) in which $\boldsymbol{\kappa}_3$ is replaced by $\mathbf{e}_3$ (and other expressions for the parameters p_i and q_i) is also valid in the case of a purely linear polarization of the scattered photon.

If two photons are linearly polarized (say, $\mathbf{e}_1 = \mathbf{e}_1^*$ and $\mathbf{e}_2 = \mathbf{e}_2^*$), then

$$A_1 = A_2 = A_3^* = A_4^*, \quad \mathrm{Im}\, A_1 = -\tfrac{1}{2}\xi_3(\mathbf{e}_1\mathbf{e}_2)(\boldsymbol{\kappa}_3[\mathbf{e}_1\mathbf{e}_2]),$$

and the cross section contains six terms:

$$M_{J_iJ_f} = \varphi_0 + \varphi_1(\mathbf{e}_1\mathbf{e}_2)^2 + \varphi_2|\mathbf{e}_1\mathbf{e}_3|^2 + \varphi_3|\mathbf{e}_2\mathbf{e}_3|^2 + \varphi_4\ \mathrm{Re}\, A_1 + \varphi_5\xi_3(\mathbf{e}_1\mathbf{e}_2)(\boldsymbol{\kappa}_3[\mathbf{e}_1\mathbf{e}_2]), \tag{17}$$

where the coefficients φ_0 to φ_5 are related in a simple manner to the parameters f_i and g_i in (14).

2.3. Selection rules for three-photon transitions

As the general expression (9) for an amplitude implies, the selection rules for three-photon transitions coincide with those of the matrix elements of a rank-3 tensor constructed from three polar vectors **d**. As is known,[15] each rank-3 tensor is decomposable into seven irreducible tensors with ranks x ranging from 0 to 3, but such decomposition is non-unique due to the ambiguity in the choice of the tensors of ranks 1 and 2. Hence the representation (10) of A_{fi}, which corresponds to this decomposition, and formula (13) contain additional summation with respect to index y. As a result, (10) contains two tensor products of rank $x=2$ (with $y=1,2$) and three tensors of rank 1 (with $y=0,1,2$), whose interference in (13) is the source of polarization anomalies in the transition cross sections.

The selection rules for the separate terms in (13) are the same as those for the electric octupole $(x=3)$, magnetic quadrupole $(x=2)$, and electric dipole $(x=1)$ radiations and a pseudoscalar $(x=0)$, respectively. In particular, transitions are possible only between states $|i\rangle$ and $|f\rangle$ of opposite parity, and the subscripts x and y in the partial amplitudes Q_{xy} obey the following selection rules:

$$3 \geqslant x \geqslant |J_i - J_f|, \quad 2 \geqslant y = x, x\pm 1, \quad \Delta J \equiv |J_i - J_f| \leqslant 3. \tag{18}$$

As a result, for given angular momenta J_i and J_f with $\Delta J \leqslant 3$ the number of nonzero parameters Q_{xy} and independent coefficients in (14) depends on the values of J_i and J_f. The cross section of a transitions between states with maximum possible angular momenta difference $\Delta J = 3$ has the simplest structure. In this case in (13) only the octupole part of the cross section is nonzero, with the cross section determined by the partial amplitude Q_{32}, and the corresponding tensor product of the vectors contains all combinations of vectors (1) and (2), with the exception of $\mathrm{Im}\, A_i$. This is the only type of transition in which there is no circular dichroism for open dissipation channels, with the cross section determined solely by the atomic parameter $|Q_{32}|^2$. For transitions with $\Delta J = 2$ the only nonzero partial amplitudes are Q_{21}, Q_{22}, and (for $J_i, J_f > 0$) Q_{32}. In this case the cross sections contains only one dichroic term $\sim \mathrm{Im}\, Q_{22}Q_{21}^*$ (see Appendix B) caused by the interference of the magnetic quadrupole amplitudes Q_{2y}. The amplitude Q_{01} enters into the cross sections only for transitions with $\Delta J = 0$, with the result that all seven partial amplitude Q_{xy} contribute to the cross sections only for transitions with $J_i = J_f \geqslant 3$.

The general results become simpler when both J_i and J_f are small. Earlier we noted that for transitions with $J_i = 0$, $J_f = 3$ and $J_i = 0$, $J_f = 2$ only one and two amplitudes

Q_{xy} are nonzero, respectively. For $J_i=0$, $J_f=1$ and $J_i=J_f=\frac{1}{2}$ only three and four parameters Q_{xy} are nonzero, respectively, with $\mathrm{Im}\, g_3=0$, and circular dichroism is described by the interference of the real and imaginary parts of the quantities α_i introduced in Appendix B, which are combinations of electric dipole amplitudes Q_{1y}.

3. THREE-PHOTON TRANSITIONS WITH IDENTICAL PHOTONS

Of special interest is the case, important for practical reasons, when two of the three photons are identical (belong to the same light beam). To make matters simple, we assume the photons are totally polarized and set, for the sake of simplicity, $(\mathbf{e},\boldsymbol{\kappa}_1)=(\mathbf{e}_2,\boldsymbol{\kappa}_2)\equiv(\mathbf{e},\boldsymbol{\kappa})$ and $(\mathbf{e}_3,\boldsymbol{\kappa}_3)\equiv(\mathbf{e}',\boldsymbol{\kappa}')$. The transition amplitude in this case is (cf. (9))

$$\begin{aligned}\widetilde{A}_{fi}(\mathbf{e},\omega;\mathbf{e}'^*,-\omega')&=\langle\gamma_f J_f M_f|\{(\mathbf{e}'^*\mathbf{d})G_{E_i+2\omega}(\mathbf{ed})G_{E_i+\omega}(\mathbf{ed})\\&\quad+(\mathbf{ed})G_{E_i+\omega-\omega'}(\mathbf{e}'^*\mathbf{d})G_{E_i+\omega}(\mathbf{ed})\\&\quad+(\mathbf{ed})G_{E_i+\omega-\omega'}(\mathbf{ed})G_{E_i-\omega'}(\mathbf{e}'^*\mathbf{d})\}|\gamma_i J_i M_i\rangle\end{aligned}\tag{19}$$

and can be written in the form (10). Here the terms with $y=1$ are not present in the sum over y, so that the cross sections are determined only by the four (generally complex-valued) partial amplitudes Q_{10}, Q_{12}, Q_{22}, and Q_{32}, which are expressed by relations of the form (11) with $T^y_{J_1J_2}(\alpha,\beta;\gamma,\delta)$ replaced by

$$\widetilde{T}_{J_1J_2}(\alpha,\beta)=\langle\gamma_f J_f\|\mathbf{d}G^{J_1}_{E_i+\alpha}\mathbf{d}G^{J_2}_{E_i+\beta}\mathbf{d}\|\gamma_i J_i\rangle.$$

The combinations of vectors (2) in this case also become simpler:

$$A_2=A_1^*=l(\mathbf{ee}')(\mathbf{ee}'^*),\quad A_3=|\mathbf{ee}'|^2,\quad A_4=|\mathbf{ee}'^*|^2,$$

with

$$\mathrm{Im}\, A_2=\xi l\{l'(\boldsymbol{\epsilon\epsilon}')(\boldsymbol{\kappa}[\boldsymbol{\epsilon\epsilon}'])+\tfrac{1}{2}(1-l')(\boldsymbol{\epsilon\kappa}')(\boldsymbol{\epsilon}[\boldsymbol{\kappa\kappa}'])\},\tag{20}$$

$$\begin{aligned}2\,\mathrm{Re}\, A_2=l\{&2l'(\boldsymbol{\epsilon\epsilon}')^2-(l-1)(l'-1)(1+(\boldsymbol{\kappa\kappa}')^2)\\&+(l-1)[\boldsymbol{\kappa\epsilon}']^2-(l'-1)[\boldsymbol{\epsilon\kappa}']^2\}.\end{aligned}$$

Hence the polarization–angular structure of the cross section is given by the following expression (cf. (14)):

$$\begin{aligned}\widetilde{M}_{J_iJ_f}=&\,a_1+a_2l^2+a_3|\mathbf{ee}'|^2+a_4|\mathbf{ee}'^*|^2\\&+a_5\,\mathrm{Re}\, A_2+a_6\,\mathrm{Im}\, A_2,\end{aligned}\tag{21}$$

where

$$a_1=\frac{1}{3}\left(\frac{1}{7}|Q_{32}|^2+\frac{2}{5}|Q_{22}|^2\right),$$

$$a_2=\frac{1}{15}\left(\frac{1}{3}|Q_{12}-\sqrt{5}Q_{10}|^2-\frac{1}{7}|Q_{32}|^2=|Q_{22}|^2\right),$$

$$a_3=\frac{2}{3}\left(\frac{1}{7}|Q_{32}|^2-\frac{2}{5}|Q_{22}|^2-\frac{1}{5}|Q_{21}|^2\right),$$

$$a_4=\frac{1}{15}\left(3|Q_{12}|^2-\frac{4}{7}|Q_{32}|^2-|Q_{22}|^2\right),$$

$$a_5=\frac{2}{15}\left(\sqrt{5}\,\mathrm{Re}\{Q_{10}Q_{12}^*\}-\frac{2}{7}|Q_{32}|^2+|Q_{22}|^2-|Q_{12}|^2\right),$$

$$a_6=\frac{2\sqrt{5}}{15}\,\mathrm{Im}\{Q_{10}Q_{12}^*\}.$$

Equations (20) show that the last term in (21) is nonzero only when the pump field is elliptically polarized, with $0<|\xi|<1$, and changes sign when ξ is replaced by $-\xi$, so that it leads to a dependence in the cross sections on the sense of rotation of the pump electric field vector (elliptical dichroism). As is the case with circular dichroism, elliptical dichroism is determined by the interference of the real and imaginary parts of the partial amplitudes, and its measurement provides data on the dissipative parameters of the medium, inaccessible in experiments with linear or circular polarization of the pump photons.

The selection rules (18) for the amplitudes Q_{xy} are also valid for transitions involving identical photons, with the additional requirement $Q_{x1}=0$. Hence elliptical dichroism is absent not only from transitions with $\Delta J=3$ but also from transitions with $\Delta J=2$. In the latter case the cross section is determined by two noninterfering amplitudes, Q_{32} and Q_{22}. All four parameters Q_{xy} enter into the cross section only for transitions with $J_i+J_f\geqslant 3$ and $\Delta J=0, 1$. At $J_i=0$, $J_f=1$ and $J_i=J_f=\frac{1}{2}$ the cross sections are determined only by two parameters, Q_{10} and Q_{12}.

In the case of three-photon excitation by the field of a single light beam, all three photons are identical. Here the cross section is determined only by two atomic factors with arbitrary J_i and J_f:

$$M_{J_iJ_f}=a+bl^2,\tag{22}$$

where

$$a=\tfrac{1}{7}|Q_{32}|^2,\quad b=\tfrac{1}{45}|\sqrt{5}Q_{10}+2Q_{12}|^2-\tfrac{3}{35}|Q_{32}|^2.$$

To find the atomic parameters of the three-photon transitions, the reduced matrix elements (12) in Eq. (11) for Q_{xy} must be calculated with allowance for the scheme of angular momentum coupling in the specific atom. For instance, for the hyper-Raman transition $|\gamma_i S_{1/2}\rangle+2\omega\rightarrow|\gamma_f P_{1/2}\rangle+\omega'$ in an atom with one electron in the outer shell (e.g., for alkali atoms), the partial amplitudes $Q_{10,12}$ have the form

$$\begin{aligned}\sqrt{3}Q_{10}=&\,R^s_{1/2}(\omega,\omega)+R^s_{1/2}(\omega,-\omega')+2[R^s_{3/2}(\omega,\omega)\\&+R^d_{1/2}(\omega,-\omega')]+\tfrac{1}{3}[2R^d_{3/2}(-\omega',\omega)+4R^d_{1/2}\\&\times(-\omega',\omega)+4R^s_{3/2}(-\omega',\omega)-R^s_{1/2}(-\omega',\omega)],\end{aligned}$$

$$\begin{aligned}\sqrt{\frac{3}{5}}\,Q_{12}=&\,R^d_{1/2}(\omega,\omega)+R^d_{1/2}(\omega,-\omega')+\frac{1}{5}\left[R^d_{3/2}(\omega,\omega)\right.\\&\left.+R^d_{3/2}(\omega,-\omega')-\frac{2}{3}R^d_{3/2}(-\omega',\omega)\right]\end{aligned}$$

$$+\frac{1}{3}\,[-R^d_{1/2}(-\omega',\omega)+R^s_{3/2}(-\omega',\omega)+2R^s_{1/2}(-\omega',\omega)].$$

Here

$$R^s_J(\alpha,\beta)=\langle\gamma_f P_{1/2}|dg_{1/2,0}(E_i+\alpha+\beta)dg_{J,1}(E_i+\beta)d|\gamma_i S_{1/2}\rangle,$$

$$R^d_J(\alpha,\beta)=\langle\gamma_f P_{1/2}|dg_{3/2,2}(E_i+\alpha+\beta)dg_{J,1}\times(E_i+\beta)d|\gamma_i S_{1/2}\rangle$$

are the radial composite matrix elements corresponding to the two values $J=1/2,3/2$ of the total angular momentum in the radial part $g_{JL}(E;r,r')$ of the Green's function.

The numerical calculation of the $R^{s,d}_J(\alpha,\beta)$ can be carried out using standard methods for calculating multiphoton cross sections by perturbation techniques;[1,2] the numerical values of cross sections for a number of three-photon transitions in specific atoms can be found, for instance, in Ref. 16. For above-threshold frequencies ($E_i+\alpha=E>0$ or $E_i+\alpha+\beta=E>0$), the matrix elements $R^{s,d}(\alpha,\beta)$ have imaginary parts proportional to the product of amplitudes of one- and two-photon ionization from states $|i\rangle$ and $|f\rangle$ into the same continuum state with energy E. Generally, $\mathrm{Re}\,R^{s,d}$ and $\mathrm{Im}\,R^{s,d}$ are of the same order of magnitude, so that the circular and elliptical dichroism effects are of order unity (for a similar calculation of the matrix elements for circular dichroism in two-photon dipole-forbidden above-threshold scattering of light see Ref. 10).

In resonant processes the effects of circular dichroism prove to be small, since the interference terms in the cross section proportional to Γ are small compared to the purely resonant part of the cross section. For instance, if Δ is the fine-structure interval of the resonant level of width Γ, the dichroism terms are of order Γ/Δ, the same as the dissipation-induced orientational effects (see Sec. 4) in dipole-allowed two-photon scattering.[17] For this reason, as in the case of two-photon scattering,[10] circular and elliptic dichroism effects in resonant three-photon processes may become significant only for a resonance on a dipole-forbidden transition. Here the polarization–angular structure of the cross section is even more cumbersome than (14), since the resonant part of the amplitude incorporates the wave vectors $\mathbf{k}_{1,2}$ originating in the nondipole corrections to the atom–photon interaction.

Here we do not give the corresponding formulas because they can be written in a simpler form for each concrete geometry of the experiment.

4. DISSIPATION-INDUCED ATOMIC-ORIENTATION EFFECTS IN THREE-PHOTON SCATTERING

Earlier we have seen that dichroism effects are specific interference effects and emerge because of the presence in the problem of the T-odd pseudoscalar $\xi\boldsymbol{\kappa}=i[\mathbf{e}^*\mathbf{e}]$ inherent in an elliptically polarized wave and a T-odd (dissipative) atomic parameter. Another class of interference phenomena caused by dissipation processes is related to the polarization of atoms in multiphoton transitions in fields with a zero degree of circular polarization. In this case the axial T-odd vector is the average angular momentum $\mathbf{J}$ of the atom in the initial ($\mathbf{J}_i$) or final ($\mathbf{J}_f$) state.

The dependence of the cross sections of two-photon processes on the orientation of the atom in the initial state and the emergence of a preferred orientation in the final state in the case of a unpolarized initial states have been analyzed in Ref. 17 and are due to the interference term in the cross section of the following type ($\mathbf{j}=\mathbf{J}_i$ or $\mathbf{J}_f$):

$$\Gamma(\boldsymbol{\kappa}_1\boldsymbol{\kappa}_2)([\boldsymbol{\kappa}_1\boldsymbol{\kappa}_2]\mathbf{j}) \tag{23a}$$

for unpolarized photons and

$$\Gamma(\mathbf{e}_1\mathbf{e}_2)([\mathbf{e}_1\mathbf{e}_2]\mathbf{j}) \tag{23b}$$

for linearly polarized photons.

Obviously, dissipation-induced orientational phenomena can also occur in multiphoton processes involving three or more photons. The most interesting is the case of double-frequency fields with polarization vectors $\mathbf{e}_1$ and $\mathbf{e}_2$. Earlier we noted that two effects are possible here: the dependence of cross sections on the orientation of the initial atomic state with a varying orientation of the final state (this requires that the gaseous medium be pre-oriented, say, by optically pumping the medium), and the emergence of orientation in the final state with the initial state of the gas unpolarized. In the latter case the orientation in processes of the light-scattering type is of a correlational nature, with the result that it can be fixed only in the process of registering the photons scattered in a given direction $\boldsymbol{\kappa}_2$ and depends on the direction of $\boldsymbol{\kappa}_2$.

Below we give the results for the simplest case of a hyper-Raman transition between states with $J_i=J_f=1/2$ (e.g., for the $nS_{1/2}-n'P_{1/2}$ transition in alkali-metal atoms) accompanied by the absorption of two photons $\mathbf{e}$, $\boldsymbol{\kappa}$, ω and the emission of one photon $\mathbf{e}'$, $\boldsymbol{\kappa}'$, ω'. The transition probability averaged over the orientations $M_i=\pm 1/2$ of the initial state is calculated for a fixed projection M_f of the angular momentum of the final state onto the given direction $\mathbf{N}$ in space (the axis of the detector measuring the orientation of the excited atom). The vector $\mathbf{j}=(M_f)_{\mathrm{av}}\mathbf{N}$, where $(M_f)_{\mathrm{av}}$ is the average value of M_f, characterizes the orientation of the atom in the final state ($0\leqslant|\mathbf{j}|\leqslant 1/2$).

For arbitrary (elliptical) polarizations of the incident and scattered photons, the polarization–angular structure of the cross section can be written as follows:

$$2\widetilde{M}(\mathbf{e},\boldsymbol{\kappa},\mathbf{e}',\boldsymbol{\kappa}',\mathbf{j})=a_2^{(1/2)}l^2(1+2\xi'\boldsymbol{\kappa}'\mathbf{j})+a_4^{(1/2)}|\mathbf{e}\mathbf{e}'^*|^2(1-2\xi\boldsymbol{\kappa}\mathbf{j})+a_5^{(1/2)}(\mathrm{Re}\,A_2+2l\,\mathrm{Im}\,\mathbf{Bj})+a_6^{(1/2)}(\mathrm{Im}\,A_2-2l\,\mathrm{Re}\,\mathbf{Bj}). \tag{24}$$

Here $\mathrm{Re}\,A_2$ and $\mathrm{Im}\,A_2$ are given by (20),

$$\mathbf{Bj}=(\mathbf{e}\mathbf{e}'^*)([\mathbf{e}\mathbf{e}']\mathbf{j}),$$

and the coefficients $a_i^{(1/2)}$ are the same as in (21) if in the latter we put $J_i=J_f=1/2$. Here $a_1=a_3=0$, while the other coefficients can be expressed in terms of the parameters Q_{10} and Q_{12}. Thus, for a transition with $J_i=J_f=1/2$ the orientational effects are described by the same atomic parameters as in the case of an unpolarized atom. Note that Eq.

(24) can also be used to analyze three-photon transitions in a pre-oriented atom without analyzing the orientation in the final state (for a similar treatment of two-photon scattering see Ref. 17).

In the general case of arbitrary photon polarizations, the real and imaginary parts of **Bj** are fairly complicated:

$$\begin{aligned}2\,\mathrm{Re}\,\mathbf{B}\mathbf{j}=&\,2l'(\boldsymbol{\epsilon}\boldsymbol{\epsilon}')([\boldsymbol{\epsilon}\boldsymbol{\epsilon}']\mathbf{j})+\tfrac{1}{2}\xi\xi'[(\boldsymbol{\kappa}'\boldsymbol{\epsilon})([\boldsymbol{\kappa}\boldsymbol{\epsilon}]\mathbf{j})\\&+([\boldsymbol{\kappa}\boldsymbol{\epsilon}]\boldsymbol{\kappa}')(\boldsymbol{\epsilon}\mathbf{j})]+l'(l-1)(\boldsymbol{\kappa}\boldsymbol{\epsilon}')([\boldsymbol{\kappa}\boldsymbol{\epsilon}']\mathbf{j})\\&+l(l'-1)(\boldsymbol{\kappa}'\boldsymbol{\epsilon})([\boldsymbol{\epsilon}\boldsymbol{\kappa}']\mathbf{j})+\tfrac{1}{2}(l-1)(l'-1)\\&\times(\boldsymbol{\kappa}\boldsymbol{\kappa}')([\boldsymbol{\kappa}\boldsymbol{\kappa}']\mathbf{j}),\end{aligned}\tag{25}$$

$$\begin{aligned}2\,\mathrm{Im}\,\mathbf{B}\mathbf{j}=&\,\xi'\{l([\boldsymbol{\epsilon}\boldsymbol{\kappa}'][\boldsymbol{\epsilon}\mathbf{j}])+\tfrac{1}{2}(l-1)[\boldsymbol{\kappa}\boldsymbol{\kappa}'][\boldsymbol{\kappa}\mathbf{j}]\}\\&+\xi\{l'(2(\boldsymbol{\epsilon}\boldsymbol{\epsilon}')([\boldsymbol{\kappa}\boldsymbol{\epsilon}][\boldsymbol{\epsilon}'\mathbf{j}])+[\boldsymbol{\epsilon}'\boldsymbol{\kappa}][\boldsymbol{\epsilon}'\mathbf{j}])\\&+\tfrac{1}{2}(l'-1)(2(\boldsymbol{\kappa}'\boldsymbol{\epsilon})([\boldsymbol{\kappa}\boldsymbol{\epsilon}][\boldsymbol{\kappa}'\mathbf{j}])+[\boldsymbol{\kappa}'\boldsymbol{\kappa}]\\&\times[\boldsymbol{\kappa}'\mathbf{j}])\}.\end{aligned}\tag{26}$$

Equations (24)–(26) imply that there are two mechanisms by which an initially unpolarized atom becomes oriented in the process of three−photon scattering. In the first three terms in (24), the orientation terms ($\sim\mathbf{j}$) disappear at $\xi=0$ and $\xi'=0$ but describe normal orientational phenomena in fields with a finite degree of circular polarization. Here we note that when the photon **e** is only circularly polarized, the orientation terms $\sim\xi'\boldsymbol{\kappa}'\mathbf{j}$ and $\sim\mathrm{Im}\,\mathbf{B}\mathbf{j}$ vanish from the cross section; they contribute to the cross section only if the photon **e** is either linearly or elliptically polarized. The term with $\mathrm{Re}\,\mathbf{B}\mathbf{j}$ containing the coefficient $a_6\sim\mathrm{Im}\{Q_{10}Q_{12}^*\}$ is finite even for $\xi=\xi'=0$ and describes dissipation-induced orientation in two linearly polarized light fields. Equation (24) becomes much simpler in this case:

$$\begin{aligned}2\widetilde{M}(\boldsymbol{\epsilon},\boldsymbol{\kappa},\boldsymbol{\epsilon}',\boldsymbol{\kappa}',\mathbf{j})=&\,a_2^{(1/2)}+(a_4^{(1/2)}+a_5^{(1/2)})(\boldsymbol{\epsilon}\boldsymbol{\epsilon}')^2\\&-2a_6^{(1/2)}(\boldsymbol{\epsilon}\boldsymbol{\epsilon}')([\boldsymbol{\epsilon}\boldsymbol{\epsilon}']\mathbf{j}).\end{aligned}\tag{27}$$

We see that the orientation term has the same vector structure (23b) as in the case of two-photon scattering. For collinear light beams the orientation vector points in the direction of light propagation and is the longest when the angle between the direction of linear photon polarization amounts to $\pi/4$. Orientation appears also in the case of an unpolarized photon $\mathbf{e}'$ ($\xi'=l'=0$), but for a photon **e** with a nonzero degree of linear polarization l. The corresponding results can easily be obtained as a particular case of Eqs. (24)–(26). Here the orientation vector is perpendicular to the plane containing the vectors $\boldsymbol{\epsilon}$ and $\mathbf{k}'$, is the longest when the angle between these vectors is $\pi/4$, and vanishes in the case of collinear light beams. If the photon **e** is also unpolarized, then orientation vanishes (see Eq. (23a)), in contrast to the case of two-photon scattering. This fact is specific only to the case of two identical photons, so that if all three photons are different, orientation emerges for unpolarized photons and is described by vector combinations of the type

$$(\boldsymbol{\kappa}_1\boldsymbol{\kappa}_2)(\boldsymbol{\kappa}_1\boldsymbol{\kappa}_3)([\boldsymbol{\kappa}_2\boldsymbol{\kappa}_3]\mathbf{j})$$

plus the terms obtained via permutations of the subscripts 1, 2, and 3. As for the numerical value of the dissipation-induced orientational effects, the same ideas concerning the circular and elliptical dichroism effects are valid here: at above-threshold frequencies the magnitude of the effect is of order unity, and in the resonant-frequency range the case of dipole-forbidden one- or two-photon resonances is the most promising for observing dissipation-induced orientational phenomena.

The present work was sponsored by the Russian Fund for Fundamental research (Grant No. 96-02-16251-a) and the Fundamental Natural Science Grant Center at the St. Petersburg State University (Grant No. 95-0-5.3-25).

APPENDIX A: CALCULATING TENSOR PRODUCTS OF VECTORS

In using the quantum angular-momentum technique to separate the kinematic and dynamic factors in the process cross sections containing several vectors (photon polarization vectors, particle momenta, etc.) one is forced to deal with scalar products of irreducible tensors composed of these vectors. Such constructions in the most general case are

$$(\{\ldots\{\mathbf{a}\otimes\mathbf{a}_1\}_{i_1}\otimes\,..\,\mathbf{a}_k\}_l\{\ldots\{\mathbf{b}\otimes\mathbf{b}_1\}_{j_1}\otimes\ldots\mathbf{b}_{k'}\}_l)\tag{A1}$$

and, being scalars, can br expressed in terms of combinations of scalar and mixed products of vectors $\mathbf{a}_i$ and $\mathbf{b}_j$. The respective formulas for (A1) containing three and four vectors can be found in Ref. 15. Below we describe the procedure of simplifying (A1) for arbitrary k, k', and l.

Let us first introduce a special notation for the equal vectors, where the rank of this tensor coincides with the number of the constituent vectors:

$$\{\ldots\{\{\mathbf{a}\otimes\mathbf{a}\}_2\otimes\mathbf{a}\}_3\otimes\ldots\mathbf{a}\}_{jm}\equiv\{\mathbf{a}\}_{jm}\,.\tag{A2}$$

We also list the properties of irreducible tensors, which can be verified by applying the rules governing the change of the coupling scheme in tensor products.[15]

(1) All tensors whose ranks coincide with the number of the constituent vectors are independent on the coupling scheme of these vectors. This assertion becomes obvious if we turn to the following relation (R_{i-1} and S_j are arbitrary tensors):

$$\{\{R_{i-1}\otimes\mathbf{a}\}_i\otimes S_j\}_{i+j,m}=\{R_{i-1}\otimes\{\mathbf{a}\otimes S_j\}_{j+1}\}_{i+j,m}\,.\tag{A3}$$

(2) If the rank of the tensor product is equal to the difference of ranks of the constituent tensor, the coupling scheme for these tensors can be changed according to the following relationship:

$$\{\{R_{i-1}\otimes\mathbf{a}\}_i\otimes S_j\}_{j-i,m}=\{R_{i-1}\otimes\{\mathbf{a}\otimes S_j\}_{j-1}\}_{j-i,m}\,.\tag{A4}$$

(3) For the scalar product of the tensors $\{\mathbf{a}\}_{jm}$ and $\{\mathbf{b}\}_{jm}$ we have the following relation:

$$(\{\mathbf{a}\}_j\{\mathbf{b}\}_j)=P_j(x)\,\frac{j!}{(2j-1)!!}\,(ab)^j,\tag{A5}$$

which follows from the well-known representation of the spherical function $Y_{jm}(\mathbf{a}/a)$ in terms of $\{\mathbf{a}\}_{jm}$ (Ref. 15) and the addition theorem for $Y_{jm}(\mathbf{a}/a)$; here $x=\mathbf{ab}/ab$ and P_j is the Legendre polynomial.

Let us now prove that for each tensor construction the following is true:

$$\{\ldots\{\{\mathbf{a}\otimes\mathbf{a}_1\}_{i_1}\otimes\mathbf{a}_2\}_{i_2}\otimes\ldots\mathbf{a}_k\}_{i_k m_k} = \hat{O}(\mathbf{a};\mathbf{a}_1,\ldots,\mathbf{a}_k)\{\mathbf{a}\}_{i_k m_k}, \tag{A6}$$

where $\hat{O}$ is a scalar differential operator containing the vectors $\mathbf{a}_1,\ldots,\mathbf{a}_k$ and the gradient operator $\nabla_{\mathbf{a}}\equiv\partial/\partial\mathbf{a}$. Indeed, such an operator can be built for each set $i_1,\ i_2,\ldots,i_k$ if we employ the following relations:

$$\{\{\mathbf{a}\}_l\otimes\mathbf{b}\}_{l-1\ m}=\frac{a^{2l+1}(\mathbf{b}\nabla_{\mathbf{a}})}{\sqrt{(2l-1)(2l+1)}}\ a^{-2l+1}\{\mathbf{a}\}_{l-1\ m},$$

$$\{\{\mathbf{a}\}_l\otimes\mathbf{b}\}_{l+1\ m}=\frac{1}{l+1}\ (\mathbf{b}\nabla_{\mathbf{a}})\{\mathbf{a}\}_{l+1\ m}, \tag{A7}$$

$$\{\{\mathbf{a}\}_l\otimes\mathbf{b}\}_{lm}=-\frac{i}{\sqrt{l(l+1)}}\ (\mathbf{b}[\mathbf{a}\nabla_{\mathbf{a}}])\{\mathbf{a}\}_{lm},$$

which can be verified if we allow for the rule of changing the coupling scheme.

These results are sufficient for simplifying constructions of type (A1). Using (A7), we can combine the vectors $\mathbf{a}_1,\ldots,\mathbf{a}_k$ and $\mathbf{b}_1,\ldots,\mathbf{b}_{k'}$ into a scalar (i.e., containing only scalar and mixed products) differential operator (see (A6)) acting on the scalar product (A5) with $j=l$. Calculating the result of the action of this operator on the Legendre polynomial written explicitly in terms of the scalar products $\mathbf{ab}$, we arrive at the final result. When the number of the vectors $\mathbf{a}_i$ and $\mathbf{b}_i$ is small, the most widely used are the formulas that follow from (A7) with $l=1,\,2$:

$$\{\{\mathbf{a}\otimes\mathbf{a}\}_2\otimes\mathbf{b}\}_{2m}=-\frac{i}{\sqrt{6}}\ (\mathbf{b}[\mathbf{a}\nabla_{\mathbf{a}}])\{\mathbf{a}\otimes\mathbf{a}\}_{2m},$$

$$\{\mathbf{a}\otimes\mathbf{b}\}_{2m}=\frac{1}{2}\ (\mathbf{b}\nabla_{\mathbf{a}})\{\mathbf{a}\otimes\mathbf{a}\}_{2m},$$

$$\{\{\mathbf{a}\otimes\mathbf{a}\}_2\otimes\mathbf{b}\}_1=\frac{a^5}{\sqrt{15}}\ (\mathbf{b}\nabla_{\mathbf{a}})\ \frac{1}{a^3}\ \mathbf{a}.$$

This procedure can be illustrated by the following example:

$$\begin{aligned}
&(\{\mathbf{a}\otimes\mathbf{a}_1\}_2\{\mathbf{b}\otimes\mathbf{b}_1\}_2)\\
&\quad=\tfrac{1}{6}(\mathbf{a}_1\nabla_{\mathbf{a}})(\mathbf{b}_1\nabla_{\mathbf{b}})a^2b^2P_2(\cos\,\theta)\\
&\quad=\tfrac{1}{4}(\mathbf{a}_1\nabla_{\mathbf{a}})(\mathbf{b}_1\nabla_{\mathbf{b}})\{(\mathbf{ab})^2-\tfrac{1}{3}a^2b^2\}\\
&\quad=\tfrac{1}{2}(\mathbf{a}_1\nabla_{\mathbf{a}})\{(\mathbf{ab})(\mathbf{b}_1\mathbf{a})-\tfrac{1}{3}a^2(\mathbf{b}_1\mathbf{b})\}\\
&\quad=\tfrac{1}{2}(\mathbf{ba}_1)(\mathbf{b}_1\mathbf{a})+\tfrac{1}{2}(\mathbf{ab})(\mathbf{a}_1\mathbf{b}_1)-\tfrac{1}{3}(\mathbf{bb}_1)(\mathbf{aa}_1),
\end{aligned}$$

which coincides with the result of Ref. 15.

In conclusion we note that this technique of transforming scalar products of the form (A1) can be adopted in simplifying tensors R_{kq} of low rank k that are constructed from vectors. For instance, to simplify the tensor R_{1q} of rank 1 we must first calculate the scalar product $(R_1\mathbf{r})$, where $\mathbf{r}$ is an arbitrary vector, and then obtain an explicit expression for R_{1q} by applying the operator $(\nabla_{\mathbf{r}})_q$ to $(R_1\mathbf{r})$.

APPENDIX B: INVARIANT ATOMIC PARAMETERS FOR THREE-PHOTON PROCESSES

If for the combinations of the atomic parameters Q_{xy} we employ the notation

$$\alpha_1=\frac{1}{2}\ (\sqrt{3}Q_{12}+\sqrt{5}Q_{11}),\quad \alpha_2=\frac{1}{2}\ (\sqrt{3}Q_{12}-\sqrt{5}Q_{11}),$$

$$\alpha_3=\frac{1}{\sqrt{3}}\ (\sqrt{5}Q_{10}-Q_{12}),\quad \beta_1=\frac{1}{2}\ (Q_{22}+\sqrt{3}Q_{21}),$$

$$\beta_2=\frac{1}{2}\ (Q_{22}-\sqrt{3}Q_{21}),\quad \gamma_1=|Q_{01}|^2,\quad \gamma_2=\frac{1}{7}\ |Q_{32}|^2,$$

we can write the atomic factors f_i and g_i in (14) as follows:

$$\begin{aligned}
f_0&=\tfrac{1}{6}(\gamma_1+\gamma_2+\tfrac{8}{15}|\beta_1+\beta_2|^2-\tfrac{8}{15}\ \mathrm{Re}(\beta_1\beta_2^*)),\\
f_1&=\tfrac{1}{15}(|\alpha_3|^2-\gamma_2-|\beta_1+\beta_2|^2),\\
f_2&=\tfrac{1}{6}(-\gamma_1+\gamma_2+\tfrac{4}{15}|\beta_1+\beta_2|^2+\tfrac{8}{15}\ \mathrm{Re}(\beta_1\beta_2^*)),\\
f_3&=\tfrac{1}{6}(-\gamma_1+\gamma_2+\tfrac{4}{15}[|\beta_1-\beta_2|^2-3|\beta_2|^2]),\\
f_4&=\tfrac{1}{15}(|\alpha_2|^2-\gamma_2-|\beta_1|^2),\\
f_5&=\tfrac{1}{6}(-\gamma_1+\gamma_2+\tfrac{4}{15}[|\beta_1-\beta_2|^2-3|\beta_1|^2]),\\
f_6&=\tfrac{1}{15}(|\alpha_1|^2-\gamma_2-|\beta_2|^2),\\
g_1&=\tfrac{2}{15}(\alpha_2^*\alpha_3-\gamma_2+|\beta_1|^2+\beta_1^*\beta_2),\\
g_2&=\tfrac{2}{15}(\alpha_1\alpha_3^*-\gamma_2+|\beta_2|^2+\beta_1^*\beta_2),\\
g_3&=\tfrac{1}{3}(\gamma_1+\gamma_2-\tfrac{4}{15}[|\beta_1-\beta_2|^2+3\beta_1\beta_2^*]),\\
g_4&=\tfrac{2}{15}(\alpha_1^*\alpha_2-\gamma_2-\beta_1\beta_2^*).
\end{aligned}$$

The parameters p_i and q_i in (16) have the following form:

$$\begin{aligned}
p_0&=\tfrac{1}{30}(8\gamma_2+|\alpha_1|^2+|\alpha_2|^2+|\beta_1|^2+|\beta_2|^2),\\
p_1&=\frac{1}{15}\left(-3\gamma_2+\mathrm{Re}\sum_{i=1}^{3}\ \alpha_1\alpha_3^*\right),\\
p_2&=\tfrac{1}{45}[12\gamma_2+3\ \mathrm{Re}(\alpha_1\alpha_2^*)-5\ \mathrm{Re}(\beta_1\beta_2^*)],\\
p_3&=\tfrac{1}{60}(5\gamma_1-3\gamma_2-2|\alpha_2|^2+\tfrac{2}{3}|\beta_1+2\beta_2|^2),\\
p_4&=\tfrac{1}{60}(5\gamma_1-3\gamma_2-2|\alpha_1|^2+\tfrac{2}{3}|2\beta_1+\beta_2|^2),\\
q_1&=\tfrac{1}{15}[2\gamma_2-\alpha_3(\alpha_1^*+\alpha_2^*)-|\beta_1+\beta_2|^2],\\
q_2&=\tfrac{1}{30}[-5\gamma_1-3\gamma_2+\tfrac{4}{3}|\beta_1-\beta_2|^2-\alpha_1\alpha_2^*\\
&\qquad+2\ \mathrm{Re}(\beta_1\beta_2^*)+i6\ \mathrm{Im}(\beta_1\beta_2^*)].
\end{aligned}$$

*E-mail: manakov@thp.vucnit.voronezh.su

---

[1] N. B. Delone and V. P. Krainov, *Multiphoton Processes in Atoms*, Springer, Berlin (1994).

[2] N. L. Manakov, V. D. Ovsiannikov, and L. P. Rapoport, Phys. Rep. **141**, 319 (1986); A. G. Fainshtein, N. L. Manakov, V. D. Ovsiannikov, and L. P. Rapoport, Phys. Rep. **210**, 112 (1992).

[3] S. Baier, A. N. Grum-Grzhimailo, and N. M. Kabachnik, J. Phys. B **27**, 3363 (1994).

[4] N. A. Cherepkov, V. V. Kuznetsov, and V. A. Verbitskii, J. Phys. B **26**, 1221 (1995).

[5] J. Berakdar, H. Klar, A. Huetz, and P. Selles, J. Phys. B **26**, 1463 (1993).
[6] N. M. Kabachnik and V. Schmidt, J. Phys. B **28**, 233 (1995).
[7] N. L. Manakov, S. I. Marmo, and A. V. Meremianin, J. Phys. B **29**, 2711 (1996).
[8] N. L. Manakov, S. I. Marmo, and V. V. Volovich, Phys. Lett. A **204**, 42 (1995).
[9] N. L. Manakov, S. I. Marmo, and A. G. Faĭnshteĭn, Zh. Éksp. Teor. Fiz. **108**, 1569 (1995) [JETP **81**, 860 (1995)].
[10] N. L. Manakov, Zh. Éksp. Teor. Fiz. **106**, 1286 (1994) [JETP **79**, 696 (1994)].
[11] N. L. Manakov and A. V. Meremianin, in: *Contributed Papers of the 5th ECAMP*, Edinburgh (1995), Part II, p. 635.
[12] V. B. Berestetskiĭ, E. M. Lifshitz, and L. P. Pitaevskiĭ, *Quantum Electrodynamics*, 2nd ed., Pergamon Press, Oxford (1982).
[13] V. L. Strizhevskiĭ and V. M. Klimenko, Zh. Éksp. Teor. Fiz. **53**, 244 (1967) [Sov. Phys. JETP **26**, 163 (1968)].
[14] N. L. Manakov and V. D. Ovsyannikov, Opt. Spektrosk. **48**, 651 (1980) [Opt. Spectrosc. **48**, 359 (1980)].
[15] D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskiĭ, *Quantum Theory of Angular Momentum*, World Scientific, Singapore (1987).
[16] N. L. Manakov and V. D. Ovsyannikov, Opt. Spektrosk. **48**, 838 (1980) [Opt. Spectrosc. **48**, 463 (1980)].
[17] M. Ya. Agre and N. L. Manakov, J. Phys. B **29**, L5 (1996).

Translated by Eugene Yankovsky

Instability and disruption of precision cumulation of cavities and mass flows in the field of gravitational and inertial forces and other dipole perturbations

G. A. Askar'yan and I. V. Sokolov

Institute of General Physics, Russian Academy of Sciences, 117942 Moscow, Russia

(Submitted 30 September 1996)
Zh. Éksp. Teor. Fiz. **111**, 2001–2015 (June 1997)

The effect of gravitational and inertial forces on the precision cumulation of cavities, fast mass flows, and detonation and shock waves is studied in stages. The possible effects of these forces on the motion of masses and waves (the velocity, direction and structure of the waves can change) and on the properties of the medium in front of the waves that influence the change in the parameters and refraction of the waves are noted. It is shown that it is advantageous to perform precision cumulation experiments in zero-gravity. [S1063-7761(97)00606-9]

1. INTRODUCTION

In the last few years, cumulation—one of the most interesting directions of science and technology—has entered new fields of small-scale processes and ultrahigh densities, temperatures, and pressures. Besides the numerous studies of laser-driven fusion, it is sufficient to mention the successful attempts to obtain bursts of thermonuclear neutrons with fast converging mass flows accelerated by the cumulation by an explosive initiated by a simultaneous detonation of its outer surface,[1] the production of superstrong magnetic fields by compression of liners which sweep up the magnetic field with and without plasma, the already long-matured possibilities of producing microcritical masses of fissioning materials and micronuclear explosions,[2] particle acceleration, meson and neutrino production, and so on.[2,3]

Actually, the idea here, by switching to small foci and high energy densities, is to increase the energy density a thousandfold and more above the cumulation levels previously achieved in the explosion, military, and ultrasonic fields. However, a number of difficulties due to the destruction of the symmetry of the cumulation process arise on the path to such small-scale processes. Most of these difficulties are trivial and can be eliminated technologically. We shall be interested primarily in the influence of gravitational and inertial forces, which is difficult to eliminate, on the processes which occur.

The acceleration g due to gravitational and inertial forces can vary over wide limits—from $g=g_0=10^3$ cm/s^2 (earth's gravitational field) up to $g\sim(10-10^4)g_0$, in the case when a body moving with a high velocity decelerates in a gaseous or dense medium, and down to $g=0$ in zero-gravity.

The effect of these forces can be expressed as a direct action on the elements of the medium which are entrained into the cumulation process through a change in the initial properties of the medium through which the shock or detonation wave travels. Finally, the action of the force on the exterior part of the flow can cause the disturbance to accumulate and instability to develop in the interior regions.[4]

The above-mentioned instability has been found to be characteristic for many cumulation systems and it does not reduce to Rayleigh–Taylor or Richtmyer–Meshkov instabilities. It leads to growth of the dipole harmonics of the disturbance and is very closely related to the force of gravity—on the one hand the instability can be excited by the action of the gravitational force and on the other the presence of the dipole disturbance, even if it is present for different reasons (for example, presence of a gradient of the external pressure), can itself often be taken into account by introducing into the equations an effective, gravitational-type mass force. In this latter case the effective acceleration g can reach $(10^5-10^6)g_0$.

It is easy to estimate the effect of an acceleration g on the very simple ''impact'' system in which noninteracting particles, initially arranged on a sphere with an initial radius R_0, fall toward the center with velocity V_0. They will collide at a point displaced downwards from the center of symmetry by the distance

$$\Delta x=\frac{1}{2}g\tau^2,\quad \tau=R_0/V_0,$$

with velocities differing by $\delta V\sim 2V_0\Delta x/R_0$. When colliding particles or waves are used to compress a target, such a dissonance can be manifested at the last stage of compression where a high degree of symmetry of the action is essential.

In hydrodynamic cumulation the situation is fundamentally different in the sense that the analogous estimate $\Delta x\sim g\tau^2$, where τ is a characteristic time, may be incorrect even in order of magnitude. In addition, gravity can shift the cumulation point upwards and not downwards.

The initial deformation due to gravity is also a direct action of gravity. This effect is not discussed in the present paper, though the problem of eliminating it appears to be nontrivial in some cases. For example, consider a bubble in a liquid. The collapse of the bubble is an example of a cumulative flow. The bubble can be assumed to be spherical only if it is small compared with the so-called capillary constant $k=\sqrt{2\sigma/g\rho_0}$, where σ is the surface tension and ρ_0 is the density. For water $k\cong 0.4$ cm. if the initial bubble size is $\geqslant 0.1$ cm, then the distortions of the bubble shape cannot be regarded as small, especially since, as will be shown below, these distortions grow as the bubble collapses. If an attempt is made to stabilize the initial shape and the initial position of the bubble with the aid of shells, then the problem arises of taking account of the effect of the residues of these shells on the final result of the cumulation process.

2. EFFECT OF THE GRAVITATIONAL FORCE ON THE SYMMETRY OF COLLAPSE OF A BUBBLE IN A LIQUID

One of the oldest cumulation problems is the collapse of a bubble in a liquid under the action of an external pressure[5,6] or resonance vibrational action down to small sizes by a pulse train.[7] Such dynamic bubbles have long proved to be sources of strong shock waves in acoustics, ultrasonic technology, and maraine practice. They are used for working or breaking down surfaces, accelerating chemical reactions, acting on a medium and on physical and biophysical processes and structures, and so on. But this was up to kbar pressures. How will such bubbles behave in attempts to obtain thousands of times higher pressures?

First, we shall estimate the initial and final pressures and bubble sizes at which high collapse rates should be expected.

For purposes of estimation, let us consider an incompressible liquid. The condition of incompressibility is not satisfied at pressures $\leqslant$ Mbar for a liquid such as water, but there is, for example, mercury, whose atoms have many electrons and for this reason are poorly compressed, and there are also ''hard'' solids, which behave at high pressures as a liquid with a low compressibility. In the the compression of cavities in bodies which are initially solid but become liquid at high pressures, difficulties due to initial deformation and initial motion on account of buoyancy do not arise.

We shall present an expression, derived from energy considerations and helpful for further derivations, for the rate of collapse. Let an empty spherical bubble with initial radius R_0 and instantaneous radius $R(t)$ collapse under a pressure P_0 applied to the distant outer surface of an incompressible liquid with density ρ_0. The work performed by the external pressure equals at each moment in time the product of P_0 by the change in the volume bounded by the outer surface of the liquid, equal, taking compressibility into account, the compression volume of the cavity:

$$A=4\pi P_0(R_0^3-R^3)/3.$$

This work equals the kinetic energy K of the liquid set in motion during the collapse of the bubble. Since the velocity $v(r)$ of a spherically symmetric flow in an incompressible liquid satisfies at any given moment in time the relation $vr^2=\text{const}$, where r is the distance from the point to the center, integration over the volume gives

$$K=4\pi\rho R^3\,\frac{\dot{R}^2}{2}.$$

Equating A and K, we easily obtain the well-known expression for the velocity of the cavity walls:[5,6]

$$\dot{R}=\sqrt{\frac{2P_0}{3\rho_0}\left(\frac{R_0^3}{R^3}-1\right)}.$$

For small R we obtain $\dot{R}\sim R^{3/2}$ and $K=\text{const}$. Therefore a bubble in an incompressible liquid acts like an ideal converter that converts the work performed by an external pressure into the kinetic energy of the liquid localized in a small volume $\sim 4\pi R^3$. The acceleration of the inner surface increases as

$$\ddot{R}=\dot{R}\,\frac{d\dot{R}}{dR}=-\frac{P_0R_0^3}{\rho_0R^4}.$$

The pressure, which equals zero on the inner surface, grows in the direction into the liquid with gradient $\partial P/\partial r\sim\rho_0\ddot{R}$, so that a pressure

$$P\sim R\partial P/\partial r\sim P_0(R_0/R)^3.$$

arises at distances $\sim R$ from the surface. As $R\to 0$, we have $P\gg P_0$ by a fantastic factor (but, recall, that the incompressible liquid model is imperfect). The growing inertial compression of the near-surface layer of matter under a high pressure $P\gg P_0$ is analogous to the first (adiabatic) stage of compression of matter by a programmable laser pulse, under which the compression grow and accelerates up to the peak action on the supercompressed matter.

It is easy to take account of the effect of the gas present inside a bubble. For collapse velocities greater than the velocity of sound in the gas the formation of a shock wave in the gas must be taken into account (numerical calculations of such a flow were performed in Ref. 8). If the bubble boundary compressing the gas accelerates toward the center, then, as is well known from the theory of the pinch effect in plasma, the gas swept up by the shell is localized in a thin layer, which is separated by the shock front from the unperturbed gas, near the shell. The kinetic energy of this layer can be set equal to $4\pi\rho_g(R_0^3-R^3)\dot{R}^2/3$, where ρ_g is the initial gas density. The internal energy of the gas compressed by the strong shock wave is usually of the order of the kinetic energy. For $R\leqslant R_m=R_0(\rho_g/\rho_0)^{1/3}$ the effective mass of the liquid set in motion becomes less than the mass of the gas. A large fraction of the kinetic energy is transferred to the gas and the increase in the energy density is limited by the finiteness of the mass of the gas. The converging shock wave in the gas on these scales should detach from the boundary of the bubble and its further cumulation[5,6,9] is not associated with the motion of the bubble walls but inherits all the distortions of the bubble shape.

The existence of a critical scale R_m leads to an important limitation on the degree of volume compression of the gas in the bubble: It cannot be much greater than the ratio of the initial densities of the liquid and the gas or, equivalently, the maximum achievable average gas density in the bubble is of the order of the density of the liquid compressing it. This limitation is more stringent than the case, ordinarily analyzed in hydroacoustics, of adiabatic compression of an air bubble in water.

Taking $P_0\sim 1$ Mbar, $\rho_g\sim 10^{-2}$ g/cm^3, $\rho_0\sim 10$ g/cm^3, and $R_0\sim 10$ cm for purposes of estimation, we would obtain on scales of the order of 1 cm collapse velocities of the order of 10^7 cm/s and compressed gas density ~ 10 g/cm^3, which could be of interest for controlled thermonuclear fusion.

We took these figures as typical for the example of explosive initiation, keeping in mind that both the compressibility of the liquid and, generally speaking, a number of plasma effects occurring at the liquid–plasma interface will have to be taken into account. For all its limitations, the incompressible-liquid model nonetheless makes it possible to take into account simply the effect of gravity on the bubble cumulation process.

Cumulation can break down because of not only the direct effect of gravity but also the pressure differential on the outer surface of the liquid, i.e., a small asymmetry in the distribution of P_0. As a result of such an asymmetry, the total force which acts on the liquid and under which the liquid as a whole acquires an inertial acceleration, may be different from zero. The action of such an acceleration can be described, on the basis of the principle of equivalence, by an effective gravitational acceleration which is determined, as is easily understood, by the expression

$$\mathbf{g}_{eq}=-\frac{\oint P_0\mathbf{n}dS}{\int\rho_0 dV},$$

where the surface integral is taken over the outer surface of the liquid, the unit vector points in the direction of the normal to the surface into the liquid, and the volume integral extends over the entire volume of the liquid. Another possibility is the case when the pressure on the outer surface, though uniform, is switched on in different sections of the surface at different times. In this case, the effective acceleration acts during the switch-on time.

To study the action of the acceleration of the gravitational force, real or effective, on the collapse of a bubble let us examine the following problem. Let an empty spherical cavity of radius R_0 be present initially in an incompressible liquid which is initially at rest at $t=0$. The pressure in the liquid sets the liquid in motion and the cavity collapses. We shall find the flow that arises, taking account of the gravitational force.

Since the gravitational force is a potential force and the vorticity in the liquid is initially zero, the flow will be a potential flow. The conditions of irrotationality and incompressibility and Euler's equation are[6]

$$\mathbf{u}=\text{grad }\varphi,\quad \Delta\varphi=0,\quad \frac{\partial\mathbf{u}}{\partial t}+(\mathbf{u}\cdot\nabla)\mathbf{u}=-\frac{1}{\rho}\text{ grad }P+\mathbf{g}. \tag{2.1}$$

Here $\mathbf{u}$, φ, P, and ρ are, respectively, the velocity, the velocity potential, the pressure, and the density. We integrate Euler's equation over space, assuming that at large distances from the bubble the velocity potential vanishes and the pressure distribution equals the hydrostatic pressure distribution:

$$P\big|_{|x|\to\infty}=P_0+\rho\mathbf{g}\cdot\mathbf{x},$$

where $P_0=\text{const}$ and $\mathbf{x}$ is the radius vector. Then

$$\frac{\partial\varphi}{\partial t}+\frac{1}{2}\text{ grad}^2\ \varphi+\frac{P-P_0}{\rho}=\mathbf{g}\cdot\mathbf{x}. \tag{2.2}$$

We introduce a system of spherical coordinates r,θ initially centered at the center of the spherical cavity, and let the $\theta=0$ axis be oriented upwards (in a direction opposite to the gravitational force). The azimuthal angle is not introduced, since the flow obviously does not depend on it. The equation of the bubble surface can be represented in the form $r=r_m(\theta,t)$. We expand the function $1/r_m(\theta,t)$ in a series of Legendre polynomials $P_n(\cos\theta)$ and at the same time expand the velocity potential, which is a harmonic function (satisfying Laplace's equation), in a series in spherical harmonics:

$$\frac{1}{r_m}=\sum_{n=0}^{\infty}a_n(t)P_n(\cos\theta),$$

$$\varphi=\sum_{n=0}^{\infty}b_n(t)P_n(\cos\theta)\left(\frac{1}{r}\right)^{n+1}. \tag{2.3}$$

To obtain equations for the coefficients a_n and b_n we employ the identity

$$\frac{\partial}{\partial t}\left(\frac{1}{r_m}\right)=-\left(\frac{1}{r_m}\right)^2\left(\frac{\partial\varphi}{\partial r}+\frac{\partial}{\partial\theta}\left(\frac{1}{r_m}\right)\frac{\partial\varphi}{\partial\theta}\right)\Bigg|_{r=r_m}. \tag{2.4}$$

To derive Eq. (2.4) it is necessary to transform temporarily on the surface $r=r_m$ to Lagrangian coordinates (t,ξ) (see Ref. 6), which possess the property that a point with fixed ξ on the boundary coincides at all times with the same element of the liquid and therefore moves in space with velocity $\mathbf{u}$. Using the well-known differential expression for transforming from the coordinates (θ,l) to coordinates (ξ,l),

$$\left(\frac{\partial r_m}{\partial t}\right)_\theta=\left(\frac{\partial r_m}{\partial t}\right)_\xi-\left(\frac{\partial r_m}{\partial\theta}\right)_\theta\left(\frac{\partial\theta}{\partial t}\right)_\xi,$$

and expressing the velocity of an element of the liquid lying on the boundary (the derivatives at constant ξ), in terms of the velocity potential at the boundary, we obtain Eq. (2.4). Substituting the expansions (2.3) into Eqs. (2.4) and (2.2) and recalling that $P=0$ holds at $r=r_m$, we obtain the system of equations

$$\sum_{n=0}^{\infty}\frac{da_n}{dt}P_n(\cos\theta)=r_m^{-2}\left(\sum_{n=0}^{\infty}\frac{b_n(n+1)}{r_m^{n+2}}P_n(\cos\theta)-\sum_{n=1}^{\infty}a_n\frac{dP_n(\cos\theta)}{d\theta}\times\sum_{n=1}^{\infty}\frac{b_n}{r_m^{n+1}}\frac{dP_n(\cos\theta)}{d\theta}\right), \tag{2.5}$$

$$\sum_{n=0}^{\infty}\frac{db_n}{dt}\frac{P_n(\cos\theta)}{r_m^{n+1}}+\frac{1}{2}\left(\sum_{n=0}^{\infty}\frac{(n+1)b_nP_n(\cos\theta)}{r_m^{n+2}}\right)^2+\frac{1}{2}\left(\sum_{n=1}^{\infty}\frac{b_n}{r_m^{n+2}}\frac{dP_n(\cos\theta)}{d\theta}\right)^2=\frac{P_0}{\rho}-gr_m\cos\theta. \tag{2.6}$$

In Eqs. (2.5) and (2.6) r_m must be expressed in terms of the series (2.3). Then all derivatives of the Legendre polynomials must be represented as sums of Legendre polynomials, and equating to zero the total term of the expansion in front of each polynomial gives an infinite system of ordinary differential equations for a_n and b_n. In practice, this can be done only for the case when all modes with $n\neq 0$ can be regarded as a perturbation of the main motion with $n=0$.

This is admissable if the initial disturbances of the bubble shape and the velocity field are small; an additional condition for the gravitational force to be weak is presented below. Linearizing Eqs. (2.5) and (2.6) with respect to the quantities a_n, b_n $(n\geqslant 1)$ gives the system

$$\dot{a}_0 = a_0^4 b_0, \quad a_0 \dot{b}_0 + \frac{1}{2} b_0^2 a_0^4 = \frac{P_0}{\rho}, \tag{2.7}$$

$$\dot{a}_1 = 2b_1 a_0^5 + 4b_0 a_0^3 a_1, \quad \dot{b}_0 a_1 + \dot{b}_1 a_0^2 + 2b_1 b_0 a_0^5 + 2b_0^2 a_1 a_0^3 = -g/a_0, \tag{2.8}$$

$$\dot{a}_n = (n+1) b_n a_0^{n+4} + 4b_0 a_n a_0^3, \quad \dot{b}_0 a_n + \dot{b}_n a_0^{n+1} + (n+1) b_n b_0 a_0^{n+4} + 2b_0^2 a_0^3 a_1 = 0 \quad (n\geqslant 2). \tag{2.9}$$

The equations (2.7) describe the well-known process of spherical cumulation of a bubble and have a solution of the form

$$b_0 = \sqrt{\frac{2P_0}{3\rho}\left(\frac{R_0^3}{a_0} - \frac{1}{a_0^4}\right)}, \quad t = \int_{R_0^{-1}}^{a_0} \frac{da_0}{a_0^4 b_0(a_0)}. \tag{2.10}$$

Near the center (for $a_0 \gg R_0^{-1}$) we obtain from Eqs. (2.10)

$$R(t) = a_0^{-1}(t) \approx \left(\frac{25(\tau - t)^2 P_0 R_0^3}{6\rho}\right)^{1/5},$$

$$\tau = \sqrt{\frac{3\rho}{2P_0}} \int_{R_0^{-1}}^{\infty} \frac{da_0}{\sqrt{R_0^3 a_0^7 - a_0^4}}. \tag{2.11}$$

Here we have introduced the notation τ for the collapse time.

Next, Eqs. (2.8) describe in a linear approximation the growth of a dipole disturbance on account of the action of the gravitational force on the bubble collapse process. Eliminating from Eq. (2.8) b_0 and b_1 with the aid of Eqs. (2.7) and (2.8) we find that in this approximation the action of the gravitational force on the bubble causes the bubble to collapse, and the upward displacement $\zeta(t) = -a_1 R^2$ of the bubble is described by the simple equation

$$\frac{d}{dt}\left(R^3 \frac{d\zeta}{dt}\right) = 2gR^3. \tag{2.12}$$

The equation (2.12) has a clear physical meaning: A bubble with reduced mass $2\pi R^3 \rho/3$ (this expression is given in Ref. 6) rises under the action of the buoyancy $-4\pi R^3 \rho \mathbf{g}/3$. The reduced mass vanishes over the collapse time τ, and the impulse of the buoyancy is finite and equals

$$\int_0^\tau \frac{4\pi}{3} gR^3 dt = \frac{5\pi}{6} g\tau R_0^3. \tag{2.13}$$

Therefore the rise velocity grows as

$$\frac{d\zeta}{dt} = \frac{5R_0^3 \tau g}{4R^3},$$

i.e., much more rapidly than the collapse velocity

$$\frac{dR}{dt} \sim \frac{R_0^{5/2}}{R^{3/2}\tau}.$$

The linear approximation is applicable for $R \gg l \sim R_0^{1/3}(g\tau^2)^{2/3}$. This condition holds at the initial stage of collapse for bubbles which are not too large (such that $g\rho R_0 \ll P_0$). However, at distances $R \sim l \sim R_0^{1/3}(g\tau^2)^{2/3}$ near the center the rise and collapse velocities are equal, and the model of spherical cumulation of a bubble is no longer applicable. We note that $l \gg g\tau^2$, and for not too small initial bubble sizes the limits imposed on cumulation by the gravitational force can appear earlier than other limiting factors (compressibility, counterpressure).

Finally, the equations (2.9) describe the growth of initial disturbances of the bubble shape. To find the growth rate of the disturbances near the center, we substitute the relations (2.11) into Eq. (2.9). We obtain for the amplitude $\zeta_n = -a_n R^2$ the asymptotic behavior

$$\zeta_n \propto R^\alpha, \quad \alpha = -\frac{1}{4} \pm i\sqrt{\frac{24n - 25}{16}}. \tag{2.14}$$

These modes are gravity waves, in which the acceleration of the bubble surface during collapse ($\ddot{R}$) plays the role of the acceleration of gravity. For large n the frequency of these waves is proportional, as usual, to the square root of the wave number. On account of the compression of the surface occupied by the waves, the amplitude increases

$$|\zeta_n| \propto R^{-1/4},$$

but much more slowly than the mode $n = 1$ examined above, for which $\zeta \propto R^{-1/2}$. The increase in the amplitude of the gravity waves is a consequence of the conservation of an adiabatic invariant; in the process the energy of the waves grows as $R^{-5/2}$.

It seems to us that this result is of great interest for the general problem of the stability of cumulative flows since, first, the growing perturbations are perturbations on a Rayleigh–Taylor stable surface (the inertial acceleration is directed in the direction of the heavy liquid) and, second, the dipole mode of the perturbations has the maximum growth velocity.

In the linear approximation the gravitational force affects the highest modes only through the deformation of the initial shape of the bubble. One can attempt to eliminate these initial imperfections of the cavity shape by confining the cavity in a thin light spherical shell, but the residues of the shell can degrade cumulation.

If the dipole mode is excited by an effective gravitational force, then it must be kept in mind that this effective force can change in time and then the gravitational acceleration in the expression (2.13) for the impulse of the buoyancy cannot be removed from the integrand. Moreover, if the effective force varies not only in magnitude but also in direction, then the impulse of the buoyancy can vanish so that the instability will be appreciably suppressed. We also note that an instability of the dipole mode can develop in the absence of a gravitational force if initially the bubbles have a small velocity relative to the liquid. In this latter case the role of the impulse of the buoyancy is simply transferred to the initial momentum of the bubble (product of velocity by the reduced mass).

Finally, we shall estimate the comparative role of the effect of a counterpressure and the gravitational force on cumulation during the rapid (nonadiabatic) collapse of gas in a bubble. Comparing the expressions for the radius R_m, at which the counterpressure starts to have an effect (see above), and for l shows that the gravitational force dominates when

$$g \geqslant \frac{P_0}{\rho_0 R_0} \sqrt{\frac{\rho_g}{\rho_0}}. \tag{2.15}$$

At megabar pressures the effect of the gravitational force itself ($g = g_0$) is hardly important. However, when the pressure on the outer surface of the liquid is initiated by means of an explosion, substantial inertial forces can arise since the detonation wave producing this pressure is itself subjected to the effect of the gravitational force.

3. ACTION OF A GRAVITATIONAL FORCE ON A CONVERGING DETONATION WAVE

In the case when the motion occurs in media produced by a converging detonation or shock wave, an estimate of the direct action gives $\Delta \sim g(R_0/D)^2$, where D is the detonation velocity. But taking the effect of gravitational forces on the detonation process in detail gives a sharper dependence on g, especially for gaseous media. We shall prove this.

We consider first a plane detonation wave propagating upwards or downwards in a gas with specific heat ratio γ in a gravitational field. Let us assume that the initial equilibrium state is isothermal and the heat of reaction does not depend on the pressure ahead of the front. Then the detonation velocity determined from the Chapman–Jouguet (CJ) condition does not depend on the height.

We use the Lagrangian equations of motion

$$\frac{\partial \rho}{\partial t} + \frac{\rho^2}{\rho_0} \frac{\partial u}{\partial \xi} = 0, \tag{3.1}$$

$$\frac{\partial u}{\partial t} + \frac{1}{\rho_0} \frac{\partial P}{\partial \xi} = g, \tag{3.2}$$

$$\frac{\partial (P/\rho^\gamma)}{\partial t} = 0. \tag{3.3}$$

Here we have introduced the Lagrangian coordinate ξ (see Ref. 6), which we shall take to be the initial vertical coordinate of a fluid particle measured from the location of detonation initiation in the direction of propagation of the detonation. Correspondingly, we take $g = \mp |\mathbf{g}|$ for upward or downward motion. Let the subscript ''0'' label values of the parameters ahead of the wave front and the subscript ''1'' label values behind the wave front. Since P_0 and ρ_0 depend on ξ,

$$P_0 \propto \rho_0 \propto \exp(g\xi(\partial \rho/\partial P)_T),$$

P_1 and ρ_1 also depend on ξ and for this reason the entropy conserved in the flow behind the front is different for different values of ξ:

$$P(t,\xi)/\rho^\gamma(t,\xi) = P_1(\xi)/\rho_1^\gamma(\xi) \neq \text{const}. \tag{3.4}$$

Let $c = \sqrt{\gamma P/\rho}$ be the speed of sound. We write the equations of motion in the characteristic form

$$\left(\frac{\partial}{\partial t} \pm \frac{\rho c}{\rho_0} \frac{\partial}{\partial \xi}\right)\left(u \pm \frac{2}{\gamma_1} c\right) = -\frac{\rho c^2 \partial}{\rho_0 \gamma \partial \xi} \ln(\rho_1^{\gamma/(\gamma-1)} P_1^{-1/(\gamma-1)}) + g. \tag{3.5}$$

On account of the entropy gradient on the right-hand side of Eq. (3.5) there appears, besides g, a specific gradient acceleration

$$-\frac{\rho c^2}{\rho_0 \gamma} \frac{\partial}{\partial \xi} \ln(\rho_1^{\gamma/(\gamma-1)} P_1^{-1/(\gamma-1)}) = -\frac{\rho c^2}{\rho_0} g \left(\frac{\partial \rho_0}{\partial P_0}\right)_T, \tag{3.6}$$

which is directed opposite to $\mathbf{g}$ and whose magnitude for a strong detonation wave is many times greater than g. The effect of this acceleration is similar to that of the curvature of the front on the flow behind a detonation wave, described by the same equation as Eq. (3.5) but with the right-hand replaced by Kcu, where K is the curvature. Like the curvature, the gradient acceleration leads either to overcompressed detonation or singularities near the front, depending on the sign of the gradient acceleration.

The effect of the gravitational force on a plane wave can be easily calculated taking into account the smallness of the degree of overcompression, which is characterized by the small deviations δ of the quantities behind the front from the values prescribed by the Chapman–Jouguet condition. The well-known properties of a normal detonation wave satisfying the Chapman–Jouguet condition lead to simplifications. In this state the entropy behind the front has its maximum value and the velocity D of the wave has its minimum value compared with all states which are permitted by the conditions at the discontinuity. Therefore the deviations $\delta(P_1/\rho_1^\gamma)$ and δD are second-order infinitesimals, which makes it possible to obtain, by means of the law of conservation of mass at the discontinuity ($\delta\rho_1(D_{CJ} - u_{1CJ}) = \delta u_1 \rho_1$), the relations

$$\frac{\delta c_1}{c_{1CJ}} = \frac{\delta u_1}{D_{CJ} - u_{1CJ}} \frac{\gamma - 1}{2}, \quad \frac{\delta P_1}{P_{CJ}} = \frac{2\gamma}{\gamma - 1} \frac{\delta c_1}{c_{1CJ}},$$

$$\frac{\delta \rho_1}{\rho_{1CJ}} = \frac{\delta u_1}{D_{CJ} - u_{1CJ}} \tag{3.7}$$

(The index CJ designates the values of quantities which satisfy the Chapman–Jouguet condition).

Further, the velocity of a normal wave relative to the gas behind the wave equals the local velocity of sound:

$$D_{CJ} - u_{1CJ} = c_{1CJ},$$

which makes it possible to eliminate u_{1CJ} from Eq. (3.7). Moreover, since for a normal detonation no perturbations in the region of the flow behind the front can overtake the front, it follows that only processes occurring in a narrow region, whose width d is of the same order of smallness as the degree of overcompression, behind the wave front can affect the propagation of a weakly overcompressed detonation wave.

Integrating Eq. (3.5) on the C characteristic (i.e., Eq. (3.5) with the lower signs taken on the left-hand side) over a region of width d, we obtain, taking account of Eq. (3.7), the relation

$$u-2c/(\gamma-1)=\text{const},$$

which holds, to within second-order infinitesimals, near the front. Eliminating with the aid of this relation u from Eq. (3.5) on the C_+ characteristic and expressing ρ in the latter equation in terms of c with the aid of the relation (3.4), we have to the same accuracy

$$\frac{\partial U}{\partial t}+U\,\frac{\partial U}{\partial \xi}=-\,\frac{\gamma+1}{4\gamma}\,U^2 g\left(\frac{\partial \rho_0}{\partial P_0}\right)_T,$$
$$U=D_{CJ}\left(\frac{c}{c_{CJ}}\right)^{(\gamma+1)/(\gamma-1)}. \tag{3.8}$$

Since for strong waves we have $D_{CJ}^2 g(\partial\rho_0/\partial P_0)_T \gg g$, the term g in Eq. (3.5) is dropped on switching to the relations (3.8).

The behavior of the solutions of Eq. (3.8) depends strongly on the sign of the gradient acceleration. For $g>0$ (the wave moves downwards) the acceleration is negative, and the C_+ characteristics are outgoing with respect to the detonation front. The detonation is normal, and the solution of Eq. (3.8) near the front can be sought in the form of a wave propagating with velocity D_{CJ}

$$D_{CJ}-U=F(D_{CJ}t-\xi),$$

where F is the desired function. Determining F with the aid of Eq. (3.8), we have

$$D_{CJ}-U=\sqrt{\frac{\gamma+1}{2\gamma}\,D_{CJ}^2 g\left(\frac{\partial\rho_0}{\partial P_0}\right)_T (D_{CJ}t-\xi)}. \tag{3.9}$$

The velocity of the gas and the velocity of sound behind the wave front, which moves downwards, remain constant. The pressure behind the front equals the Chapman–Jouguet pressure,

$$P_1=P_{CJ}.$$

For an upward propagating wave, when $g<0$ holds and the right-hand side of Eq. (3.8) is positive, the C_+ characteristics are incoming with respect to the front and modify the conditions at the discontinuity, giving rise to overcompression of the detonation. Exact integration of Eq. (3.8) along characteristics with the initial condition $U|_{t=0}=U_0(\xi)$ gives

$$\frac{1}{U}=\frac{1}{U_0(\xi_0)}+\frac{\gamma+1}{4}\,gt\left(\frac{\partial\rho_0}{\partial P_0}\right)_T, \tag{3.10}$$

$$\xi=\xi_0+\frac{4}{(\gamma+1)g}\left(\frac{\partial P_0}{\partial\rho_0}\right)_T \ln\left[1+\frac{\gamma+1}{4}\,g\left(\frac{\partial\rho_0}{\partial P_0}\right)_T tU_0(\xi_0)\right], \tag{3.11}$$

where ξ_0 is conserved along each characteristic and equals the value of ξ on this characteristic at time $t=0$. Measuring time from the moment of detonation initiation, we obtain (see Ref. 6) that U_0 is different from zero only in a very narrow region of values of ξ_0. The width of this region is negligibly small, which makes it possible to drop the term ξ_0 in Eq. (3.11).

Retaining in Eq. (3.11) the first two terms in the series expansion in the small quantity $gtD_{CJ}(\partial\rho_0/\partial P_0)_T$, which takes account of the curvature of the characterisitcs under the action of the gradient acceleration, we find the value of U_0 for the characteristic arriving at the time t at the detonation front

$$\frac{1}{U_0}=\frac{1-[(\gamma+1)/8]gt(\partial\rho_0/\partial P_0)_T}{D_{CJ}}. \tag{3.12}$$

Substituting the expression (3.12) into Eq. (3.10) makes it possible to find the value of U and then all other hydrodynamic quantities behind the detonation front. The expression for δP_1 is especially simple:

$$P_1=P_{CJ}+\delta P_1,\quad \frac{\delta P_1}{P_{CJ}}=\frac{1}{4}\,D_{CJ}t|\mathbf{g}|\left(\frac{\partial\rho_0}{\partial P_0}\right)_T. \tag{3.13}$$

The degree of overcompression of the upward moving wave increases with the distance $D_{CJ}t$ which the wave has traveled. We find that the pressure behind upward and downward moving detonation waves is different, the difference accumulating as the waves continue to propagate. This effect, unimportant in practice for a plane wave, can play an important role for converging cylindrical or spherical detonation waves, since the wave elements arriving at the target from different sides bring different pressures.

Let us now examine a spherical converging detonation wave. The detonation on the entire front is overcompressed because of the curvature, and the gradient acceleration leads to a pressure difference $\delta P(R)$, which increases with convergence, at the top and bottom points of the front. At the initial stage of convergence the degree of overcompression is small and the expressions (3.13) can be used. For $R\ll R_0$ (R and R_0 are the instantaneous and initial radii of the front), the effect of the gravitational force is small but δP continues to increase because the cumulation process is unstable with respect to multidimensional disturbances. Treating the overcompressed detonation wave near the center as a shock wave, we estimate the rate of growth of the disturbances as[10]

$$\delta P/P\propto R^{-1}.$$

For a distance l at which $\delta P/P\sim 1$ we obtain the estimate

$$l\sim R_0^2|\mathbf{g}|\left(\frac{\partial\rho_0}{\partial P_0}\right)_T. \tag{3.14}$$

The foregoing analysis can be applied to aerosol explosions, initiated over a spherical surface, of a large volume in the atmosphere. These explosions have a great deal of energy, powerful shock waves, and highly uniform convergence, which makes it possible to use them for experiments on controlled thermonuclear fusion and for generating strong magnetic fields. The expressions presented here can be used to estimate the maximum size of the focus of convergence of detonation waves. We call attention to the fact that the maximum achievable pressure increases with decreasing cloud size

$$P \propto (R_0/l)^{0.8} \propto R_0^{-0.8}.$$

The estimate (3.14) is also applicable for a converging spherical shock wave, if the wave is produced so that at $R=R_0$ the wave velocity is the same along the entire wave front.

In a liquid explosive, besides taking into account exactly all effects determining the quantity $(\partial P/\partial \rho)_T$, it may be necessary to take into consideration as well the possible dependence of the detonation velocity on the hydrostatic pressure P_0. Then $(\partial \rho_0/\partial P_0)_T$ is replaced everywhere by $(\partial \rho_0/\partial P_0)_T + 3\rho_0(\partial \ln D_{CJ}/\partial P_0)_T$.

The action of the gravitational force on a converging detonation wave has been observed in remarkable experiments on gas-dynamic thermonuclear fusion.[1] A spherical volume of a liquid explosive with radius $R_0 \sim 0.5$ m was initiated simultaneously over the entire surface. The neutron yield from a target placed near the center was measured. The maximum neutron yield (far from the expected value) corresponded to a target position $\cong 0.1$ mm below the center of symmetry. The quantity $\lambda = (\partial \ln D_{CJ}/\partial P_0)_T \approx 5 \cdot 10^{-9}$ cm^2/dyne was measured separately.

The detonation front should move over the convergence time upwards by $\Delta_r \sim R_0^2 \rho_0 g \lambda \sim 0.1$ mm. To achieve simultaneity of action the target should be displaced upwards by Δ_r. On the other hand, we have shown that when the wave converges, the pressure at the bottom of the front is higher than at the top. Hence it follows that, conversely, to achieve pressure uniformity on the target, the target should be displaced downwards from the center so that the maximum convergence of the detonation wave in the radial direction would be achieved at the points of the front with minimum $\delta P(R)$. Therefore, in searching for the maximum, an empirical adjustment of the target position from simultaneous multidimensional action to nonsimultaneous uniform action does not make it possible to eliminate completely the influence of the gravitational force. In addition to displacing the target downward for pressure uniformity (obviously, even lower than in the experiment performed with the maximum neutron yield), simultaneity of front arrival at the target must also be ensured (by programming a delay in the initiation at the bottom with respect to the top or a small change in the shape of the charge).

4. ON THE SPECIAL ROLE OF THE HYDRODYNAMIC INSTABILITY OF THE DIPOLE DISTURBANCE MODE IN INERTIAL CONTROLLED THERMONUCLEAR FUSION

Before Ref. 1 appeared, virtually no attention was devoted in the literature to the role of the hydrodynamic instability of cumulative flows with respect to dipolar disturbances. But the mode with $n=1$ has a higher growth rate than all other modes and its instability impedes the realization of controllable thermonuclear fusion in large-scale cumulative setups, specifically, gas-dynamic thermonuclear fusion.

Let us consider the problem of the collapse of a liquid spherical shell with outer radius R_e and inner radius $R(t)$ (see the notation in Sec. 2) under the action of the pressure

$$P_e = P_0 + \delta P_1 \cos \theta,$$

which includes a small perturbation with $n=1$ $(\delta P_1 \ll P_0)$. For simplicity, let $R_e \gg R$. Then, first, $R_e \approx$ const and, second, the problem is described by the same equations as in Sec. 1 but with **g** replaced by the effective inertial acceleration, equal to

$$\mathbf{g}_{\text{eff}} = \frac{\delta P_1}{\rho R_e} \mathbf{e}_\theta. \tag{4.1}$$

(The unit vector $\mathbf{e}_0$ is directed along the ray $\theta=0$, so that the effective acceleration is oriented along the gradient of the external pressure).

Indeed, the velocity potential (2.3) in a doubly-connected volume of liquid can be supplemented by terms of the form $G(t) r \cos \theta$. Substituting the potential modified in this manner into Eq. (2.2) with $\mathbf{g}=0$ gives Eq. (2.2), in which the notation φ is maintained for the series (2.3) and the effective acceleration equals $-G'(t)\mathbf{e}_\theta$. The boundary condition for Eq. (2.2) at the outer surface gives

$$G = -\delta P_1 t/(\rho R_e),$$

which leads to Eq. (4.1).

Using the results of Sec. 2, we obtain an estimate for the final compression

$$l \sim (g_{\text{eff}} \tau^2 R_0^2)^{1/3} \sim \left(\frac{\delta P_1 R_0^4}{P_e R_e} \right)^{1/3}. \tag{4.2}$$

If the collapse of a shell is used to compress matter located in an interior cavity, then the maximum achievable pressure of the compressed material is

$$P_{\max} \sim \frac{P_e R_0^3}{l^3} \sim \frac{P_e^2 R_e}{\delta P_1 R_0}. \tag{4.3}$$

The hydrodynamic instability of the dipole mode plays an especially large role in large-scale cumulation setups because the growth of the pressure perturbation on account of the effect of the gravitational force in the outer layers of the system leads to large values of the effective acceleration in the interior regions.

Let us consider the following rough idealized model of explosive thermonuclear fusion. Imagine a thick, DT-filled, spherical shell consisting of a heavy material, which we shall consider to be an incompressible liquid. The shell is compressed by a converging detonation wave in a spherical explosive setup of radius $R_b \gg R_e$. We shall find the condition under which the action of the gravitational force on the system does not impede ignition.

The conditions that the thermonuclear temperature must reach $T \sim 3$ keV and the thermonuclear α particles must heat the DT by $\rho_{\text{DT}} l \sim 0.3$ g/cm^2 require that

$$l P_{\max} \geq 1 \quad \text{Gbar} \cdot \text{cm}. \tag{4.4}$$

We shall determine δP_1 from the relations employed in the derivation of Eq. (3.14):

$$\delta P_1 \sim P_e \frac{R_b^2}{R_e L}, \quad \frac{1}{L} = \left[3\rho_0 \left(\frac{\partial \ln D_{CJ}}{\partial P_0} \right)_T + \left(\frac{\partial \rho_0}{\partial P_0} \right)_T \right] g. \tag{4.5}$$

The effective acceleration (4.1), corresponding to Eq. (4.5) and acting in the liquid shell, is several orders of magnitude greater than the gravitational acceleration, so that the direct action of the gravitational force on the cumulation of the shell is completely unimportant. From Eqs. (4.2)–(4.5) we obtain the necessary condition

$$P_e \left(\frac{L^2 R_e^4 R_0}{R_b^4} \right)^{1/3} \geqslant 1 \quad \text{Gbar}\cdot\text{cm}. \tag{4.6}$$

For $P_e \sim 1$ Mbar, $R_b \sim 10^2$ cm, $R_e \sim R_0 \sim 10^1$ cm, and $L \sim 10^5$ cm (the last value corresponds to the value of $(\partial \ln D_{CJ}/\partial P_0)_T$ presented in Ref. 1) the condition (4.6) is not satisfied, i.e., the effect of the gravitational force on the flow makes ignition of a thermonuclear reaction impossible, though one-dimensional estimates neglecting instability of the $n=1$ mode are completely favorable.

It follows from Eq. (4.6) that for fixed R_e and R_0 it makes no sense to increase the outer radius too much, since the pressure on the target then grows more slowly than $R_b^{4/3}$, which prevents the inequality (4.6) from being satisfied.

We have seen that the effect of the gravitational force in many cases is to degrade the quality of precision cumulation. On account of the seriousness of the associated limitations, it is relevant to remark that the condition for ideal cumulation and maximum neutron yield is achieved in zero-gravity. Even short-time weightlessness (dropping from a low height) after averaging or decay of sound waves arising when the force of gravity is removed could be helpful.

5. CONCLUSIONS

1. A specific instability with respect to disturbances of the dipolar mode operates in large cumulative devices intended for achieving ultrahigh energy densities. This instability can be excited by the effect of the gravitational force.

2. This instability is capable of limiting the neutron yield and impeding ignition to such an extent that it may be useful to perform experiments under conditions of zero gravity (for example, during free fall).

We thank E. E. Meshkov for a discussion of Ref. 1.

This work was supported by the Russian Fund for Fundamental Research.

[1] A. N. Anisimov, A. N. Arinin, M. I. Arifov *et al.*, in *Abstracts of Reports at the International Conference ''3rd Zababakhin Scientific Lectures''* [in Russian], Chelyabinsk-70, 1991, pp. 20, 21, 96.

[2] G. A. Askar'yan, V. A. Namiot, and M. S. Rabinovich, JETP Lett. **17**, 424 (1973).

[3] G. A. Askar'yan, JETP Lett. **28**, 296 (1978); Usp. Fiz. Nauk **128**, 727 (1979) [Sov. Phys. Usp. **22**, 666 (1979)].

[4] G. A. Askar'yan and I.V. Sokolov, JETP Lett. **63**, 786 (1996).

[5] E. I. Zababakhin and I. E. Zababakhin, *Unbounded Cumulation Phenomena* [in Russian], Nauka, Moscow, 1989.

[6] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics*, Pergamon Press, N. Y. [Russian original, Nauka, Moscow, 1985].

[7] R. I. Nigmatulin, V. Sh. Shagapov, N. K. Vakhitova, and R. T. Lekhi, Dokl. Akad. Nauk **341**, 37 (1995).

[8] C. C. Wu and P. H. Roberts, Phys. Rev. Lett. **70**, 3424 (1993).

[9] I. V. Sokolov, Usp. Fiz. Nauk **160**, 143 (1990) [Sov. Phys. Usp. **33**, 960 (1990)].

[10] K. V. Brushlinskiĭ, Preprint, M. V. Keldysh Institute of Applied Mathematics, Soviet Academy of Sciences, 1980.

Translated by M. E. Alferieff

Phase equilibrium in a current-carrying liquid conductor in *Z*-pinch geometry

V. S. Vorob'ev and S. P. Malyshenko

Institute of High Temperatures, Russian Academy of Sciences, 127412 Moscow, Russia

(Submitted 10 September 1996)
Zh. Éksp. Teor. Fiz. **111**, 2016–2029 (June 1997)

A cylindrical liquid conductor at constant temperature with an axisymmetric dc current generating an azimuthal magnetic field is studied. The magnetic field introduces an additional field component that is different for each phase in the total chemical potential. It is shown that treatment of the field components can be reduced to equating the chemical potentials of the matter in different phases without a field at different pressures. As a result of the shift appearing in the parameters of phase equilibrium, the liquid not heated up to boiling coexists with gas at pressures corresponding to the metastable branch of the state parameters. For this reason, there exists a limiting value of the current above which phase equilibrium is impossible in this geometry. In the latter case, a transition of the system into a dispersed state (drops in vapor) followed by mechanical breakup of the fluid column on account of the energy stored in the magnetic field occur. The expansion rate, drop size, and current at which this phenomenon occurs are estimated. The phenomenon can occur in the final stage of the fast electrical explosion of conductors. [S1063-7761(97)00706-3]

1. INTRODUCTION

As is well known, strong electric or magnetic fields penetrating into a solid strongly influence the thermodynamic properties of the solid. The general thermodynamic relations for a material in a field are presented in Refs. 1 and 2. An elegant example of the application of these relations is given in Ref. 3, where the problem of the condensation of an electrically charged drop is examined. At the same time, the question of the effect of a strong magnetic field, produced by an electric current flowing through the matter, on the thermodynamics of phase transitions has never been fully clarified, though this problem is of practical importance for investigations of the compression of matter in pinch-effect systems and the electrical explosion of conductors and in a number of other applied problems.

Recently a series of works has appeared (see, for example, Refs. 4–6) on the possible effects associated with the displacement of the phase equilibrium in a magnetic field. Nonetheless, no systematic analysis of this question has been performed.

To study the effect of the magnetic field generated by an internal current on the thermodynamic properties, the simplest magnetic field, which is realized in *Z*-pinches, is ordinarily studied. It is presupposed that a constant axisymmetric current, which generates an azimuthal magnetic field, flows along a liquid conductor at constant temperature. The density in such a field varies only in the radial direction. This makes the problem much simpler than in the case of an arbitrary magnetic field, when the density varies in all coordinate directions.

The present paper also employs this magnetic-field geometry. The state of a single-phase conductor is studied first. An expression for the total thermodynamic potential of the matter–field system is written as a functional of the density and field strength on the basis of the general equilibrium condition for a body in an external magnetic field. Minimizing this function with respect to the appropriate variables yields a conservation condition for the local value of the total chemical potential and a mechanical equilibrium condition for a current-carrying liquid cylinder in which the electrodynamic force balances the pressure gradient. The behavior of the thermodynamic functions is analyzed on the basis of these conditions. It is shown that the role of the magnetic field reduces not only to producing a nonuniform pressure profile in the liquid, but in the presence of a field the total chemical potential acquires an additional component which depends on the field intensity, the matter density, and the radius of the conductor. The thermodynamic relations and the equation of state in the absence of a field hold at each point for the local values of the chemical potential of the matter and the pressure and temperature dependence of the density.

The question of phase transitions in a current-carrying conductor is also studied. The corresponding functional of the total thermodynamic potential is written down for a two-phase system. Minimizing this functional with respect to the position of the interface yields new phase-equilibrium conditions. The essence of these conditions is that the total chemical potentials of the phases must be equal at the interface. The chemical potentials of the matter in phases under identical pressures and temperatures may be different. The equality of the total chemical potentials of the phases in the matter–field system can be formally reduced to an equality of the chemical potentials of the matter in the phases without a field but under different pressures. The pressure difference is due to the difference of the magnetic pressures in the phases. The possibility that the chemical potentials of two states of matter which have the same temperatures and different pressures are equal is analyzed qualitatively as well as with the aid of the van der Waals equation. It is shown that the densities and pressure of the coexisting phases are higher in the presence of a field than with no field. In the case when the system is in thermodynamic equilibrium, the less dense phase is in a state which in the absence of a field corresponds to a metastable state (supercooled vapor). However, when this phase reaches a point on the spinodal, phase equilibrium in this geometry becomes impossible. The system must expand sharply and pass into a new state. It is shown that this state can be a finely dispersed mixture of liquid droplets in vapor. The drop sizes and expansion rates are estimated. It is conjectured that a transition of this kind could be directly responsible for the sharp increase in the radius and resistance in the final stage of the electrical explosion of conductors which occurs in fast regimes.[7–10] An estimate of the critical current for this phenomenon gives a relation which was previously known as an empirical relation and indicated that the action integral[9] is proportional to parameters which depend only on the properties of the conductor. It is shown that the coefficient of proportionality in this relation is determined by the critical temperature of the metal.

2. EQUILIBRIUM OF A ONE-PHASE LIQUID-CONDUCTOR IN A MAGNETIC FIELD

Let us examine an infinitely long axially uniform cylindrical column of unmagnetized liquid material at constant temperature T in a cylindrical coordinate system (r,φ,z) with the z axis oriented along the axis of the filament. There are no striction effects in the liquid during quasistatic processes.

In such a filament all quantities depend only on the radial coordinate r. Current flows along the axis and has only one component $j_z \equiv j$. In the general case the current density j can depend on r. The magnetic field generated by this current will have only an azimuthal component $H_\varphi \equiv H$. The components j and H are related by the corresponding Maxwell equation

$$j=\frac{c}{4\pi r}\frac{d}{dr}(rH), \tag{1}$$

where c is the speed of light. The solution of Eq. (1) with the boundary condition $H(0)=0$ has the form

$$H(r)=\frac{4\pi}{c}\frac{1}{r}\int_0^r jr'dr'. \tag{2}$$

The derivative of the magnetic-field intensity from Eq. (2) can be represented in the form

$$\frac{dH}{dr}=\frac{4\pi}{c}j-\frac{H}{r}. \tag{3}$$

In an external field the total chemical potential μ^{I} of the matter–field system, including, besides the chemical potential of the matter, a field component as well, must be constant.[1,2] Therefore, its total differential must vanish at equilibrium

$$\rho d\mu^{\mathrm{I}}=\rho d\mu+\frac{1}{c}jH\,dr=0, \tag{4}$$

where μ is the chemical potential per unit mass of the material in the absence of the field and ρ is the density of the material. The first term on the right-hand side of Eq. (4) is the change in the chemical potential of a volume element of the material. The second term is the change in the field part of the chemical potential, equal to the work performed by the electrodynamic forces acting on this element. Using the expression (3), the equation (4) can be rewritten in the form

$$\rho d\mu+\frac{1}{4\pi}HdH+\frac{H^2}{4\pi r}dr=0. \tag{5}$$

A somewhat different interpretation of the equilibrium conditions follows from Eq. (5). For a nonuniform liquid conductor in equilibrium in a field, the change in the chemical potential of a volume element of the liquid is compensated by a change in the energy of the magnetic field and the work performed by the magnetic tension forces directed toward the center of curvature of the flux lines.[11]

Since in our case the material depends on only one coordinate, the equation (4) can be easily integrated in r. The result is

$$\mu(r)+\frac{1}{c}\int_0^r vjH\,dr=\mu_0=\mu^{\mathrm{I}}, \tag{6}$$

where μ_0 is the chemical potential of the material on the axis for $H=0$ and $v=1/\rho$ is the specific volume. The equation (6) is the desired condition expressing that the chemical potential of the liquid-conductor–field system is constant. The integral on the left-hand side of Eq. (6) taken with the opposite sign is the potential of the mass density of the electrodynamic force performing work on a unit mass of the material:

$$A(r)=-\frac{1}{c}\int_0^r vjH\,dr. \tag{7}$$

The functional Ψ, related to the total thermodynamic potential Φ of the system by $\Psi=\Phi-\lambda m$, has the form

$$\Psi=2\pi\int_0^R r\,dr\rho\left[\mu(r)+\frac{1}{c}\int_0^r vjH\,dr-\lambda\right], \tag{8}$$

where R is the outer radius, m is the mass per unit length of the conductor, and λ is a undetermined Lagrange multiplier, introduced in the functional (8) so as to take account of the fact that the total mass of the body is conserved. Setting to zero the first variation of Ψ gives the necessary conditions for a minimum at a fixed temperature. Let δr be an arbitrary displacement of the material elements in Eq. (8). Then, for the first variation of Ψ to vanish it is necessary that

$$\frac{\delta\rho}{\delta r}\left[\mu(r)+\frac{1}{c}\int_0^r vjHdr-\lambda\right]+\rho\left[\frac{d\mu}{dr}+\frac{1}{c}vjH\right]=0. \tag{9}$$

Hence it follows that the expressions in both sets of brackets in Eq. (9) must vanish. The first condition gives the above-noted constancy of the chemical potential, where $\lambda=\mu_0$. The second condition gives, taking account of the thermodynamic relation $d\mu=vdP$, the condition of mechanical equilibrium of a liquid conductor in a field:

$$\frac{dP}{dr}=-\frac{1}{c}jH, \tag{10}$$

where P is the pressure of the matter. The meaning of this relation is that the pressure gradient in the liquid is balanced by Ampere's electrodynamic force.

So, the necessary conditions of equilibrium of a conductor in a magnetic field follow from the condition of a minimum for the functional (8).

Integrating Eq. (10) along r yields the radial pressure distribution

$$P(r)=P_0-\frac{1}{c}\int_0^r jH\,dr', \tag{11}$$

where P_0 is the pressure on axis. The quantity

$$P_m(r)=\frac{1}{c}\int_0^r jH\,dr' \tag{12}$$

is termed the magnetic pressure. It is essentially the potential of the volume density of the electrodynamic force. In the particular case $j=\text{const}$, the well-known expression

$$P_m(r)=\frac{\pi}{c^2}j^2r^2 \tag{13}$$

follows from Eq. (12). It follows from Eqs. (10) and (11) that the pressure cannot become an increasing function of radius for any current distribution. We also note that the pressure distribution is determined by the current distribution and does not explicitly depend on the state of the matter.

The change in the chemical potential can be expressed in terms of the thermodynamic variables by replacing $-jH/c$ in Eq. (10) by dP:

$$\mu(P(r),T)=\mu(P_0,T)-\int_{P(r)}^{P_0} v\,dP. \tag{14}$$

The formula (14) does not contain r explicitly and is a standard thermodynamic formula relating the chemical potentials to different pressures on an isotherm. When there is no current, the pressure is constant over the cross section, the integral on the right-hand side of Eq. (14) vanishes, and $\mu(P(r),T)=\mu(P_0,T)$. Since the function $\mu(P(r),T)$ is determined by the equation of state, which holds in the absence of a field, it increases monotonically with pressure or decreases with increasing radius. The distributions of the other thermodynamic parameters in a current-carrying liquid conductor can be obtained from Eq. (14).

Let us now examine the question of phase transitions in matter containing a current.

3. PHASE EQUILIBRIUM

Assume that phase separation has occurred. The high-density phase 1, which we conventionally term a liquid, lies inside a cylinder of radius r_1. It coexists with a low-density phase 2, occupying an outer annular layer of thickness $R-r_1$. We conventionally term it a gas. We shall neglect the curvature of the interface and capillary effects, and we assume only that an interphase boundary is established at $r=r_1$. We shall use subscripts 1 and 2 to denote the parameters of the first and second phases, respectively.

The functional Ψ for the two-phase system will have the form

$$\Psi=2\pi\left\{\int_0^{r_1} r\,dr\,\rho_1[\mu_1-A_1-\lambda_1]\right.$$
$$\left.+\int_{r_1}^{R} dr'r'\rho_2[\mu_2-A_2-\lambda_2]\right\}, \tag{15}$$

where A_2 is the potential of the mass density of the forces in the second phase,

$$A_2(r')=\frac{1}{c}\left[A_2(r_1)+\int_{r_1}^{r'} v_2j_2H_2dr\right], \tag{16}$$

and $A_2(r_1)$ is the value of this potential at the interphase boundary. Minimizing the functional (16) with respect to δr and $\delta r'$ yields the equilibrium conditions (6) and (10) for each phase. Then $\lambda_1=\mu_{10}$. The quantity λ_2 equals

$$\lambda_2=\mu_2(r_1)-A_2(r_1). \tag{17}$$

As is clear from the relations presented above, in a Z-pinch geometry it is impossible for the pressure to increase with r independently of the ratio of the currents j_1 and j_2. Within each phase the pressure decreases monotonically with increasing r. There is no pressure jump at the interphase boundary; this corresponds to the condition that the boundary is in mechanical equilibrium.

Since the total chemical potentials of each phase are constant, we write the total thermodynamic potential of the system in the form

$$\Phi=m_1\mu_1^{\mathrm{I}}+m_2[\mu_2(r_1)-A_2(r_1)], \tag{18}$$

where m_1 and m_2 are the masses of the phases. Minimizing the potential (18) with respect to m_1 under the condition that the total mass of both phases is constant ($dm_1=-dm_2$), we obtain the condition of equilibrium of a two-phase system:

$$\mu_1(r_1)-A_1(r_1)=\mu_2(r_1)-A_2(r_1). \tag{19}$$

As expected, the total chemical potentials of the phases (taking account of the correction for the field) are equal at the interphase boundary. The difference of the potentials of the mass density of the forces at the interphase boundary can be written in the form

$$A_1(r_1)-A_2(r_1)=\Delta\mu=-\frac{1}{c}\int_0^{r_1}(v_1j_1-v_2j_2)H\,dr. \tag{20}$$

Replacing the force differentials in Eq. (20) by pressure differentials according to the relation (10), we obtain the conditions of phase equilibrium in a different form

$$\mu_2(P+P_{m2},T)=\mu_1(P+P_{m1},T), \tag{21}$$

where P_{m1} and P_{m2} are the magnetic pressures in the first and second phases, respectively, at the point of contact of the phases and P is the pressure at the interface.

The relations (19)–(21) are the conditions of phase equilibrium in the presence of a magnetic field. It follows from Eq. (19) that if the potentials of the mass density of the forces at the interphase boundary are different, then the chemical potentials of the matter in the phases under identical temperatures and pressures are different. This circumstance indicates that phase equilibrium in external fields is

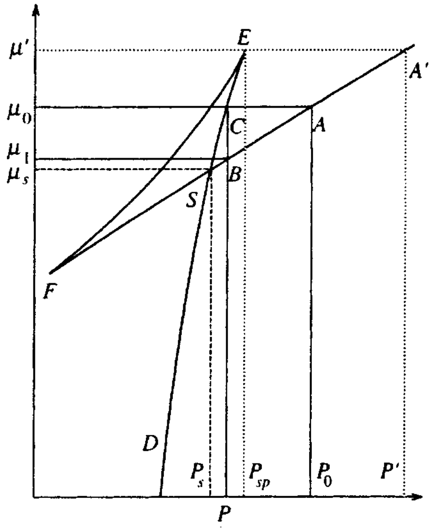


FIG. 1. Chemical potential versus pressure at constant temperature. The dotted lines mark the values of the pressure P_s and chemical potential μ_s in the absence of a field. Solid lines—values of the pressures (P and P_0) and chemical potentials (μ_1 and μ_0) corresponding to phase equilibrium in the presence of magnetic pressure; dotted lines—limiting values of the pressures (P_{sp} and P') and chemical potential (μ') at the center of the liquid for which phase equilibrium is still possible.

fundamentally different from equilibrium in uniform matter in the absence of a field. We note that if the force per unit mass acting on the interphase boundary varies continuously, then despite the presence of a field the chemical potentials of the matter in the phases will be the same at the interphase boundary. It is easy to show that this happens for a gravitational field. However, for the case considered here we have $A_1 \gg A_2$, since $j_1 \gg j_2$, and for this reason there is a field-associated shift of the phase-equilibrium parameters.

The relation (21) makes it possible to give a somewhat different interpretation of the results obtained. It follows from this equation that the equality of the total chemical potentials of the phases can be reduced to the equality of the chemical potentials of the matter in the phases under different pressures at a point where there is no field. As one can see from Eq. (21), this pressure difference depends on the magnetic pressure in the phases.[1)] In a current-carrying conductor the magnetic pressure satisfies $P_{m1} \gg P_{m2}$, so that the latter pressure can be neglected. This approximation will be used below, so that P_m will be the magnetic pressure of the first phase. Then Eq. (21) states that the chemical potential of the liquid on the axis of the conductor with pressure $P_0 = P + P_m$ equals the chemical potential of the gas at the interphase boundary with pressure P. In the limit $P_m \to 0$ the relations (19)–(21) degenerate into the equality of the chemical potentials of phases with identical temperatures and pressures.

The graphical interpretation of the conditions of equilibrium (19)–(21) is shown in Fig. 1, where a curve of the chemical potential of the material versus pressure at a fixed temperature below the critical temperature is constructed. The separate sections of this curve correspond to the following branches: DS—gas; SE—supercooled gas; SA—liquid; FS—superheated liquid; FE—absolutely unstable state. The point S corresponds to the conditions of phase equilibrium in the absence of a field. The corresponding pressure and chemical potential are designated as P_s and μ_s. The straight line $\mu = \mu_0 > \mu_s$ gives the value of the total chemical potential, which remains constant over the entire system. The point A of intersection of this straight line with the branch SA corresponding to the liquid determines the state of the liquid at the center with pressure P_0. The point C of intersection with the branch SE is the state of a supercooled gas at the interphase boundary with pressure P and the same value of the chemical potential, and $P_0 - P = P_m$. The straight line $P = \text{const}$ intersects the segment SA of the liquid state at the point B, which determines the parameters of the liquid on the interphase boundary at pressure P with chemical potential μ_1, where $\mu_0 - \mu_1 = \Delta\mu$. The pressure drop in the gas phase from the values P at the interphase boundary to P_R at the outer boundary is small, since it is due to the work performed by the electrodynamic forces in the gas phase, and is not shown in Fig. 1.

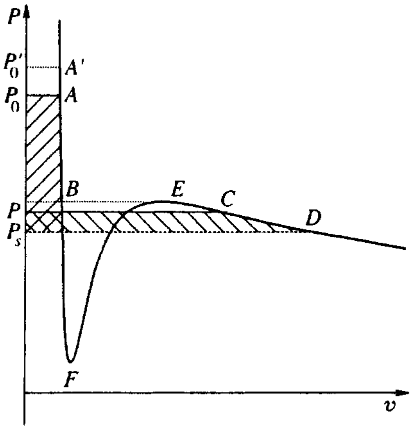


FIG. 2. Pressure versus volume. The hatched areas correspond to equality of the chemical potentials of the liquid at the center and gas on the interphase boundary. The dotted lines mark the limiting values of the liquid and gas pressures at which phase equilibrium is still possible.

The state of the material changes as follows. The pressure in the liquid phase drops from P_0 at the center to P at the interphase boundary. The segment AB describes the change in the chemical potential of the liquid. A jump in the chemical potential up to the point C on the gas phase branch occurs at the point B corresponding to P. Since the magnetic pressure in the gas phase is low, the state of the gas for $r > r_1$ corresponds to the point C.

It follows from analysis of Fig. 1 that for each temperature there exists a limiting value P' and corresponding values $\Delta\mu'$ and P_m' for which phase equilibrium is still possible. In Fig. 1 the dotted lines correspond to these values. In this case the state of the gas corresponds to the point E on the gas branch of the spinodal with pressure $P_{sp}(T)$. At this point the condition $(\partial P/\partial v)_{T,\mu} < 0$ of thermodynamic stability of the gas phase breaks down, and the liquid is not heated to boiling at any point.

Similar conclusions can be drawn on the basis of the $P - v$ phase diagram. A typical subcritical pressure isotherm is displayed in Fig. 2. In this figure, a pressure P above P_s

corresponds to equilibrium at the interphase boundary. Maxwell's rule of areas does not hold here, since the chemical potentials of the phases are different. The state of the liquid is described by the branch AB and the state of the gas is described by the point C. Equality of the areas of the curvilinear trapezoids corresponds to equality of the chemical potentials of the liquid at the center of the liquid and the gas at the interphase boundary. Hence, the area of the trapezoid between the straight lines $P=P_0$ and $P=P_s$ and the branch of the isotherm of the liquid phase should equal the area of the trapezoid formed by the straight lines P and $P=P_s$ and the branch of the gas phase.

One can see that equilibrium exists so long as $P<P_{sp}$ holds. The dotted curves represent the limiting values of the pressures for which equilibrium is still possible ($P=P_{sp}$). Phase equilibrium becomes impossible for $P>P_{sp}$. Figures 1 and 2 elucidate qualitatively the effect of a magnetic field on phase equilibrium.

At pressures above the critical pressure phase equilibrium becomes impossible in the geometry under study and the system should pass into a new state. In some sense such a situation arises in the investigation of the equilibrium state in the normal and superconducting phases in type-I superconductors.[1] As is well known, for some value of the external magnetic field a superconducting body cannot be in a superconducting or normal state. An intermediate state, in which a structure consisting of alternating layers of normal and superconducting phases is characteristic, is realized.

In our case, a two-phase system with a cylindrical interphase boundary should also pass into a new state with a different geometry. This transition can be initiated by spontaneous condensation of an unstable vapor phase at the interphase boundary. A decrease of the current density and magnetic pressure will destroy the existing balance of the pressure gradient in the liquid by Ampere forces, i.e., it will destroy the mechanical stability of the system and lead to disintegration of the conductor. It should be noted that the kinetics of such a transition is the subject of a separate investigation. It is important to underscore that the final equilibrium state of the system with practically no current can correspond only to a dispersed liquid–vapor system (liquid drops in vapor) with the fluid pressure equal to the pressure P_{sp} at the boundary of the cylinder prior to disintegration. In this case, the chemical potentials of a drop and the vapor in equilibrium with it equalize and become equal to the chemical potential of the liquid at the interphase boundary prior to the disintegration of the conductor. From this condition we find the gas pressure P_G after disintegration:

$$\int_{P_s}^{P_{sp}} v_1 dP = \int_{P_s}^{P_G} v_2 dP. \tag{22}$$

It follows from Eq. (22) that $P_G-P_s \approx (P_{sp}-P_s)v_1/v_2 \ll P_s$. Far from the critical point we have $v_1/v_2 \ll 1$, i.e., $P_G \approx P_s$. The pressure difference between the liquid and the gas is balanced by surface effects. The characteristic drop size can be found from Laplace's relation

$$r=\frac{2\sigma}{P_{sp}-P_G}, \tag{23}$$

where σ is the surface tension. The number of such drops per unit length of the cylinder is given by the expression

$$N=2\pi r_1^2 \Big/ \frac{4\pi}{3} r^3. \tag{24}$$

The difference of the chemical potentials of the gas and liquid which is present before the transition and is associated with the work performed by the electrodynamic forces goes into the work required to form the drop surfaces and imparting kinetic energy to the drops. Writing the corresponding energy balance, we obtain

$$2\pi r_1^2 \Delta\mu = N\left(4\pi r^2 \sigma + \frac{4\pi\rho_1}{3} r^3 \frac{V^2}{2}\right), \tag{25}$$

where V is the expansion velocity. Using the fact that $\Delta\mu \sim P_m/\rho_1$ and Eqs. (23) and (24), we obtain from Eq. (25)

$$V=\sqrt{2\frac{P_m-3(P_{sp}-P_s)/2}{\rho_1}}. \tag{26}$$

An estimate on the basis of the van der Waals equation shows that $P_{sp}-P_s \sim 0.2P_c$ at temperatures $\sim(0.5-0.6)T_c$, where P_c and T_c are the critical pressure and temperature. Using the value $\sigma \sim 100$ dynes/cm^2 characteristic for liquid metals and these conditions and assuming that $P_m \sim P_c \sim 8000$ atm, as for Al and Cu, we have from Eq. (23) that the characteristic drop size equals $\sim 10^{-7}-10^{-6}$ cm, and from Eq. (26) we obtain that the characteristic expansion velocities equal hundreds of meters per second. These values agree in order of magnitude with the experimental data obtained in an electrical explosion of conductors.[7–10]

In concluding this section, we underscore that in the systems studied here thermodynamic instability occurs not as a result of the liquid reaching limiting thermodynamic states (the liquid is not heated to boiling) but rather as a result of the appearance of a mechanical instability of the system and disintegration and dispersion of the system as a result of the work performed on it by the electromagnetic field.

We shall use the van der Waals equation to investigate in greater detail the state of matter on the interphase boundary in the presence of a magnetic field.

4. INVESTIGATION OF PHASE EQUILIBRIUM IN A CURRENT-CARRYING CONDUCTOR ON THE BASIS OF THE VAN DER WAALS EQUATION

The van der Waals equation in units scaled to the critical values has the form

$$P'=\frac{8T'}{3}\frac{1}{v'-1/3}-\frac{3}{v'^2}. \tag{27}$$

Here $P'=P/P_c$, $T'=T/T_c$, $v'=v/v_c$ and P_c, T_c, and v_c are, respectively, the critical pressure, temperature, and specific volume. In what follows the primes will be dropped. Quantities referring to the center of the cylinder will be labelled with a subscript 0. The pressure P at the interphase boundary is related to the pressure P_0 at the center by $P_0 = P + P_m$. Using Eq. (27) in the last relation, after some transformations, we obtain

$$P_0 = \rho_0\rho_2(3-\rho_0-\rho_2) + \frac{P_m(3-\rho_2)\rho_0}{3(\rho_0-\rho_2)}, \tag{28}$$

$$P = \rho_0\rho_2(3-\rho_0-\rho_2) + \frac{P_m(3-\rho_0)\rho_2}{3(\rho_0-\rho_2)}, \tag{29}$$

$$T = \frac{1}{8}(\rho_0-\rho_2)(3-\rho_0)(3-\rho_2) + \frac{P_m}{3(\rho_0-\rho_2)(\rho_0+\rho_2)}, \tag{30}$$

$$T = \frac{1}{8}(\rho_1-\rho_2)(3-\rho_1)(3-\rho_2), \tag{31}$$

$$P = \rho_1\rho_2(3-\rho_1-\rho_2). \tag{32}$$

The system (28)–(32) determines the temperature and pressure at the center and at the interphase boundary in terms of the corresponding densities. The quantity P_m is a parameter. Specifically, if $P_m = 0$ and $\rho_0 = \rho_1 = \rho_2 = 1$, then the correct values of the temperature and pressure at the critical point follow from the system (28)–(32): $T = 1$ and $P_0 = P = 1$.

Another equation supplementing the system (28)–(32) will be the equality of the chemical potentials of the liquid at the center of the conductor and the gas at the interphase boundary. To obtain this equation we write the free energy in the van der Waals approximation

$$f = \frac{8T}{3}\ln\left[\frac{\varphi(T)}{v-1/3}\right] - \frac{3}{v}, \tag{33}$$

where $\varphi(T)$ is a function of temperature. Using Eqs. (27) and (30) and the thermodynamic relation $\mu = f + Pv$, the condition of phase equilibrium (21) can be put into the form

$$\frac{8T}{3}\ln\left(\frac{v_2-1}{v_0-1}\right) - (\rho_0-\rho_2)\left[6 - \frac{8T}{(3-\rho_0)(3-\rho_2)}\right] = 0. \tag{34}$$

The system of equations (28)–(32) and (34) makes it possible to calculate the temperature and pressure at the center and on the interphase boundary as well as the corresponding densities for some value of the parameter P_m.

First, we shall calculate the limiting values of the parameter P'_m which admits equilibrium. The gas phase reaches a point on the spinodal where $(\partial P/\partial v)_T = 0$. Applying this condition to Eq. (27), we obtain that the temperature and pressure on the spinodal satisfy the equations

$$P_{sp} = \rho_2^2(3-2\rho_2), \quad T_{sp} = \rho_2(3-\rho_2)^2/4. \tag{35}$$

One can see that the spinodal passes through the critical point. So, if $\rho_2 = 1$, then it follows that $P_{sp} = 1$ and $T_{sp} = 1$ from the relations (35). Equating the temperatures from Eq. (35) and (30) gives the following expression for P'_m:

$$P'_m = \frac{3(\rho_0-\rho_2)}{3-\rho_0}(\rho_0^2+\rho_0\rho_2+3\rho_2-3\rho_0-2\rho_0^2). \tag{36}$$

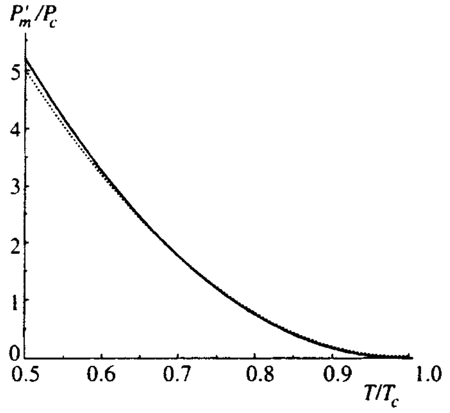


FIG. 3. Limiting magnetic pressure versus temperature.

Substituting the relation (36) into Eq. (30) and the expression so obtained into Eq. (34), we find an equation containing only the densities ρ_0 and ρ_2. Solving this equation numerically, we find from Eqs. (30) and (33) the temperature and the maximum magnetic pressure. The curve of P'_m versus temperature obtained in this manner is displayed in Fig. 3. As expected from an analysis of Figs. 1 and 2, as the temperature approaches the critical value, the values of P'_m approach zero and as the temperature decreases, they increase. The temperature dependence of the maximum values P'_m are approximated well by the simple parabola

$$P'_m \cong 80(1-T)^2. \tag{37}$$

This approximation is shown in Fig. 3 by the dotted line.

Plots of the temperature dependence of the densities and pressures of the phases for fixed values of the parameter P_m are displayed in Fig. 4. The case $P_m = 0$ (curves *1*) corresponds to no field. The curves *2–4* correspond to increasing values of P_m. Three density curves are presented for each value of P_m. The lower curve corresponds to gas, the upper curve corresponds to the liquid at the center, and the middle curve corresponds to the liquid at the interphase boundary. On the scale of the figure, it is seen only for $P_m = 0.1$. At low temperatures the curves $\rho(T)$ approach the curve *1*. As temperature increases, the curves *2–4* start to deviate from the curve *1*, but very little for the liquid phase. The largest deviations occur when the temperature at which the gas phase reaches the point on the spinodal is reached (the gas spinodal in Fig. 4b is plotted with a thin solid line). At higher temperatures, phase equilibrium is impossible. As one can see, for each value of P_m there exists a maximum temperature at which the phase curves terminate. This limiting temperature decreases with increasing P_m. In Fig. 4b, two curves correspond to each value of P_m: The lower curve refers to the interphase boundary and the upper curve refers to the center of the liquid. As the temperature decreases, the pressure on the interphase boundary approaches the value without a field. The pressure at the center becomes equal to P_m. When the gas-phase pressures reach a value corresponding to the gas spinodal of the phase, the curves terminate.

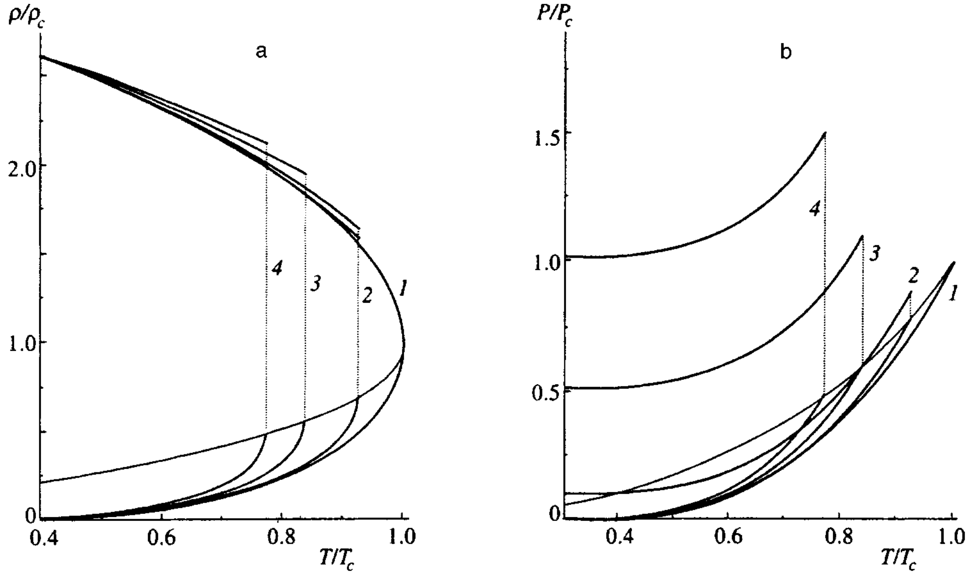


FIG. 4. Phase densities (a) and pressures (b) versus temperature for different values of the parameter P_m : *1*—$P_m=0$; *2*—0.1; *3*—0.5; *4*—1. The dotted lines connect points at the maximum temperature.

5. CONSEQUENCES FOR THE ELECTRICAL EXPLOSION OF CONDUCTORS

Some results concerning the thermodynamics of phase equilibria in liquid conductors carrying a current can be used to describe the electrical explosion of conductors. Here we shall be concerned with fast regimes, when the penetration depth of the electric field in a fused conductor is greater than the radius of the wire, which makes the skin effect unimportant. At the same time, the current density is high enough so that the characteristic heating time is much shorter than the growth time of instabilities which destroy the shape symmetry of the conductor.[7–10] Despite the long investigative history and the existence of different hypotheses, there is still no concensus about the reason why a conductor explodes. The gradual heating of a conductor by a current and the onset of surface evaporation[12] are processes which in themselves are continuous and follow one another in a definite sequence—they cannot lead to a sudden, extremely rapid, increase in the expansion rate as well as electrical resistance. The latter processes occur over a time which is much shorter than the time of the preceding processes. This is the situation observed experimentally.

The explosion of a conductor is a nonstationary process, where as the current increases, the temperature and magnetic pressure increase and the outer layers of the conductor evaporate. However, estimates show that at each moment in time the matter is in local thermodynamic equilibrium. At a definite stage of the evolution, the liquid-metal core, along which the main current generating the magnetic pressure flows, is in a thermodynamic equilibrium with the characteristic products of vaporization. This makes it possible to use the results obtained above to analyze this equilibrium. The most important result is that for each value of the temperature there exists a maximum current (magnetic pressure) for which equilibrium is still possible between the liquid and gas phases. Here we conjecture that the attainment of this value of the current is a direct cause of the sharp transition from the state of the matter where a liquid-metal shell is surrounded by vapor to a state of a rapidly-expanding finely-dispersed mixture of liquid droplets and vapor, i.e., a mechanical explosion of the conductor. Naturally, the expansion rate and the electrical resistance of the sample increase sharply at this transition. This conjecture is supported by the experimental data of Ref. 13, where the state of matter in the immediate vicinity of the point of the explosion was investigated. It was found that at the moment of the explosion the conductor separates into fragments of size $\sim 10^{-7}$ cm. The measured rates of expansion of the conductor at the moment of explosion[7–10] equal hundreds of meters per second. These figures are in complete agreement with the estimates made above.

Let us now estimate the critical current at which this can happen. The simplest equation governing the heating of a conductor by an electric current has the form

$$T=\frac{I^2\tau}{\pi^2 r_1^4 C_v \rho\sigma}, \tag{38}$$

where C_v is the heat capacity of the liquid at constant volume and τ is the characteristic heating time. Substituting the current obtained from this expression into Eq. (13) gives

$$P'_m=3\beta T/Z_c T_c, \tag{39}$$

where $\beta=2\sigma\pi/c^2\tau r_1^2$ is the skin-layer parameter and

$$Z_c=P_c m/kT_c\rho_c \tag{40}$$

is the compressibility at the critical point. For the van der Waals equation $Z_c=3/8$. Substituting the expression (39) into the approximation (37), we obtain a quadratic equation for T/T_c, whose solution has the form

$$T/T_c=1-\sqrt{\beta'}\sqrt{2+\beta'}+\beta', \tag{41}$$

where $\beta'=\beta/20$.

Using Eqs. (38) and (41), we find the critical current

$$I^2\tau=A^2\sigma\rho C_v T_c f(\beta), \tag{42}$$

where A is the transverse cross-sectional area of the conductor and $f(\beta)$ is the function on the right-hand side of Eq. (41). For an ac current Eq. (42) can be written in the form

$$\int_0^\tau I^2(t)dt=f(\beta)A^2C_v\sigma\rho T_c, \tag{43}$$

where t is the running time and τ is the time at the moment of the explosion. The integral on the left-hand side of Eq. (43) has been termed the action integral in investigations of an electrical explosion of conductors. It does not depend on the parameters of the circuit, but rather it is determined only by the properties of the material of the conductor and is proportional to the transverse cross-sectional area of the conductor. This fact was well-known empirically.[11] However, it is important that the coefficient of proportionality in Eq. (43) is $f(\beta)T_c$; this is a direct consequence of the mechanism, discussed here, of the abrupt transition of the material into a dispersed state. In this connection, we call attention to the fact that Eq. (43) provides the grounds for considering new methods for determining the critical temperature experimentally. The action integral is ordinarily easy to determine from the experimental data. The electrical conductivity, density, and ''visible'' radius of the conductor immediately preceding the explosion can also be measured.

This work is supported by the Russian Fund for Fundamental Research under Projects Nos. 96-02-16676 and 96-02-17546.

[1)] We note that in Ref. 4 a functional for the free energy is used to calculate the work required to form a nucleus of a new phase in the presence of a magnetic field. However, only the change occurring in the energy of the magnetic field in connection with the formation of a nucleus of a definite size is minimized. This approach does not give the necessary conditions for the equilibrium of a conductor in a magnetic field and leads to a number of erroneous conclusions.

[1] L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media*, Pergamon Press, N. Y. [Russian original, Nauka, Moscow, 1992].
[2] L. D. Landau and E. M. Lifshitz, *Statistical Physics*, Pergamon Press, N. Y. [Russian original, Nauka, Moscow, 1964].
[3] M. A. Leontovich, *Introduction to Thermodynamics: Statistical Physics* [in Russian], Nauka, Moscow, 1983.
[4] Yu. Dolinsky and T. Elperin, Phys. Rev. B **47**, 14778 (1993).
[5] Yu. Dolinsky and T. Elperin, Phys. Rev. B **50**, 52 (1994).
[6] Yu. Dolinsky and T. Elperin, J. Appl. Phys. **78**, 2253 (1995).
[7] N. N. Stolovich, *Electric-Explosion Energy Converters* [in Russian], Nauka i tekhnika, Minsk, 1983.
[8] V. A. Burtsev, N. V. Kalinin, and A. V. Luchinskiĭ, *Electrical Explosion of Wires and Its Applications in Electrophysical Systems* [in Russian], Énergoatomizdat, Moscow, 1990.
[9] W. G. Chace, *Conference on the Exploding Wire Phenomenon,* Plenum, New York, 1965.
[10] F. B. Bennet, in *Progress in High Temperature Physics and Chemistry,* edited by C. A. Rose, Pergamon, Oxford, 1968, Vol. 2, p. 1.
[11] I. E. Golant, A. P. Zhilinskiĭ, and I. E. Sakharov, *Fundamentals of Plasma Physics*, Wiley, New York, 1980.
[12] V. S. Vorob'ev, Zh. Tekh. Fiz. **66**, 35 (1996) [Tech. Phys. **41**, 17 (1996)].
[13] T. Sahashi, Y. Hioki, and J. Yamada, Jpn. J. Appl. Phys. **30**, 780 (1991).

Translated by M. E. Alferieff

Contribution to the statistical theory of wave localization in a two-layered medium

N. V. Gryanik

Institute of Atmospheric Physics, Russian Academy of Sciences, 109017 Moscow, Russia

V. I. Klyatskin

Institute of Atmospheric Physics, Russian Academy of Sciences, 109017 Moscow, Russia;
Pacific Ocean Oceanological Institute, Far-East Division of the Russian Academy of Science, 690041 Vladivostok, Russia

(Submitted 10 September 1996)
Zh. Éksp. Teor. Fiz. **111**, 2030–2043 (June 1997)

A very simple system of stochastic boundary-value wave equations that describes the interaction of two types of waves in a randomly inhomogeneous medium is studied. The statistics of the reflection and transmission coefficients for the incident and excited waves are discussed. It is shown that the excitation of waves is statistically equivalent to switching on damping for the initial incident waves which are localized in separate specific realizations. The parameters of the length of such localization are estimated in terms of the spectral density of the variations of the medium. It is also shown that for excited waves there is no dynamical localization, and the transmission coefficients for them are estimated.
[S1063-7761(97)00806-8]

1. INTRODUCTION

The localization of plane waves in layered media described on the basis of a one-dimensional Helmholtz equation for a random index of refraction has now been studied in detail (see, for example, Refs. 1–4). There are several approaches for analyzing such problems. The traditional approach is based on the analogy between the Helmholtz equation and the time-independent Schrödinger equation with a random potential, to analyze which the spectral properties of this equation are studied.[1] Another approach, which is also often used, is based on analytical and numerical methods of analysis of the Lyapunov exponents, which likewise characterize the spatial localization of the eigenfunctions of a homogeneous boundary-value wave problem.[1,5] However, an entirely different approach was used in Refs. 2–4. In this approach specific physical boundary-value problems of wave generation in random media and the spatial structure of the wave field in such a medium are analyzed.

However, investigators often encounter a multidimensional situation where waves of one type can create waves of a different type on account of the dependence of the parameters of the problem on the spatial coordinates. In many cases such a problem can be parameterized by dividing the medium in some direction into layers which are characterized by a discrete set of some parameters while other parameters in these layers vary in a continuous manner. An example is the equation of large-scale and low-frequency motions in the earth's atmosphere and oceans (for example, Rossby waves). These motions can be described on the basis of a quasigeostrophic model, in which the atmosphere and ocean are regarded as thin, multilayered films characterized in the vertical direction by the thicknesses of the layers and the values of the density in them.[6] The parameters of the medium within the layers vary in a continuous manner. A possible source of localization of Rossby waves is spatial variations of the bottom topography in a horizontal plane. The simplest one-layer model, equivalent to the one-dimensional Helmholtz equation, describes the barotropic motions of the medium,[5,7] and a two-layer model takes account of the baroclinic effects.[8,9] The possibility of wave localization has virtually not been studied in a two-layer model of a medium described by a system of Helmholtz equations. In the present paper this problem is studied for the example of the simplest system of wave equations, and it is shown that complete (for all types of waves) localization does not occur.

2. FORMULATION OF THE PROBLEM

A very simple model of the propagation of interacting waves in a two-layer medium is described by the system of equations[9]

$$\frac{d^2}{dx^2}\,\psi_1+k^2\psi_1-\alpha_1F(\psi_1-\psi_2)=0,$$

$$\frac{d^2}{dx^2}\,\psi_2+k^2[1+\varepsilon(x)]\psi_2+\alpha_2F(\psi_1-\psi_2)=0, \tag{1}$$

where $\alpha_1=1/H_1$ and $\alpha_2=1/H_2$ (H_1 and H_2 are the thicknesses of the top and bottom layers), F characterizes the interaction of the waves, and the function $\varepsilon(x)$ describes the inhomogeneities of the medium in the bottom layer. We assume that the function $\varepsilon(x)$ is a random function and is different from zero only in the region (L_0,L). The geometry of the problem is presented in Fig. 1. The boundary conditions for Eq. (1) are radiation conditions at infinity and continuity of the wave fields and their derivatives at the boundaries L and L_0 of the region.

We note that the numerical quantity F, characterizing the parameterization of the medium in the vertical direction, appears in the system of equations (1) as a horizontal scale responsible for the generation of an additional wave. The character of the wave interaction (and, specifically, the de-

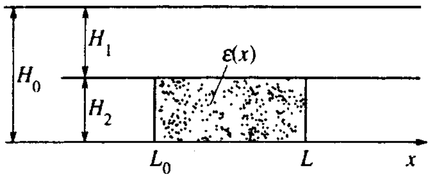


FIG. 1. Geometry of the stochastic problem (1).

pendence of the parameters α_i on the layer thicknesses) in the system (1) corresponds to the problems of geophysical hydrodynamics.[6] For problems of different types these relations can change, which is not important for the further analysis (see Conclusions). Only the linearity of the interaction between the waves is of fundamental importance.

The transition to a one-layer model is made by setting $F=0$ and $\Psi_1=0$. The corresponding wave equation assumes the form of the Helmholtz equation

$$\frac{d^2}{dx^2}\,\psi+k^2[1+\varepsilon(x)]\psi=0. \tag{2}$$

For definiteness, we shall call a wave of this type a k wave. The transition to the one-dimensional problem can also be made by passing to the limit $H_1\to 0$; then $\psi_1=\psi_2$. It should be kept in mind, however, that the limits $L_0\to-\infty$ (transition to a half-space) and $H_i\to 0$ are not interchangeable in the statistical problem.[9] In what follows, we shall be interested in the case of a half-space and so the layer thicknesses H_i, though small, can be finite.

2.1. Point sources of wave generation inside the layers

Let us consider the system of equations for the Green's functions:

$$\frac{d^2}{dx^2}\,\psi_1+k^2\psi_1-\alpha_1F(\psi_1-\psi_2)=-v_1\delta(x-x_0),$$

$$\frac{d^2}{dx^2}\,\psi_2+k^2[1+\varepsilon(x)]\psi_2+\alpha_2F(\psi_1-\psi_2)=-v_2\delta(x-x_0), \tag{3}$$

for the corresponding problem of wave excitation in the top and bottom layers. With the vector notation $\psi(x;x_0)=\{\psi_1(x;x_0),\psi_2(x;x_0)\}$ and $\mathbf{v}=\{v_1,v_2\}$, the system (3) can be rewritten in the vector form

$$\left[\frac{d^2}{dx^2}+A^2+k^2\varepsilon(x)\Gamma\right]\psi(x;x_0)=-\mathbf{v}\delta(x-x_0),$$

$$A^2=\begin{bmatrix} k^2-\alpha_1F & \alpha_1F \\ \alpha_2F & k^2-\alpha_2F \end{bmatrix},\quad \Gamma=\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix},$$

$$A=k\begin{bmatrix} \widetilde{\alpha}_2+\lambda\widetilde{\alpha}_1 & (1-\lambda)\widetilde{\alpha}_1 \\ (1-\lambda)\widetilde{\alpha}_2 & \widetilde{\alpha}_1+\lambda\widetilde{\alpha}_2 \end{bmatrix},$$

$$A^{-1}=\frac{1}{k\lambda}\begin{bmatrix} \widetilde{\alpha}_1+\lambda\widetilde{\alpha}_2 & -(1-\lambda)\widetilde{\alpha}_1 \\ -(1-\lambda)\widetilde{\alpha}_2 & \widetilde{\alpha}_2+\lambda\widetilde{\alpha}_1 \end{bmatrix},$$

where we have introduced the parameter

$$\lambda^2=[1-(\alpha_1+\alpha_2)F/k^2], \tag{5}$$

describing a mode which we shall call a λ wave (we assume $\lambda^2>0$) and the relative thicknesses of the layers are

$$\widetilde{\alpha}_1=\frac{\alpha_1}{\alpha_1+\alpha_2}=\frac{H_2}{H_0},\quad \widetilde{\alpha}_2=\frac{\alpha_2}{\alpha_1+\alpha_2}=\frac{H_1}{H_0},$$

$$\widetilde{\alpha}_1+\widetilde{\alpha}_2=1. \tag{5$'$}$$

Let us consider the matrix Ψ described by the equation

$$\left[\frac{d^2}{dx^2}+A^2+k^2\varepsilon(x)\Gamma\right]\Psi(x;x_0)=-E\,\delta(x-x_0), \tag{6}$$

in terms of which the vector function $\psi(x;x_0)$ is defined according to the relation

$$\psi(x;x_0)=\Psi(x;x_0)\mathbf{v}=\begin{bmatrix} \psi_{11} & \psi_{12} \\ \psi_{21} & \psi_{22} \end{bmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$$

$$=\begin{pmatrix} v_1\psi_{11}+v_2\psi_{12} \\ v_1\psi_{21}+v_2\psi_{22} \end{pmatrix}. \tag{7}$$

In consequence, the components of the matrices $\{\psi_{11},\psi_{21}\}$ and $\{\psi_{12},\psi_{22}\}$ describe waves generated by the sources $\{v_1,0\}$ and $\{0,\,v_2\}$ in the top and bottom layers, respectively. The boundary conditions for Eq. (6) are

$$\left(\frac{d}{dx}-iA\right)\Psi(x;x_0)|_{x=L}=0,$$

$$\left(\frac{d}{dx}+iA\right)\Psi(x;x_0)|_{x=L_0}=0. \tag{6$'$}$$

Following Ref. 9, we shall simplify the problem even further. Specifically, we shall assume that the source of plane waves lies at the boundary $x_0=L$ of the irregular region. In this case, using the condition for a jump in the wave field at the source point x_0,

$$\frac{d}{dx}\,\Psi(x;x_0)|_{x=x_0+0}-\frac{d}{dx}\,\Psi(x;x_0)|_{x=x_0-0}=-E,$$

we obtain the boundary-value problem

$$\left[\frac{d^2}{dx^2}+A^2+k^2\varepsilon(x)\Gamma\right]\Psi(x;L)=0,$$

$$\left(\frac{d}{dx}-iA\right)\Psi(x;L)|_{x=L}=E,$$

$$\left(\frac{d}{dx}+iA\right)\Psi(x;L)|_{x=L_0}=0. \tag{8}$$

2.2. Generation of *k* and λ waves

To put the problem into a physically more illuminating form, we shall study, instead of the matrix $\Psi(x;L)$, the matrix function

$$U(x;L)=-2iK\Psi(x;L)K^{-1}B, \tag{9}$$

where

$$B=k\begin{bmatrix}\lambda & 0\\ 0 & 1\end{bmatrix},\quad K=\begin{bmatrix}1 & -1\\ \widetilde{\alpha}_2 & \widetilde{\alpha}_1\end{bmatrix},\quad K^{-1}=\begin{bmatrix}\widetilde{\alpha}_1 & 1\\ -\widetilde{\alpha}_2 & 1\end{bmatrix},$$

which makes it possible to switch to the boundary-value problem for $U(x;L)$ with a diagonal matrix B,

$$\left[\frac{d^2}{dx^2}+B^2+k^2\varepsilon(x)\widetilde{\Gamma}\right]U(x;L)=0,$$

$$\left(\frac{d}{dx}-iB\right)U(x;L)|_{x=L}=-2iB,$$

$$\left(\frac{d}{dx}+iB\right)U(x;L)|_{x=L_0}=0, \tag{10}$$

where

$$\widetilde{\Gamma}=K\Gamma K^{-1}=\begin{bmatrix}\widetilde{\alpha}_2 & -1\\ -\widetilde{\alpha}_1\widetilde{\alpha}_2 & \widetilde{\alpha}_1\end{bmatrix}.$$

The boundary-value problem (10) describes the incidence of k and λ waves with unit amplitudes on the medium, where an incident λ wave U_{11} generates a k wave U_{21} and an incident k wave U_{22} generates a λ wave U_{12}.

It follows from the system (10) that the amplitude of the generated k wave U_{21} is proportional to the parameter

$$\delta=\lambda\widetilde{\alpha}_1\widetilde{\alpha}_2=\lambda\frac{H_1H_2}{H_0^2}, \tag{11}$$

i.e.,

$$U_{21}=\delta\widetilde{U}_{21}. \tag{11$'$}$$

In the general case the parameter $\delta<\lambda/4$. However, since in models describing real media we ordinarily have $\widetilde{\alpha}_1\widetilde{\alpha}_2\ll 1$ (for example, for the atmosphere we assume $H_2\ll H_1$ or $\widetilde{\alpha}_1\ll 1$, $\widetilde{\alpha}_2\cong 1$ and for the ocean we assume $H_1\ll H_2$ or $\widetilde{\alpha}_2\ll 1$, $\widetilde{\alpha}_1\cong 1$), a small parameter δ appears in the problem under study. For models of a medium with $H_2/H_1\simeq 1$ we have $\delta\ll 1$ for $\lambda\ll 1$.

We now introduce the matrices $R(L)=U(L;L)-E$ and $T(L)=U(L_0;L)$. Then R_{ij} and T_{ij} will be complex reflection and transmission coefficients for the incident ($i=j$) and excited ($i\neq j$) λ and k waves, respectively.

The system (10) implies the existence of two integrals corresponding to conservation of the energy flux density of the λ and k waves:

$$\widetilde{\alpha}_1\widetilde{\alpha}_2\left[U_{11}^*(x)\frac{d}{dy}U_{11}(x)-U_{11}(x)\frac{d}{dy}U_{11}^*(x)\right]$$
$$+U_{21}^*(x)\frac{d}{dy}U_{21}(x)-U_{21}(x)\frac{d}{dy}U_{21}^*(x)=\text{const},$$

$$\widetilde{\alpha}_1\widetilde{\alpha}_2\left[U_{12}^*(x)\frac{d}{dy}U_{12}(x)-U_{12}(x)\frac{d}{dy}U_{12}^*(x)\right]$$
$$+U_{22}^*(x)\frac{d}{dy}U_{22}(x)-U_{22}(x)\frac{d}{dy}U_{22}^*(x)=\text{const}.$$

They are described in terms of the reflection and transmission coefficients in the form of the inequalities

$$\delta[1-|R_{11}|^2-|T_{11}|^2]=|R_{21}|^2+|T_{21}|^2,$$

$$1-|R_{22}|^2-|T_{22}|^2=\delta[|R_{12}|^2+|T_{12}|^2], \tag{12}$$

which, on the strength of Eq. (11$'$), can be rewritten in the form

$$1-|R_{11}|^2-|T_{11}|^2=\delta[|\widetilde{R}_{21}|^2+|\widetilde{T}_{21}|^2],$$

$$1-|R_{22}|^2-|T_{22}|^2=\delta[|R_{12}|^2+|T_{12}|^2]. \tag{12$'$}$$

If complete localization of the waves occurs in the region (L_0,L) of inhomogeneities of the medium, then all transmission coefficients T_{ij} must approach zero as the size $L-L_0$ of the region increases.

The equalities (12) and (12$'$) determine the relation between the transmission and reflection coefficients, which, by means of the method of embedding, can be described by a closed system of equations. The method of embedding makes it possible to switch from a problem for the matrix function $U(x)$ with boundary conditions to a system of equations for the matrix functions $U(x;L)$ and $U(L;L)$ with initial conditions with respect to the parameter L (in this case the variable x is treated as a parameter):^{9}

$$\frac{\partial}{\partial L}U(x;L)=iU(x;L)B$$
$$+\frac{i}{2}k^2\varepsilon(L)U(x;L)B^{-1}\widetilde{\Gamma}U(L;L),$$

$$U(x;L)|_{L=x}=U(x;x),$$

$$\frac{d}{dL}U(L;L)=-2iB+i[U(L;L)B+BU(L;L)]$$
$$+\frac{i}{2}k^2\varepsilon(L)U(L;L)B^{-1}\widetilde{\Gamma}U(L;L),$$

$$U(L;L)|_{L=L_0}=E. \tag{13}$$

The last equation for the matrix $R(L)=U(L;L)-E$ can be rewritten in the form of the matrix Riccati equation

$$\frac{d}{dL}R(L)=i[R(L)B+BR(L)]$$
$$+\frac{i}{2}k^2\varepsilon(L)[E+R(L)]B^{-1}\widetilde{\Gamma}[E+R(L)],$$

$$R(L)|_{L=L_0}=0. \tag{13$'$}$$

Writing Eq. (13$'$) R_{ij}, it is easy to see that there exists an additional integral $R_{21}=\delta R_{12}$, i.e., $\widetilde{R}_{21}=R_{12}$. This makes it possible to study a system of three equations for R_{11}, R_{12}, and R_{22}, which we represent in the form

$$\frac{d}{dL}R_{11}=2i\lambda kR_{11}+i\widetilde{\varepsilon}(L)[\widetilde{\alpha}_2(1+R_{11})^2$$
$$-2\delta(1+R_{11})R_{12}+\delta\lambda\widetilde{\alpha}_1R_{12}^2],$$

$$\frac{d}{dL}R_{22}=2ikR_{22}+i\widetilde{\varepsilon}(L)[\lambda\widetilde{\alpha}_1(1+R_{22})^2$$
$$-2\delta(1+R_{22})R_{12}+\delta\widetilde{\alpha}_2R_{12}^2],$$

$$\frac{d}{dL}R_{12}=ik(1+\lambda)R_{12}+i\tilde{\varepsilon}(L)$$

$$\times\{[\tilde{\alpha}_2(1+R_{11})+\lambda\tilde{\alpha}_1(1+R_{22})-\delta R_{12}]R_{12}$$

$$-(1+R_{11})(1+R_{22})\},\qquad(14)$$

where

$$\tilde{\varepsilon}(L)=\frac{k}{2\lambda}\,\varepsilon(L).$$

3. STATISTICAL ANALYSIS

Thus far no assumptions have been made about the character of the nonuniformities of the random function $\varepsilon(x)$. We shall assume below that $\varepsilon(x)$ is a homogeneous Gaussian random process with zero mean ($\langle\varepsilon(x)\rangle=0$) and correlation and spectral functions

$$B_\varepsilon(\xi)=\langle\varepsilon(x)\varepsilon(x')\rangle,\quad \Phi_\varepsilon(q)=\int_{-\infty}^{\infty}d\xi B_\varepsilon(\xi)\exp(iq\xi),$$

$$\xi=x-x',\qquad(15)$$

characterized by the correlation length l_0. Here and below $\langle\ldots\rangle$ denotes averaging with respect to an ensemble of realizations of the random function $\varepsilon(x)$.

3.1. Some results of a statistical analysis of the one-layer model

We note that for a one-layer model of the medium the equation for the reflection coefficient $R=\psi-1$ follows from Eq. (2) and the corresponding boundary conditions for it

$$\frac{d}{dL}R_L=2ikR_L+i\,\frac{k}{2}\,\varepsilon(L)(1+R_L)^2,\quad R_{L_0}=0.\qquad(16)$$

For this model the conservation of the energy flux density gives

$$1-|R_L|^2=|T_L|^2,\qquad(16')$$

and in the diffusion approximation with additional averaging over fast oscillations the Fokker–Planck equation for the probability density $P_L(W)=\langle\delta(W(L)-W)\rangle$ for the random variable $W(L)=|R_L|^2$ has the form[2–4]

$$\frac{\partial}{\partial L}P_L(W)=D\,\frac{\partial}{\partial W}\left[-(1-W)^2+\frac{\partial}{\partial W}W(1-W)^2\right]P_L(W)\qquad(17)$$

with a diffusion coefficient with the dimension of inverse length:

$$D=\frac{k^2}{4}\,\Phi_\varepsilon(2k).\qquad(17')$$

The diffusion approximation is valid with certain restrictions. Specifically, the random fluctuations of the function $\varepsilon(x)$ should not affect the propagation of a wave on scales of the order of the correlation length l_0, i.e., the condition $Dl_0\ll 1$ must be satisfied. In other words, a wave on such scales does not ‘‘feel’’ the inhomogeneities of the medium and propagates as it would in free space. The condition of applicability of averaging over fast oscillations is $kD\gg 1$.

For small-scale inhomogeneities, for which $kl_0\ll 1$, of a medium we have

$$D=\frac{k^2}{4}\,\Phi_\varepsilon(0),\qquad(17'')$$

and the solution of the statistical problem does not depend on the form of the spectral function.

It follows from Eq. (17) that if the region of inhomogeneities of a medium is sufficiently large, specifically, $D(L-L_0)\gg 1$, the probability density is $P_L(W)=\delta(W-1)$ and so we have $|T|^2=0$ with probability 1, which corresponds to dynamical wave localization. In this case the intensity of the wave field inside the medium in specific realizations has the structure of an exponentially decaying curve:[3,4]

$$I(x;L)=|U(x;L)|^2\simeq 2\exp[-(L-x)/l_{\mathrm{loc}}],\qquad(18)$$

where l_{loc} is determined by the diffusion coefficient

$$l_{\mathrm{loc}}=1/D,\qquad(18')$$

and is called the localization length. The statistical moments of the intensity of the wave then grow exponentially as the wave propagates into the layer:

$$\langle I^n(x;L)\rangle\sim\exp[n(n-1)D(L-x)],\quad n=1,2,...,\qquad(18'')$$

which is due to the large random variations of the intensity relative to the curve (18) in separate realizations.[3,4]

3.2. Statistical analysis of a two-layer model

We now return to a two-layer model of the medium. In the Appendix, the Fokker–Planck equation (A5) is derived in the diffusion approximation. This equation describes the probability density of the squared absolute values of the reflection coefficients $W_{ij}=|R_{ij}|^2$ and contains, in contrast to the one-layer medium, four diffusion coefficients which are expressed in terms of the spectral function of the random process $\varepsilon(x)$ as follows:

$$D_1=\left(\frac{k}{2\lambda}\frac{H_1}{H_0}\right)^2\Phi_\varepsilon(2\lambda k),\quad D_2=\left(\frac{k}{2}\frac{H_2}{H_0}\right)^2\Phi_\varepsilon(2k),$$

$$D_3=\left(\frac{k}{2\lambda}\right)^2\Phi_\varepsilon(k(1+\lambda)),\quad D_4=\left(\frac{k}{2\lambda}\right)^2\Phi_\varepsilon(k(1-\lambda)).\qquad(19)$$

We note that for small-scale inhomogeneities of the medium ($kl_0\ll 1$) the diffusion coefficients are all expressed in terms of one coefficient, D, described by the formula (17′)

$$D_1=\left(\frac{1}{\lambda}\frac{H_1}{H_0}\right)^2D,\quad D_2=\left(\frac{H_2}{H_0}\right)^2D,\quad D_3=D_4=\frac{1}{\lambda^2}\,D.\qquad(19')$$

The diffusion approximation is applicable when

$$D_il_0\ll 1.\qquad(20)$$

The derivation of Eq. (A5) also employed additional averaging over fast oscillations, which is valid when

$$k\lambda \gg D_i . \tag{21}$$

As mentioned above, the present problem contains a parameter δ whose smallness can be used to simplify the analysis of the problem. In Eq. (A5) we neglect terms which are of second order in δ, i.e., secondary reemission of waves. In this approximation the quantities W_{11} and W_{22} are statistically independent and their probability densities $P_L(W_{11})$ and $P_L(W_{22})$ are described by the equations

$$\frac{\partial}{\partial L} P_L(W_{11}) = \left\{ \frac{\partial}{\partial W_{11}} [-D_1(1-W_{11})^2 + 2\delta(D_3+D_4)W_{11}] + D_1 \frac{\partial^2}{\partial W_{11}^2} W_{11}(1-W_{11})^2 \right\} P_L(W_{11}), \tag{22}$$

$$\frac{\partial}{\partial L} P_L(W_{22}) = \left\{ \frac{\partial}{\partial W_{22}} [-D_2(1-W_{22})^2 + 2\delta(D_3+D_4)W_{22}] + D_2 \frac{\partial^2}{\partial W_{22}^2} W_{22}(1-W_{22})^2 \right\} P_L(W_{22}),$$

which differ from Eq. (17) for the one-layer model by the presence of the term

$$2\delta(D_3+D_4) \frac{\partial}{\partial W} W P_L(W).$$

This means that the process of generation of a λ (or k) wave by an incident k (or λ) wave is statistically equivalent to switching on damping in the initial statistical problem for the incident waves U_{11} and U_{22} (i.e., a substitution $\tilde{\varepsilon}(x) \to \tilde{\varepsilon}(x) + i\delta(D_3+D_4)$ in the equations for these waves). Then, for the half-space ($L_0 \to -\infty$) there exist ''stationary'' (independent of L) solutions of Eqs. (22)[2–4]

$$P_\infty(W_{11}) = \frac{2\gamma_1}{(1-W_{11})^2} \exp\left(-\frac{2\gamma_1 W_{11}}{1-W_{11}} \right),$$

$$P_\infty(W_{22}) = \frac{2\gamma_2}{(1-W_{22})^2} \exp\left(-\frac{2\gamma_2 W_{22}}{1-W_{22}} \right). \tag{23}$$

where the parameters

$$\gamma_1 = \delta \frac{D_3+D_4}{D_1}, \quad \gamma_2 = \delta \frac{D_3+D_4}{D_2} \tag{24}$$

determine the relative role of such damping (i.e., generation of secondary waves) compared with the direct diffusion of these waves (i.e., repeated rereflection of these waves by irregularities of the medium). For small-scale irregularities of the medium the damping parameters

$$\gamma_1 = 2\lambda \frac{H_2}{H_1}, \quad \gamma_2 = \frac{2}{\lambda} \frac{H_1}{H_2} \tag{24'}$$

are determined only by the relative thicknesses of the layers (for a fixed wavelength of the λ wave) and do not depend on the statistical characteristics of the irregularities. Then $\gamma_1\gamma_2 = 4$, i.e., if one of the parameters γ is compared to 1, then the other parameter is large.

The probability distributions (23) make it possible to calculate the statistical characteristics of the reflection coefficients for the incident waves. Specifically, for $\gamma_i \ll 1$ we have

$$\langle W_{11} \rangle \approx 1 - 2\gamma_1 \ln(1/\gamma_1), \quad \langle W_{22} \rangle \approx 1 - 2\gamma_2 \ln(1/\gamma_2). \tag{25}$$

In the opposite limiting cases $\gamma_i \gg 1$ we obtain, correspondingly,

$$\langle W_{11} \rangle \approx 1/2\gamma_1, \quad \langle W_{22} \rangle \approx 1/2\gamma_2 . \tag{25'}$$

It is obvious from what has been said above that for a sufficiently large region (L_0, L) (or in the limiting case of a half-space $L_0 \to -\infty$) the quantities $|T_{11}|^2$ and $|T_{22}|^2$ equal zero with probability 1, i.e., the incident λ and k waves are localized, and their localization lengths are determined either by the diffusion coefficients, if diffusion is much stronger than the damping, or by the damping otherwise. For example, for $\gamma_1 \ll 1$ ($\gamma_2 \gg 1$),

$$l_{\text{loc}}^{(1)} = \frac{1}{D_1} = \left(\frac{\lambda H_0}{H_1} \right)^2 l_{\text{loc}},$$

$$l_{\text{loc}}^{(2)} = \frac{1}{2\delta(D_3+D_4)} = \frac{\lambda H_0}{4H_1H_2} l_{\text{loc}}, \tag{26}$$

where l_{loc} is the localization length in the one-layer problem (18′). In the case when the opposite inequality holds $\gamma_1 \gg 1$ ($\gamma_2 \ll 1$) we have

$$l_{\text{loc}}^{(1)} = \frac{1}{2\delta(D_3+D_4)} = \frac{\lambda H_0}{4H_1H_2} l_{\text{loc}},$$

$$l_{\text{loc}}^{(2)} = \frac{1}{D_2} = \left(\frac{H_0}{H_2} \right)^2 l_{\text{loc}}. \tag{26'}$$

Finding the statistics for W_{12} is a much more complicated problem, since the statistics in this case are due to the correlation of W_{12} with W_{11} and W_{22}.

To estimate the average values of the transmission coefficients for the excited waves, we shall employ the relations (12′), which we write in the form

$$1 - \langle W_{11} \rangle - \delta \langle W_{12} \rangle = \delta \langle |\tilde{T}_{21}|^2 \rangle,$$

$$1 - \langle W_{22} \rangle - \delta \langle W_{12} \rangle = \delta \langle |T_{12}|^2 \rangle. \tag{27}$$

It is evident from the Fokker–Planck equation (A5) that for the combinations $T_1 = 1 - W_{11} - \delta W_{12}$ and $T_2 = 1 - W_{22} - \delta W_{12}$, which determine the transmission coefficients for the excited waves, there are no stationary solutions of the form $P(T_j) = \delta(T_j)$ for a half-space. This means that there is no localization for the waves generated. Since Eq. (A5) is symmetric with respect to the indices 1 and 2, the average value $\langle W_{12} \rangle$ should also be symmetric with respect to these indices and therefore, to within symmetric parts, at least the order of the quantities $\langle |T_{ij}|^2 \rangle$ should be determined by the order of the asymmetric parts of the conditions (27). For example, by virtue of Eqs. (25) and (25′), for the asymptotic case $\gamma_1 \ll 1$ ($\gamma_2 \gg 1$) Eqs. (27) assume the form

$$2\gamma_1 \ln(1/\gamma_1) = \delta \langle W_{12} \rangle + \delta \langle |\tilde{T}_{21}|^2 \rangle,$$

$$1 - 1/2\gamma_2 = \delta \langle W_{12} \rangle + \delta \langle |T_{12}|^2 \rangle, \tag{28}$$

i.e.,

$$\langle|\widetilde{T}_{21}|^2\rangle \sim \frac{2}{\delta}\,\gamma_1 \ln\frac{1}{\gamma_1}, \quad \langle|T_{12}|^2\rangle \sim \frac{1}{\delta}. \tag{29}$$

Similarly, for the opposite asymptotic case $\gamma_2 \ll 1$ ($\gamma_1 \gg 1$) we obtain the estimate

$$\langle|\widetilde{T}_{21}|^2\rangle \sim \frac{1}{\delta}, \quad \langle|T_{12}|^2\rangle \sim \frac{2}{\delta}\,\gamma_2 \ln\frac{1}{\gamma_2}. \tag{29$'$}$$

Returning to the initial formulation of the problem of sources located in the top and bottom layers of the medium at the boundary $x_0 = L$ of the irregular region, let us now find the intensities of the waves at the boundary L_0 for $L_0 \to -\infty$, i.e., the transmission coefficients for this problem.

We consider two cases:

1. The source is located in the bottom layer ($v_1 = 0$, $v_2 = 1$). In this case, we obtain on the basis of the equality (7) the expressions

$$\langle|\psi_1(x)|^2\rangle|_{x=L_0} = \langle|\psi_{12}(x)|^2\rangle|_{x=L_0}$$
$$= \frac{1}{4k^2}\frac{H_2^2}{H_0^2}\left(\frac{H_1^2}{H_0^2}\langle|\widetilde{T}_{21}|^2\rangle + \frac{H_2^2}{H_0^2}\langle|T_{12}|^2\rangle\right),$$
$$\langle|\psi_2(x)|^2\rangle_{x=L_0} = \langle|\psi_{22}(x)|^2\rangle|_{x=L_0}$$
$$= \frac{1}{4k^2}\frac{H_1^2 H_2^2}{H_0^4}\left(\langle|\widetilde{T}_{21}|^2\rangle + \langle|T_{12}|^2\rangle\right). \tag{30}$$

2. The source is located in the top layer ($v_1 = 1$, $v_2 = 0$). Then we obtain similarly to Eqs. (30)

$$\langle|\psi_1(x)|^2\rangle|_{x=L_0} = \langle|\psi_{11}(x)|^2\rangle|_{x=L_0}$$
$$= \frac{1}{4k^2}\frac{H_1^2 H_2^2}{H_0^4}\left(\langle|\widetilde{T}_{21}|^2\rangle + \langle|T_{12}|^2\rangle\right),$$
$$\langle|\psi_2(x)|^2\rangle_{x=L_0} = \langle|\psi_{21}(x)|^2\rangle|_{x=L_0}$$
$$= \frac{1}{4k^2}\frac{H_1^2}{H_0^2}\left(\frac{H_2^2}{H_0^2}\langle|\widetilde{T}_{21}|^2\rangle + \frac{H_1^2}{H_0^2}\langle|T_{12}|^2\rangle\right). \tag{30$'$}$$

Therefore the transmission coefficients for waves generated by sources in both the top and bottom layers of the medium are different from zero in the entire medium, i.e., wave localization does not occur. Their specific values, however, are determined by both the ratio of the layer thicknesses and the parameter λ (see, for example, the asymptotic formulas (29) and (29′) as well as (24′)).

4. CONCLUSIONS

Thus, we have shown that on the basis of a very simple two-layer model of a medium described by the system of stochastic equations (1) with random inhomogeneities in the bottom layer, k or λ waves incident on the layer are localized, just as in the one-layer model of the medium. Incident k or λ waves generate λ or k waves which are no longer localized, i.e., their transmission coefficient does not approach zero as the size of the irregular region of the medium increases. The values of these transmission coefficients were estimated for the asymptotic case of a half-space. The lack of localization in this limiting case means that there is no localization for a finite region of the inhomogeneous medium as well.

We note that making the problem more complicated (by including inhomogeneities in the top layer as well, changing the character of the interaction of the waves, changing the model of fluctuating parameters, for example, replacing $\varepsilon(x)$ by $d\varepsilon(x)/dx$, and so on) complicates the corresponding Fokker–Planck equation and changes the dependence of the statistical characteristics on the geometric parameters, but it does not change the main result of this work—the absence of dynamic localization for waves as a whole.

This work was supported in part by the Russian Fund for Fundamental Research under projects Nos. 95-05-14247 and 96-05-65347.

APPENDIX

Diffusion approximation and the Fokker–Planck equation for the probability density of reflected-wave intensities

The equations for the intensities of the reflected waves can be obtained from the system of equations (14) ($W_{ij} = |R_{ij}|^2$):

$$\frac{d}{dL}W_{11}(L) = i\widetilde{\varepsilon}(L)[\widetilde{\alpha}_2 R_{11}^*(1 - W_{11}) - 2\,\delta R_{11}^* R_{12}$$
$$- 2\,\delta W_{11} R_{12} + \delta\lambda\,\widetilde{\alpha}_1 R_{11}^* R_{12}^2 - \text{c.c.}],$$
$$\frac{d}{dL}W_{22}(L) = i\widetilde{\varepsilon}(L)[\lambda\,\widetilde{\alpha}_1 R_{22}^*(1 - W_{22}) - 2\,\delta R_{22}^* R_{12}$$
$$- 2\,\delta W_{22} R_{12} + \delta\widetilde{\alpha}_2 R_{22}^* R_{12}^2 - \text{c.c.}],$$
$$\frac{d}{dL}W_{12}(L) = i\widetilde{\varepsilon}(L)[(\widetilde{\alpha}_2 R_{11} + \lambda\,\widetilde{\alpha}_1 R_{22} - \delta R_{12})W_{12}$$
$$- (1 + R_{11})(1 + R_{22})R_{12}^* - \text{c.c.}]. \tag{A1}$$

Their probability density is described by the formula

$$P_L(W_{11}, W_{22}, W_{12}) = \langle\Phi_L(W_{11}, W_{22}, W_{12})\rangle,$$

where the function Φ is determined by the equality

$$\Phi_L(W_{11}, W_{22}, W_{12})$$
$$= \delta(W_{11}(L) - W_{11})\,\delta(W_{22}(L) - W_{22})\,\delta(W_{12}(L) - W_{12})$$

and is described by the Liouville equation

$$\frac{\partial}{\partial L}\Phi_L(W_{11}, W_{22}, W_{12})$$
$$= -i\widetilde{\varepsilon}(L)\left\{\frac{\partial}{\partial W_{11}}[\widetilde{\alpha}_2 R_{11}^*(1 - W_{11}) - 2\,\delta R_{11}^* R_{12}\right.$$
$$- 2\,\delta W_{11} R_{12} + \delta\lambda\,\widetilde{\alpha}_1 R_{11}^* R_{12}^2 - \text{c.c.}]$$
$$+ \frac{\partial}{\partial W_{22}}\left[\lambda\,\widetilde{\alpha}_1 R_{22}^*(1 - W_{22}) - 2\,\delta R_{22}^* R_{12}\right.$$
$$- 2\,\delta W_{22} R_{12} + \delta\widetilde{\alpha}_2 R_{22}^* R_{12}^2 - \text{c.c.}$$

$$
\begin{aligned}
&+\frac{\partial}{\partial W_{12}}[(\widetilde{\alpha}_2 R_{11}+\lambda\widetilde{\alpha}_1 R_{22}-\delta R_{12})W_{12}\\
&\left.-(1+R_{11})(1+R_{22})R_{12}^{*}-\text{c.c.}]\right\}\\
&\times\Phi_L(W_{11},W_{22},W_{12}). \qquad \text{(A2)}
\end{aligned}
$$

Let us average Eq. (A2) over an ensemble of realizations of the random process $\widetilde{\varepsilon}(L)$. To decouple the correlations of $\widetilde{\varepsilon}(x)$ from the functions $W_{ij}(L)$, which are functionals of the process $\widetilde{\varepsilon}(L)$, we employ the Furutsu–Novikov formula (see, for example, Ref. 10)

$$
\langle\widetilde{\varepsilon}(L)R[\widetilde{\varepsilon}(x)]\rangle=\int_{L_0}^{L}d\xi\, B_{\widetilde{\varepsilon}}(L-\xi)\left\langle\frac{\delta}{\delta\widetilde{\varepsilon}(\xi)}R[\widetilde{\varepsilon}(x)]\right\rangle, \quad \text{(A3)}
$$

which holds for a Gaussian random process with zero mean and an arbitrary functional of it.

To calculate the variational derivatives

$$
\frac{\delta}{\delta\widetilde{\varepsilon}(\xi)}R[\widetilde{\varepsilon}(x)]
$$

we shall vary Eq. (14). Taking account of the fact that the main contribution to the integral (A3) comes from the region of integration over ξ of the order of the correlation length l_0, where the effect of random inhomogeneities on the dynamics of the waves is very small (the diffusion approximation), we obtain for the variational derivatives equations with initial conditions

$$
\begin{aligned}
&\frac{d}{dL}\frac{\delta}{\delta\widetilde{\varepsilon}(\xi)}R_{11}(L)=2i\lambda k\frac{\delta}{\delta\widetilde{\varepsilon}(\xi)}R_{11}(L),\\
&\left.\frac{\delta}{i\,\delta\widetilde{\varepsilon}(\xi)}R_{11}\right|_{L=\xi}=[\widetilde{\alpha}_2(1+R_{11})^2-2\delta(1+R_{11})R_{12}\\
&\qquad\qquad+\delta\lambda\widetilde{\alpha}_1R_{12}^2]_{L=\xi},\\
&\frac{d}{dL}\frac{\delta}{\delta\widetilde{\varepsilon}(\xi)}R_{22}(L)=2ik\frac{\delta}{\delta\widetilde{\varepsilon}(\xi)}R_{22}(L),\\
&\left.\frac{\delta}{i\,\delta\widetilde{\varepsilon}(\xi)}R_{22}\right|_{L=\xi}=[\lambda\widetilde{\alpha}_1(1+R_{22})^2-2\delta(1+R_{22})R_{12}\\
&\qquad\qquad+\delta\widetilde{\alpha}_2R_{12}^2]_{L=\xi},\\
&\frac{d}{dL}\frac{\delta}{\delta\widetilde{\varepsilon}(\xi)}R_{12}(L)=ik(1+\lambda)\frac{\delta}{\delta\widetilde{\varepsilon}(\xi)}R_{12}(L),\\
&\left.\frac{\delta}{i\,\delta\widetilde{\varepsilon}(\xi)}R_{12}\right|_{L=\xi}=[(\widetilde{\alpha}_2(1+R_{11})+\lambda\widetilde{\alpha}_1(1+R_{22})\\
&\qquad\qquad-\delta R_{12})R_{12}-(1+R_{11})(1+R_{22})]_{L=\xi},
\end{aligned}
$$

which are easily integrated. Since, in addition, on these integration scales

$$
\begin{aligned}
&R_{11}(\xi)=\exp[-2i\lambda k(L-\xi)]R_{11}(L),\\
&R_{22}(\xi)=\exp[-2ik(L-\xi)]R_{22}(L),\\
&R_{12}(\xi)=\exp[-ik(1+\lambda)(L-\xi)]R_{12}(L),
\end{aligned}
$$

the expressions for the variational derivatives can be written in the final form

$$
\begin{aligned}
&\frac{\delta}{i\,\delta\widetilde{\varepsilon}(L-\xi)}R_{11}=\{\widetilde{\alpha}_2(e^{i\lambda k\xi}+e^{-i\lambda k\xi}R_{11})^2\\
&\quad-2\delta(e^{-ik(1-\lambda)\xi}+e^{-ik(1+\lambda)\xi}R_{11})R_{12}\\
&\quad+\delta\lambda\widetilde{\alpha}_1e^{-2ik\xi}R_{12}^2\},\\
&\frac{\delta}{i\,\delta\widetilde{\varepsilon}(L-\xi)}R_{22}=\{\lambda\widetilde{\alpha}_1(e^{ik\xi}+e^{-ik\xi}R_{22})^2\\
&\qquad\qquad-2\delta(e^{ik(1-\lambda)\xi}+e^{-ik(1+\lambda)\xi}R_{22})R_{12}\\
&\qquad\qquad+\delta\widetilde{\alpha}_2e^{-2i\lambda k\xi}R_{12}^2\},\\
&\frac{\delta}{i\,\delta\widetilde{\varepsilon}(L-\xi)}R_{12}=\{-(e^{i\lambda k\xi}+e^{-i\lambda k\xi}R_{11})(e^{ik\xi}+e^{-ik\xi}R_{22})\\
&\qquad\qquad+\widetilde{\alpha}_2(1+e^{-2i\lambda k\xi}R_{11})R_{12}+\lambda\widetilde{\alpha}_1(1\\
&\qquad\qquad+e^{-2ik\xi}R_{22})R_{12}-\delta e^{-ik(1+\lambda)\xi}R_{12}^2\}.
\end{aligned}
\qquad \text{(A4)}
$$

As a result, we obtain the Fokker–Planck equation

$$
\begin{aligned}
&\frac{\partial}{\partial L}P_L(W_{11},W_{22},W_{12})\\
&\quad=\left\{\frac{\partial}{\partial W_{11}}[-D_1(1-W_{11})^2-4\delta^2D_4W_{12}+2\delta\right.\\
&\quad\times(D_3+D_4)W_{11}-\delta^2D_2W_{12}^2-4\delta^2D_3W_{11}W_{12}]+\frac{\partial}{\partial W_{22}}\\
&\quad\times[-D_2(1-W_{22})^2-4\delta^2D_4W_{12}+2\delta(D_3+D_4)W_{22}\\
&\quad-\delta^2D_1W_{12}^2-4\delta^2D_3W_{22}W_{12}]+\frac{\partial}{\partial W_{12}}[(D_1(1-W_{11})\\
&\quad+D_2(1-W_{22})+2\delta(D_3+2D_4)-\delta^2D_3W_{12})W_{12}-D_3\\
&\quad\times(1+W_{11}W_{22})-D_4(W_{11}+W_{22})]+\frac{\partial^2}{\partial W_{11}^2}W_{11}[D_1(1\\
&\quad-W_{11})^2+4\delta^2D_4W_{12}+4\delta^2D_3W_{11}W_{12}+\delta^2D_2W_{12}^2]\\
&\quad+\frac{\partial^2}{\partial W_{22}^2}W_{22}[D_2(1-W_{22})^2+4\delta^2D_4W_{12}\\
&\quad+4\delta^2D_3W_{22}W_{12}+\delta^2D_1W_{12}^2]+\frac{\partial^2}{\partial W_{12}^2}W_{12}\\
&\quad\times[W_{12}(D_1W_{11}+D_2W_{22}\\
&\quad+\delta^2D_3W_{12}-2\delta D_3)+D_3(1+W_{11}W_{22})+D_4\\
&\quad\times(W_{11}+W_{22})]+8\delta^2D_3\frac{\partial^2}{\partial W_{11}\partial W_{22}}W_{22}W_{11}W_{12}\\
&\quad-2\frac{\partial^2}{\partial W_{11}\partial W_{12}}W_{12}W_{11}[D_1(1-W_{11})+2\delta(D_3+D_4)\\
&\quad-2\delta^2D_3W_{12}]-2\frac{\partial^2}{\partial W_{22}\partial W_{12}}W_{12}W_{22}[D_2(1-W_{22})
\end{aligned}
$$

$$+2\delta(D_3+D_4)-2\delta^2 D_3 W_{12}\big]\Big\}\, P_L(W_{11},W_{22},W_{12}), \tag{A5}$$

where we have introduced the diffusion coefficients

$$D_1=2\tilde{\alpha}_2^2\int_0^\infty d\xi\; B_{\tilde{\varepsilon}}(\xi)\cos(2\lambda k\xi),$$

$$D_2=2(\lambda\tilde{\alpha}_1)^2\int_0^\infty d\xi\; B_{\tilde{\varepsilon}}(\xi)\cos(2k\xi),$$

$$D_3=2\int_0^\infty d\xi\; B_{\tilde{\varepsilon}}(\xi)\cos[k(1+\lambda)\xi],$$

$$D_4=2\int_0^\infty d\xi\; B_{\tilde{\varepsilon}}(\xi)\cos[k(1-\lambda)\xi],$$

which can be expressed in terms of the spectral function of the random process $\varepsilon(x)$ according to the formulas (19) presented in the text.

The derivation of Eq. (A5) also employed additional averaging over rapidly varying functions (which can be done if $k\lambda \gg D_i$).

[1] I. M. Lifshitz, S. A. Gredeskul, and L. A. Pastur, *Introduction to the Theory of Disordered Systems*, Wiley, N. Y., 1988 [Russian original, Nauka, Moscow, 1982].

[2] V. I. Klyatskin, *The Embedding Method in the Theory of Wave Propagation* [in Russian], Nauka, Moscow, 1986.

[3] V. I. Klyatskin and A. I Saichev, Usp. Fiz. Nauk **162**, 161 (1992) [Sov. Phys. Usp. **35**, 231 (1992)].

[4] V. I. Klyatskin, Prog. Opt. **33**, 1 (1994).

[5] D. Segupta, L. Piterbarg, and G. Reznik, Dyn. Atmos. Oceans **17**, 1 (1992).

[6] J. Pedlosky, *Geophysical Fluid Dynamics*, Springer-Verlag, N. Y., 1987, 2nd edition [Russian translation, Mir, Moscow, 1984].

[7] R. E. Thomson, J. Fluid Mech. **70**, 267 (1982).

[8] G. M. Reznik and T. B. Tsybaneva, Okeanologiya **34**, 5 (1994).

[9] V. I. Klyatskin, Izv. Ross. Akad. Nauk, Fiz. Atmos. Okeana **32**, 824 (1996).

[10] V. I. Klyatskin, *Stochastic Equations and Waves in Randomly Inhomogeneous Media* [in Russian], Nauka, Moscow, 1980.

Translated by M. E. Alferieff

Statistical theory of the propagation of optical radiation in turbulent media

V. I. Klyatskin

Institute of Atmospheric Physics, Russian Academy of Sciences, 109017 Moscow, Russia;
Pacific-Ocean Oceanological Institute, Far-East Division of the Russian Academy of Sciences,
690041 Vladivostok, Russia

I. G. Yakushkin

Institute of Atmospheric Physics, Russian Academy of Sciences, 109017 Moscow, Russia

(Submitted 29 November 1996)
Zh. Éksp. Teor. Fiz. **111**, 2044–2058 (June 1997)

The problem of the propagation of a plane light wave in a turbulent medium is studied on the basis of the ideas of statistical topography. A cluster (caustic) structure of the intensity of the wave field in a plane perpendicular to the direction of propagation of the wave is analyzed both in the region of weak intensity fluctuations and in the region of saturated fluctuations. The specific (per unit area) values of the total area of the regions where the intensity is greater than a fixed level, the fraction of the power confined in these regions, and the total perimeter and average number of such regions are estimated. It is shown that estimates of this kind can be made on the basis of a knowledge of the joint one-point probability distribution of the intensity and transverse gradient of the wave field.
[S1063-7761(97)00906-2]

1. INTRODUCTION

The spatial intensity distribution of light propagating turbulent medium can be regarded as a realization of a chaotic field. Such a realization of a two-dimensional field $I(x,\mathbf{R})$ in a fixed plane $x=\text{const}$, where x is the coordinate in the direction of propagation of the wave and $\mathbf{R}=\{y,z\}$ are the transverse coordinates, is reminiscent of a complicated mountain landscape with randomly distributed peaks, valleys, ridges, and passes.

The standard methods of statistical averaging, i.e., the calculation of average values of the form $\langle I(x,\mathbf{R})\rangle$, $\langle I(x,\mathbf{R}_1)I(x,\mathbf{R}_2)\rangle$, and so on, where $\langle ...\rangle$ denotes averaging over an ensemble of realizations of the random parameters of the medium (see, for example, Refs. 1 and 2), smooth out the qualitative features of typical realizations, and these characteristics often have nothing in common with the behavior of individual realizations. For example, the statistical average over all realizations of the intensity of the field of a plane wave is constant, while each individual realization of the field tends to become increasingly irregular in space. For example, a photograph of the transverse cross section of a laser beam propagating in a trubulent atmosphere is presented on the back of the jacket of the book Ref. 1. The appearance of a caustic structure of the wave field can be seen in this photograph. Such structures arise and are observed also in the case of the refraction and reflection of light by an agitated water surface; this corresponds to scattering by a so-called phase screen.

In summary, only when statistical averages of the type indicated above are taken together do they characterize the spatial scales of the regions where stochastic processes occur, and they say nothing about the important details of this region.

The complete statistics contains, of course, all information about the system. However, only some very simple statistical characteristics, mainly associated with one-point probability distributions, can be investigated in practice. For this reason, there arises the question of how to study the main quantitative and qualitative features of the spatial behavior of individual realizations if the local statistical characteristics of the system are known. The answer to this question is given by the methods of statistical topography of random fields.

The term ''statistical topography'' apparently was first introduced in Ref. 3, though the main ideas originate with much earlier works[4–6] (see also the review Ref. 7, where a detailed bibliography on this question is given). In these works, the statistical topography of random processes and fields was used to analyze statistical problems. The application of the methods of statistical topography to the dynamic problem of the diffusion of a passive impurity in a random velocity field was studied in Refs. 7–9. In Ref. 9 the problem of the diffusion of an impurity in a two-dimensional compressed medium was solved, a characteristic feature of which is the existence of cluster structures (compact high-concentration regions surrounded by extensive low-concentration regions).

The problem of light propagation in a random medium is mathematically similar to the problem of diffusion. In what follows, the structure of the wave field of optical radiation in a turbulent medium in a fixed plane $x=\text{const}$ will be studied on the basis of the idea of statistical topography as a function of the statistical parameters of the medium. We note that the theory of fluctuations of a random intensity field was first used in Refs. 10 and 11 (see also Ref. 12) to analyze the problem of wave propagation in a turbulent medium.

2. GENERAL REMARKS

We shall study light propagation in a randomly inhomogeneous medium on the basis of the scalar parabolic equation (see, for example, Refs. 1 and 2)

$$\frac{\partial}{\partial x} u(x,\mathbf{R}) = \frac{i}{2k} \Delta_{\mathbf{R}} u(x,\mathbf{R}) + \frac{ik}{2} \varepsilon(x,\mathbf{R}) u(x,\mathbf{R}),$$

$$u(0,\mathbf{R}) = u_0(\mathbf{R}), \tag{1}$$

where $\varepsilon(x,\mathbf{R})$ is the deviation of the permittivity from 1.

The transfer equation for the intensity $I(x,\mathbf{R}) = u(x,\mathbf{R})u^*(x,\mathbf{R})$ of the wave field follows from Eq. (1):

$$\frac{\partial}{\partial x} I(x,\mathbf{R}) = \frac{i}{2k} \nabla_R \cdot \{u^*(x,\mathbf{R})\nabla_R u(x,\mathbf{R}) - u(x,\mathbf{R})\nabla_R u^*(x,\mathbf{R})\},$$

$$I(0,\mathbf{R}) = I_0(\mathbf{R}). \tag{2}$$

If a complex phase of the wave field is now introduced according to the formula

$$u(x,\mathbf{R}) = A(x,\mathbf{R})\exp\{iS(x,\mathbf{R})\} = \exp\{\chi(x,\mathbf{R}) + iS(x,\mathbf{R})\}, \tag{1'}$$

where $\chi(x,\mathbf{R}) = \ln A(x,\mathbf{R})$ and $S(x,\mathbf{R})$ are, respectively, the amplitude and the phase of the wave, then Eq. (2) can be rewritten in the form

$$\frac{\partial}{\partial x} I(x,\mathbf{R}) + \frac{1}{k} \nabla_{\mathbf{R}} \cdot \{\nabla_{\mathbf{R}} S(x,\mathbf{R}) I(x,\mathbf{R})\} = 0,$$

$$I(0,\mathbf{R}) = I_0(\mathbf{R}). \tag{3}$$

It follows from Eq. (3) that in the general case of an arbitrary incident wave beam the total wave power in the plane $x = \text{const}$ is conserved:

$$E_0 = \int d\mathbf{R} I(x,\mathbf{R}) = \int d\mathbf{R} I_0(\mathbf{R}). \tag{4}$$

In the case when a plane wave with intensity I_0 is incident on a randomly inhomogeneous medium, we have for the spatially uniform fluctuations of the parameters of the medium, instead of the equality (4),

$$\langle I(x,\mathbf{R}) \rangle = I_0. \tag{5}$$

Equation (3) can be interpreted as a transfer equation for an impurity in a potential velocity field. However, this impurity can be regarded as passive only in the geometric-optics approximation, when the transverse gradient of the phase of the wave is described by the closed equation (see, for example, Refs. 1 and 2)

$$\left(\frac{\partial}{\partial x} + \frac{1}{k} \nabla_{\mathbf{R}} S(x,\mathbf{R}) \cdot \nabla_{\mathbf{R}} \right) \nabla_{\mathbf{R}} S(x,\mathbf{R}) = \frac{k}{2} \nabla_{\mathbf{R}} \varepsilon(x,\mathbf{R}). \tag{3'}$$

In the general case the impurity is active when diffraction effects are taken into account. We note that in the geometric-optics approximation the probability density of the formation of caustics as a function of the distance traveled by the wave was estimated in Refs. 13–16.

As indicated in the introduction, a fundamental feature of the diffusion of an impurity in a potential velocity field is the existence of a cluster structure of the field intensity, which in the case at hand appears as a caustic structure due to random focusings and defocusings in the random medium.

In the present case, the method employed in Ref. 9, based on an approximation of a random velocity field by a field which is δ-correlated in x, to find the statistical parameters characterizing this cluster structure is inapplicable, since the longitudinal correlation length for the phase of the wave is of the order of the path length. In addition, in our situation the initial dynamical equation (3) becomes virtually useless for analyzing the structure of the intensity field in a random medium. However, the statistical-topography approach is applicable if the one-point probability density for the wave intensity is known in the regions of both weak and strong fluctuations of the wave field obtained by solving Eq. (1).

3. ELEMENTS OF THE STATISTICAL TOPOGRAPHY OF A RANDOM INTENSITY FIELD

The main object of study in statistical topography, just as in ordinary topography of mountain massifs, is a system of contours—lines of intensity level in a fixed plane $x = \text{const}$ (two-dimensional case)—determined by the relation $I(x,\mathbf{R}) = I = \text{const}$.

To analyze the system of contours it is convenient to introduce the delta function which is concentrated on them

$$\Phi(x,R;I) = \delta(I(x,R) - I), \tag{6}$$

and is a functional of the parameters of the medium. The average value of the functional (6) over an ensemble of realizations determines the one-point probability density

$$P(x,R;I) = \langle \Phi(x,R;I) \rangle = \langle \delta(I(x,R) - I) \rangle. \tag{7}$$

For example, quantities such as the total area of the regions bounded by the contour lines $I(x,\mathbf{R}) > 1$

$$S(x,I) = \int \theta(I(x,\mathbf{R}) - I) d\mathbf{R} = \int_I^\infty d\tilde{I} \int d\mathbf{R} \Phi(x,\mathbf{R};\tilde{I}), \tag{8}$$

and the total power of the field in these regions,

$$E(x,I) = \int I(x,\mathbf{R}) \theta(I(x,\mathbf{R}) - I) d\mathbf{R} = \int_I^\infty \tilde{I} d\tilde{I} \int d\mathbf{R} \Phi(x,\mathbf{R};\tilde{I}), \tag{9}$$

are expressed in terms of the function (6). Here $\theta(z)$ is the Heaviside step function. The statistical averages of these quantities can be expressed in terms of the probability density (7).

Additional information about the detailed structure of the field can be obtained by including in the analysis the transverse intensity gradient $\mathbf{p}(x,\mathbf{R}) = \nabla_R I(x,\mathbf{R})$. For example, the quantity

$$L(x,I)=\int d\mathbf{R}|\mathbf{p}(x,\mathbf{R})|\,\delta(I(x,\mathbf{R})-I) \qquad (10)$$

describes the total length of the contours $I(x,\mathbf{R})=I=\text{const}$. We note that to calculate the average value in Eq. (10) it is now necessary to know the joint one-point probability density of the field $I(x,\mathbf{R})$ and its gradient $\mathbf{p}(x,\mathbf{R})$

$$P(x,R;I,\mathbf{p})=\langle\,\delta(I(x,\mathbf{R})-I)\,\delta(\mathbf{p}(x,\mathbf{R})-\mathbf{p})\rangle.$$

Including in the analysis second-order spatial derivatives makes it possible to estimate, for example, the total number of contours $I(x,\mathbf{R})=I=$ const by means of the approximate (to within non-closed lines) formula[6]

$$\begin{aligned}N(x,I)&=N_1(x,I)-N_2(x,I)\\&=\frac{1}{2\pi}\int d\mathbf{R}\kappa(x,\mathbf{R};I)|\mathbf{p}(x,\mathbf{R})|\,\delta(I(x,\mathbf{R})-I),\end{aligned} \qquad (11)$$

where $N_1(x,I)$ and $N_2(x,I)$ are the numbers of contours for which the vector $\mathbf{p}$ is directed along the inner and outer normals, respectively, and $\kappa(x,\mathbf{R};I)$ is the curvature of a level line:

$$\begin{aligned}\kappa(x,\mathbf{R};I)=\{&-p_y^2(x,\mathbf{R})\partial^2I(x,\mathbf{R})/\partial z^2\\&-p_z^2(x,\mathbf{R})\partial^2I(x,\mathbf{R})/\partial y^2\\&+2p_y(x,\mathbf{R})p_z(x,\mathbf{R})\partial^2I(x,\mathbf{R})/\partial y\partial z\}\\&\times[p^3(x,\mathbf{R})]^{-1}.\end{aligned} \qquad (12)$$

For a plane incident wave all one-point probability densities are independent of the variable $\mathbf{R}$ on account of the spatial uniformity and the corresponding statistical averages (5)–(11) (without integrating over $\mathbf{R}$) will describe the specific (per unit area) values of these quantities. A natural length scale in the plane $x=\text{const}$, which does not depend on the parameters of the medium, is the size of the first Fresnel zone $L_f(x)=\sqrt{x/k}$, determining the size of the light–shadow transition region in diffraction by the edge of an opaque screen (see, for example, Ref. 1). Then the specific average values of the contour length and the estimate of their average number will be described by the dimensionless expressions

$$\langle l(x,I)\rangle=L_f(x)\langle|\mathbf{p}(x,\mathbf{R})|\,\delta(I(x,\mathbf{R})-I)\rangle, \qquad (10')$$

$$\langle n(x,I)\rangle=\frac{1}{2\pi}\,L_f^2(x)\langle\kappa(x,\mathbf{R};I)|\mathbf{p}(x,\mathbf{R})|\,\delta(I(x,\mathbf{R})-I)\rangle. \qquad (11')$$

The formula (11′) describes the average difference of the number of contours with opposite orientation of the normal vectors in the first Fresnel zone.

4. STATISTICAL ANALYSIS OF THE CLUSTER PATTERN OF THE INTENSITY OF A WAVE FIELD

Let us assume that the random field $\varepsilon(x,\mathbf{R})$ is a Gaussian homogeneous and isotropic field with correlation and spectral functions

$$\begin{aligned}B_\varepsilon(x_1-x_2,\mathbf{R}_1-\mathbf{R}_2)&=\langle\varepsilon(x_1,\mathbf{R}_1)\varepsilon(x_2,\mathbf{R}_2)\rangle\\&=\int_{-\infty}^{\infty}dq_x\int d\mathbf{q}\Phi_s(q_x,\mathbf{q})\\&\quad\times\exp\{iq_x(x_1-x_2)+i\mathbf{q}\cdot(\mathbf{R}_1-\mathbf{R}_2)\},\\ \Phi_\varepsilon(q_x\cdot\mathbf{q})&=\frac{1}{(2\pi)^3}\int_{-\infty}^{\infty}dx\int d\mathbf{R}B_s(x,\mathbf{R})\\&\quad\times\exp\{-iq_xx-i\mathbf{q}\cdot\mathbf{R}\}.\end{aligned} \qquad (13)$$

In what follows, we shall study the incidence of a plane wave with unit intensity. In this case the wave field will be statistically homogeneous in the plane $x=\text{const}$ and its statistical characteristics will all be independent of $\mathbf{R}$.

4.1. Region of weak intensity fluctuations

In the general case the intensity of the wave field will have the structure (1′), i.e.,

$$I(x,\mathbf{R})=A^2(x,\mathbf{R})=\exp\{2\chi(x,\mathbf{R})\}. \qquad (14)$$

In the region of weak fluctuations, the amplitude level $\chi(x,\mathbf{R})$ is a Gaussian random field. Moreover, it follows from the equality (5) that the average value of the amplitude level is determined by its variance, specifically,

$$\langle\chi(x,\mathbf{R})\rangle=-\sigma_\chi^2(x).$$

We now introduce a parameter, usually termed the scintillation index (see, for example, Ref. 1),

$$\beta_0(x)=4\sigma_\chi^2(x). \qquad (15)$$

Then the variance of the intensity of the wave field for $\beta_0(x)\ll 1$ will be described by the formula

$$\sigma_I^2(x)=\langle I^2(x,\mathbf{R})\rangle-1=\beta_0(x). \qquad (16)$$

Therefore the one-point probability distribution of the field $\chi(x,\mathbf{R})$ has the form

$$\begin{aligned}P(x;\chi)&=\langle\,\delta(\chi(x,\mathbf{R})-\chi)\rangle=\sqrt{\frac{2}{\pi\beta_0(x)}}\\&\quad\times\exp\left\{-\frac{2}{\beta_0(x)}\left(\chi+\frac{1}{4}\,\beta_0(x)\right)^2\right\}.\end{aligned} \qquad (17)$$

Thus, the intensity of the wave field is a log-normal random field, and its one-point probability density is determined by the expression

$$\begin{aligned}P(x;I)&=\langle\,\delta(I(x,\mathbf{R})-I)\rangle=\frac{1}{I\sqrt{2\pi\beta_0(x)}}\\&\quad\times\exp\left\{-\frac{1}{2\beta_0(x)}\,\ln^2\!\left(I\,\exp\!\left(\frac{1}{2}\,\beta_0(x)\right)\right)\right\}.\end{aligned} \qquad (18)$$

The region of weak intensity fluctuations is bounded by the values of the parameter $\beta_0(x)\leqslant 1$.

As is well known (see, for example, Ref. 17), for an arbitrary random process $I(x)$ it is always possible to introduce a determinate curve, called a typical realization $I^*(x)$, such that on any interval (X_1,X_2), where

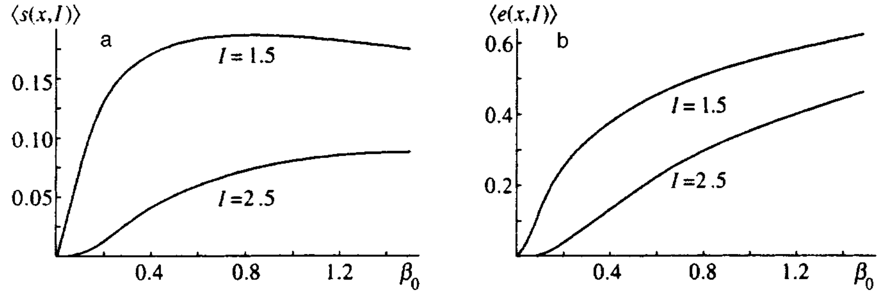


FIG. 1. Average specific area (a) and power (b) versus the parameter $\beta_0(x)$.

$X_i=\beta_0(x_i)$, the average interval where the inequality $I(x)<I^*(x)$ holds is also the average interval where the reverse inequality $I(x)>I^*(x)$ holds and equals $(X_2-X_1)/2$. For the log-normal process (18) such a typical realization of the random intensity is a curve decreasing exponentially with distance

$$I^*(x)=\exp\left\{-\frac{1}{2}\,\beta_0(x)\right\},$$

and the statistics (for example, the moment functions $\langle I^n(x,\mathbf{R})\rangle$) is formed by large fluctuations of the process $I(x,\mathbf{R})$ relative to this curve.

Furthermore, different majorant estimates also exist for realizations of a log-normal process $I(x)$.[17] For example, for separate realizations of the intensity of a wave field the inequality

$$I(x)<4\exp\left\{-\frac{1}{4}\,\beta_0(x)\right\}$$

holds with probability $p=1/2$ in the entire interval of distances $x\in(0,\infty)$. These circumstances all indicate the formation of a cluster structure of the intensity of the wave field.

As indicated above, if the probability density (18) is known, then some quantitative characteristics of such cluster formations can be determined. For example, the average specific area of the regions with $I(x,\mathbf{R})>I$ equals

$$\langle s(x,I)\rangle=\int_I^\infty d\tilde{I}P(x;\tilde{I})=\Phi\left(\frac{1}{\sqrt{2\beta_0(x)}}\times\ln\left(\exp\left(-\frac{1}{2}\,\beta_0(x)\right)\Big/ I\right)\right), \tag{19}$$

and the specific average power concentrated in these regions is described by the expression

$$\langle e(x,I)\rangle=\int_I^\infty d\tilde{I}\tilde{I}P(x;\tilde{I})=\Phi\left(\frac{1}{\sqrt{2\beta_0(x)}}\ln\left(\exp\left(\frac{1}{2}\,\beta_0(x)\right)\Big/ I\right)\right), \tag{20}$$

where

$$\Phi(z)=\frac{1}{\sqrt{\pi}}\int_{-\infty}^{z}\exp(-y^2)dy$$

is the standard error function.

The character of the spatial evolution of the cluster structure accompanying a change in the parameter $\beta_0(x)$ depends strongly on the fixed level I. In the most interesting case $I>1$, in the initial plane $\langle s(0,I)\rangle=0$ and $\langle e(0,I)\rangle=0$. As $\beta_0(x)$ increases, small cluster regions where $I(x,\mathbf{R})>I$, which remain practically unchanged over some distance and intensively attract a large fraction of the total power into their domain, are formed. Then, the areas of these regions decrease as $\beta_0(x)$ increases, and the power contained in them increases, which corresponds to an increase in the average brightness inside these regions. These processes are associated with the focusing of radiation by separate sections of the medium. Plots of the variation of the functions $\langle s(x,I)\rangle$ and $\langle e(x,I)\rangle$ as a function of the parameter $\beta_0(x)$ from the indicated range of values are displayed in Figs. 1a and 1b. The maximum average specific area is reached for $\beta_0(x)=2\ln(I)$ and

$$\langle s(x,I)\rangle_{\max}=\Phi\left(-\frac{1}{\sqrt{\ln(I)}}\right).$$

The average power for this value of $\beta_0(x)$ is $\langle e(x,I)\rangle=1/2$.

In the region of weak intensity fluctuations the spatial gradient of the amplitude level $\nabla_R\chi(x,\mathbf{R})$ is statistically independent of $\chi(x,\mathbf{R})$. This makes it possible to calculate both the specific average length of the contours $I(x,\mathbf{R})=I$ and to estimate the specific average number of such contours. Indeed, the probability density for the gradient of the amplitude level $\mathbf{q}(x,\mathbf{R})=\nabla_R\chi(x,\mathbf{R})$ in the region of weak fluctuations is a Gaussian distribution

$$P(x;\mathbf{q})=\langle\delta(\nabla_R\chi(x,\mathbf{R})-\mathbf{q})\rangle=\frac{1}{\pi\sigma_q^2}\exp\left\{-\frac{\mathbf{q}^2}{\sigma_q^2(x)}\right\}, \tag{21}$$

where $\sigma_q^2(x)=\langle\mathbf{q}^2(x,\mathbf{R})\rangle$ is the variance of the gradient of the amplitude level.

Therefore we obtain from the formula (10′) for the specific average length of the contours the expression

$$\begin{aligned}\langle l(x,I)\rangle&=L_f(x)\langle|\mathbf{p}(x,\mathbf{R})|\,\delta(I(x,\mathbf{R})-I)\rangle\\&=2L_f(x)\langle|\mathbf{q}(x,\mathbf{R})|I\delta(I(x,\mathbf{R})-I)\rangle\\&=2L_f(x)\langle|\mathbf{q}(x,\mathbf{R})|\rangle IP(x,I)\\&=L_f(x)\sqrt{\pi\sigma_q^2(x)}IP(x,I).\end{aligned} \tag{22}$$

Similarly, we have for the specific average number of contours from the formula (11′)

$$\begin{aligned}\langle n(x,I)\rangle &= \frac{1}{2\pi}L_f^2(x)\langle \kappa(x,\mathbf{R};I)\left|\mathbf{p}(x,\mathbf{R})\right|\delta(I(x,\mathbf{R})-I)\rangle\\ &= -\frac{1}{2\pi}L_f^2(x)I\langle \Delta\chi(x,\mathbf{R})\delta(I(x,\mathbf{R})-I)\rangle\\ &= -\frac{1}{\pi}L_f^2(x)\langle \mathbf{q}^2(x,\mathbf{R})\rangle I\frac{\partial}{\partial I}IP(x,I)\\ &= \frac{L_f^2(x)\sigma_{\mathbf{q}}^2(x)}{\pi\beta_0(x)}\ln\left(I\exp\left(\frac{1}{2}\beta_0(x)\right)\right)IP(x,I).\end{aligned} \tag{23}$$

We note that the expression (23) vanishes for $I=I_0(x)=\exp\{-(1/2)\beta_0(x)\}$. This means that for a given intensity level the average specific number of contours bounding the region $I(x,\mathbf{R})>I_0$ is identical to the average specific number of contours for which $I(x,\mathbf{R})<I_0$.

General expressions containing the parameter $\beta_0(x)$ characterizing the properties of the medium were obtained above. In the region of weak fluctuations the amplitude level and its gradient with respect to the transverse coordinates are described on the basis of the method of smooth perturbations (see, for example, Refs. 1 and 2) by the formulas

$$\begin{aligned}\chi(x,\mathbf{R}) &= \frac{k}{2}\int_0^x d\xi\int d\mathbf{q}e^{i\mathbf{qR}}\varepsilon_{\mathbf{q}}(\xi)\sin\frac{q^2}{2k}(x-\xi),\\ \nabla_R\chi(x,\mathbf{R}) &= i\frac{k}{2}\int_0^x d\xi\int \mathbf{q}d\mathbf{q}e^{i\mathbf{qR}}\varepsilon_{\mathbf{q}}(\xi)\sin\frac{q^2}{2k}(x-\xi),\end{aligned} \tag{24}$$

where

$$\begin{aligned}\varepsilon(x,\mathbf{R}) &= \int d\mathbf{q}e^{i\mathbf{qR}}\varepsilon_{\mathbf{q}}(x),\\ \varepsilon_{\mathbf{q}}(x) &= \frac{1}{(2\pi)^2}\int d\mathbf{q}e^{-i\mathbf{qR}}\varepsilon(x,\mathbf{R}).\end{aligned} \tag{25}$$

In the approximation of a field $\varepsilon(x,\mathbf{R})$ which is δ-correlated with respect to x, the correlation function (13) of the field is approximated by the expression (see, for example, Refs. 1 and 2)

$$B_\varepsilon(x,\mathbf{R})=\delta(x)A(\mathbf{R}),\quad A(\mathbf{R})=\int_{-\infty}^{\infty}dxB_\varepsilon(x,\mathbf{R}) \tag{26}$$

and the random field $\varepsilon_{\mathbf{q}}(x)$ can be assumed to be Gaussian with correlation function

$$\langle\varepsilon_{\mathbf{q}_1}(x_1)\varepsilon_{\mathbf{q}_2}(x_2)\rangle=2\pi\delta(x_1-x_2)\delta(\mathbf{q}_1+\mathbf{q}_2)\Phi_\varepsilon(0,\mathbf{q}_1), \tag{27}$$

where $\Phi_\varepsilon(0,\mathbf{q})$ is the three-dimensional spectral function (13) of the two-dimensional vector $\mathbf{q}$.

For a turbulent medium the function $\Phi_\varepsilon(0,\mathbf{q})$ has the form (see, for example, Refs. 1 and 2)

$$\Phi_\varepsilon(0,\mathbf{q})=AC_\varepsilon^2q^{-11/3}\exp\{-(\mathbf{q}^2/\boldsymbol{\kappa}_m^2)\}. \tag{28}$$

where $A=0.033$ is a numerical constant, C_ε^2 is the structure constant of the permittivity fluctuations which depends on the external parameters of the flows, and κ_m is a wave number corresponding to the turbulence microscale. Therefore

$$\begin{aligned}\beta_0(x)=4\sigma_\chi^2(x)=2k^2\pi^2x\int_0^\infty dqq\Phi_\varepsilon(q)\\ \times\left[1-\frac{k}{q^2x}\sin\left(\frac{q^2}{k}x\right)\right],\end{aligned} \tag{29}$$

if the turbulent medium occupies all space. If the field $\varepsilon(x,\mathbf{R})$ is different from zero only in a thin layer $\Delta x\ll x$ (random phase screen), then

$$\begin{aligned}\beta_0(x)=4\sigma_\chi^2(x)=2k^2\pi^2\Delta x\int_0^\infty dqq\Phi_\varepsilon(q)\\ \times\left[1-\cos\left(\frac{q^2}{k}x\right)\right].\end{aligned} \tag{29′}$$

If the so-called wave parameter (see, for example, Ref. 1) satisfies $D(x)=\kappa_m^2x/k\gg1$, we obtain for the parameter $\beta_0(x)$ the expressions

$$\begin{aligned}&\beta_0(x)=0.307C_\varepsilon^2k^{7/6}x^{11/6}\quad(\Delta x=x),\\ &\beta_0(x)=0.563C_\varepsilon^2k^{7/6}x^{5/6}\Delta x\quad(\Delta x\ll x).\end{aligned} \tag{30}$$

Similarly, we have for the variance of the gradient of the amplitude level with $D(x)\gg1$ in the case of a turbulent medium occupying all space

$$\begin{aligned}\sigma_{\mathbf{q}}^2(x)&=\frac{k^2\pi^2x}{2}\int_0^\infty dqq^3\Phi_\varepsilon(q)\left[1-\frac{k}{q^2x}\sin\left(\frac{q^2}{k}x\right)\right]\\ &=\frac{1.476}{L_f^2(x)}D^{1/6}(x)\beta_0(x).\end{aligned} \tag{31}$$

Now $\langle l(x,I)\rangle$ and $\langle n(x,I)\rangle$ described by Eqs. (22) and (23) can be calculated as functions of the parameters $\beta_0(x)$ and $D(x)$. The plots of these dependences on $\beta_0(x)$ are displayed in Figs. 2a and 2b.

The dependence of the average specific length of the level lines and the average specific number of contours on the turbulence microscale indicates the existence of small ripples which are superposed on a larger-scale random relief. These ripples have no effect on the redistribution of areas and powers, but they cause the level lines to be strongly irregular and lead to the appearance of small contours.

As indicated above, the description obtained is valid for $\beta_0(x)\leq1$. As the parameter $\beta_0(x)$ increases further, the method of smooth perturbations becomes invalid and the nonlinear character of the equation for the complex phase of the wave field must be taken into consideration. This region of fluctuations, called the region of strong focusings, is very difficult to investigate analytically. As the parameter $\beta_0(x)$ increases further ($\beta_0(x)\geq10$), the statistical characteristics of the intensities saturate and this range of the parameter $\beta_0(x)$ is called the region of strong intensity fluctuations.

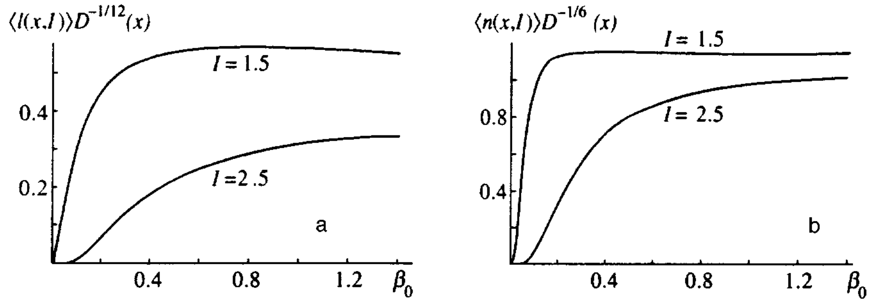


FIG. 2. Average specific contour length (a) and average specific number of contours (b) versus the parameter $\beta_0(x)$.

4.2. Region of strong intensity fluctuations

It is well known that in the region of strong fluctuations the moments of the intensity are described by the asymptotic formula (see, for example, Ref. 2)

$$\langle I^n(x)\rangle = n![1+n(n-1)(\beta(x)-1)/4], \tag{32}$$

where $\beta(x)$ is the variance of the intensity, i.e.,

$$\beta(x)=\langle I^2(x)\rangle-1.$$

For a turbulent medium we have (see, for example, Ref. 2)

$$\beta(x)=1+0.861\beta_0^{-2/5}(x) \quad (\Delta x = x),$$

$$\beta(x)=1+0.429\beta_0^{-2/5}(x) \quad (\Delta x \ll x), \tag{33}$$

where the parameter $\beta_0(x)$ is described by Eqs. (30).

The equation (32) leads to an intensity probability density with singularities. To avoid the singularities, the formula (32) can be approximated by the expression (see, for example, Ref. 18)

$$\langle I^n(x)\rangle = n!\ \exp\{n(n-1)(\beta(x)-1)/4\}, \tag{34}$$

which corresponds to a probability density of the form (see, for example, Refs. 18 and 19)

$$P(x,I)=\frac{1}{\sqrt{\pi(\beta(x)-1)}}\int_0^\infty dz \times \exp\left\{-zI-\frac{(\ln z-(\beta(x)-1)/4)^2}{\beta(x)-1}\right\}. \tag{35}$$

We note that the probability distribution (35) is inapplicable in a small neighborhood of $I\sim 0$ (the neighborhood is smaller, the larger the parameter $\beta_0(x)$). This is because Eq. (35) gives infinite moments of the quantity $1/I(x,\mathbf{R})$. However, for a finite value of $\beta_0(x)$ (regardless of how large) the quantities $\langle I^n(x,\mathbf{R})\rangle$ are finite and therefore the equality $P(x,0)=0$ should hold. Of course, the existence of such a narrow neighborhood of the point $I\sim 0$ has no effect on the behavior of the moments $\langle I^n(x,\mathbf{R})\rangle$ (34) for large values of the parameter $\beta_0(x)$.

It follows from the expression (35) that the average specific area of the regions where $I(x,\mathbf{R})>I$ holds equals

$$\langle s(x,I)\rangle=\frac{1}{\sqrt{\pi(\beta(x)-1)}}\int_0^\infty \frac{dz}{z} \times \exp\left\{-zI-\frac{(\ln z-(\beta(x)-1)/4)^2}{\beta(x)-1}\right\}, \tag{36}$$

and the specific average power concentrated in these regions is described by the expression

$$\langle e(x,I)\rangle=\frac{1}{\sqrt{\pi(\beta(x)-1)}}\int_0^\infty \frac{dz}{z}\left(I+\frac{1}{z}\right) \times \exp\left\{-zI-\frac{(\ln z-(\beta(x)-1)/4)^2}{\beta(x)-1}\right\}. \tag{37}$$

Plots of the functions (36) and (37) versus the parameter $\beta(x)$ are presented in Figs. 3a and 3b. Note the very weak dependence of the parameter $\beta(x)$ on $\beta_0(x)$. Thus, the passage to the limit $\beta_0(x)\to\infty$ corresponds to the parameter $\beta(x)=1$ and $\beta(x)=1.861$ corresponds to $\beta_0(x)=1$.

The asymptotic expressions (32)–(37) describe the transition to the region of saturated intensity fluctuations ($\beta(x)\to 1$). Correspondingly, we have in this region

$$P(I)=e^{-I}, \quad \langle s(I)\rangle = e^{-I}, \quad \langle e(I)\rangle=(I+1)e^{-I} \tag{38}$$

and therefore the fractions of the total average area and average power within the level lines depend only on the value of I. For large I these fractions are very small.

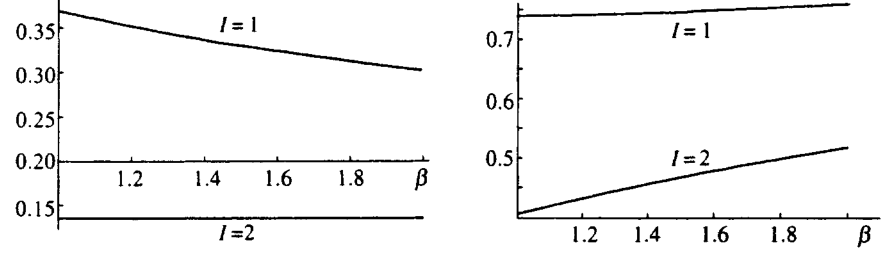


FIG. 3. Plots of the average specific area (a) and power (b) in the region of strong intensity fluctuations versus the parameter $\beta(x)$.

The exponential probability distribution (38) means that the complex field $u(x,\mathbf{R})$ is a Gaussian random field. Then

$$u(x,\mathbf{R})=A(x,\mathbf{R})e^{iS(x,\mathbf{R})}=u_1(x,\mathbf{R})+iu_2(x,\mathbf{R}), \tag{39}$$

where $u_1(x,\mathbf{R})$ and $u_2(x,\mathbf{R})$ are, respectively, the real and imaginary parts. The intensity of the wave field is

$$I(x,\mathbf{R})=A^2(x,\mathbf{R})=u_1^2(x,\mathbf{R})+u_2^2(x,\mathbf{R}).$$

The Gaussian nature of the complex field $u(x,\mathbf{R})$ means that the random fields $u_1(x,\mathbf{R})$ and $u_2(x,\mathbf{R})$ are also Gaussian, statistically independent, fields with variances

$$\langle u_1^2(x,\mathbf{R})\rangle=\langle u_2^2(x,\mathbf{R})\rangle=\frac{1}{2}. \tag{40}$$

It is natural to assume that their gradients $\mathbf{q}_1(x,\mathbf{R})=\nabla_R u_1(x,\mathbf{R})$ and $\mathbf{q}_2(x,\mathbf{R})=\nabla_R u_2(x,\mathbf{R})$ are also statistically independent of the fields $u_1(x,\mathbf{R})$ and $u_2(x,\mathbf{R})$, and in the plane $\mathbf{R}$ they are Gaussian homogeneous and isotropic fields with variances

$$\sigma_{\mathbf{q}}^2(x)=\langle \mathbf{q}_1^2(x,\mathbf{R})\rangle=\langle \mathbf{q}_2^2(x,\mathbf{R})\rangle. \tag{41}$$

Therefore the joint probability distribution of the fields u_1, u_2 and their gradients $\mathbf{q}_1$, $\mathbf{q}_2$ has the form

$$P(u_1,u_2,\mathbf{q}_1,\mathbf{q}_2;x)=\frac{1}{\pi^3\sigma_{\mathbf{q}}^4(x)}\times\exp\left\{-\left(u_1^2+u_2^2+\frac{\mathbf{q}_1^2+\mathbf{q}_2^2}{\sigma_{\mathbf{q}}^2(x)}\right)\right\}. \tag{42}$$

Let us now consider the joint probability distribution of the intensity $I(x,\mathbf{R})$ of the wave field and amplitude gradient

$$\boldsymbol{\kappa}(x,\mathbf{R})=\nabla_{\mathbf{R}}A(x,\mathbf{R})=\frac{u_1(x,\mathbf{R})\mathbf{q}_1(x,\mathbf{R})+u_2(x,\mathbf{R})\mathbf{q}_2(x,\mathbf{R})}{\sqrt{u_1^2(x,\mathbf{R})+u_2^2(x,\mathbf{R})}}.$$

We have

$$\begin{aligned}P(I,\boldsymbol{\kappa};x)&=\langle\delta(I(x,\mathbf{R})-I)\,\delta(\boldsymbol{\kappa}(x,\mathbf{R})-\boldsymbol{\kappa})\rangle_{u_1\mathbf{q}_1}\\&=\frac{1}{\pi^3\sigma_{\mathbf{q}}^4(x)}\int_{-\infty}^{\infty}du_1\int_{-\infty}^{\infty}du_2\int d\mathbf{q}_1\int d\mathbf{q}_2\\&\quad\times\exp\left\{-\left(u_1^2+u_2^2+\frac{\mathbf{q}_1^2+\mathbf{q}_2^2}{\sigma_{\mathbf{q}}^2(x)}\right)\right\}\delta(u_1^2+u_2^2-I)\\&\quad\times\delta\left(\frac{u_1\mathbf{q}_1+u_2\mathbf{q}_2}{\sqrt{u_1^2+u_2^2}}-\boldsymbol{\kappa}\right)\\&=\frac{1}{2\pi\sigma_{\mathbf{q}}^2(x)}\exp\left\{-I-\frac{\boldsymbol{\kappa}^2}{2\sigma_{\mathbf{q}}^2(x)}\right\}.\end{aligned} \tag{43}$$

Therefore the transverse amplitude gradient is statistically independent of the intensity of the wave field and is a Gaussian random field with variance

$$\langle\boldsymbol{\kappa}^2(x,\mathbf{R})\rangle=2\sigma_{\mathbf{q}}^2(x). \tag{44}$$

We note that the transverse amplitude gradient is also statistically independent of the second derivatives of the intensity of the wave field with respect to the transverse coordinates.

Therefore from Eq. (10) we obtain for the specific average length of the contours in the region of saturated intensity fluctuations the expression

$$\begin{aligned}\langle l(x,I)\rangle&=L_f(x)\langle|\mathbf{p}(x,\mathbf{R})|\,\delta(I(x,\mathbf{R})-I)\rangle\\&=2L_f(x)\sqrt{I}\langle|\boldsymbol{\kappa}(x,\mathbf{R})|\,\delta(I(x,\mathbf{R})-I)\rangle\\&=2L_f(x)\sqrt{I}\langle|\boldsymbol{\kappa}(x,\mathbf{R})|\rangle P(x,I)\\&=L_f(x)\sqrt{2\pi\sigma_{\mathbf{q}}^2(x)}Ie^{-I}.\end{aligned} \tag{45}$$

The maximum value in Eq. (45) is reached for $I=1/\sqrt{2}$.

Similarly, we have as an estimate of the average specific number of contours in this region

$$\begin{aligned}\langle n(x,I)\rangle&=\frac{L_f^2(x)}{2\pi}\langle\boldsymbol{\kappa}(x,\mathbf{R};I)|\mathbf{p}(x,\mathbf{R})|\,\delta(I(x,\mathbf{R})-I)\rangle\\&=-\frac{L_f^2(x)}{2\pi}\sqrt{I}\langle\Delta A(x,\mathbf{R})\,\delta(I(x,\mathbf{R})-I)\rangle\\&=-\frac{L_f^2(x)}{\pi}\langle\boldsymbol{\kappa}^2(x,\mathbf{R})\rangle\sqrt{I}\,\frac{\partial}{\partial I}\sqrt{I}P(x,I)\\&=-\frac{2L_f^2(x)\sigma_{\mathbf{q}}^2(x)}{\pi}\sqrt{I}\,\frac{\partial}{\partial I}\sqrt{I}e^{-I}\\&=\frac{2L_f^2(x)\sigma_{\mathbf{q}}^2(x)}{\pi}\left(I-\frac{1}{2}\right)e^{-I}.\end{aligned} \tag{46}$$

The maximum value in Eq. (46) is reached for $I=3/2$, and the level at which the average specific number of contours bounding the region $I(x,\mathbf{R})>I_0$ equals the average specific number of contours for which $I(x,\mathbf{R})<I_0$ is equal to $I_0=1/2$ in the present case.

We note that Eq. (46) is inapplicable in a small neighborhood of $I\sim 0$. For $I=0$ we have $\langle n(x,0)\rangle=0$ (refer to the discussion of Eq. (35)).

In the region of strong intensity fluctuations the second-order coherence function does not depend on diffraction phenomena and is described by the expression (see, for example, Refs. 1 and 2)

$$\begin{aligned}\Gamma_2(x,\mathbf{R}-\mathbf{R}')&=\langle u(x,\mathbf{R})u^*(x,\mathbf{R}')\rangle\\&=\langle u_1(x,\mathbf{R})u_1(x,\mathbf{R}')+u_2(x,\mathbf{R})u_2(x,\mathbf{R}')\rangle\\&=\exp\left\{-\frac{k^2x}{4}D(|\mathbf{R}-\mathbf{R}'|)\right\},\end{aligned} \tag{47}$$

where (see Eq. (26))

$$D(\mathbf{R})=A(0)-A(\mathbf{R}). \tag{48}$$

Therefore the quantity $\sigma_{\mathbf{q}}^2(x)$ from Eq. (41) is determined by the expression

$$\sigma_{\mathbf{q}}^2(x)=\frac{k^2x}{8}\Delta_{\mathbf{R}}D(\mathbf{R})\Big|_{\mathbf{R}=0}=-\frac{k^2x}{8}\Delta_{\mathbf{R}}A(\mathbf{R})\Big|_{\mathbf{R}=0}. \tag{49}$$

For turbulent fluctuations $\varepsilon(x,\mathbf{R})$ we have

$$\sigma_{\mathbf{q}}^2(x)=\frac{1.476}{L_f^2(x)}D^{1/6}(x)\beta_0(x). \tag{50}$$

As one can see from Eq. (50), in the saturated-fluctuations regime the average length of the level lines and the average number of contours continue to grow with the parameter $\beta_0(x)$, though the average areas and powers limited by them and contained in them remain constant. This is because interference of partial waves arriving from different directions plays a determining role in this regime.

The dynamical picture of the behavior of the level lines depends on the relations between the radiation-focusing and defocusing processes in separate sections of a turbulent medium.[1,20] Focusing of large-scale inhomogeneities is manifested as high peaks on a random intensity relief. In the maximum-focusing regime $(\beta_0(x)\sim 1)$, approximately half the total power of a wave is concentrated in high, narrow peaks. As the parameter $\beta_0(x)$ increases, defocusing of radiation predominates and results in broadening of the high peaks and the formation of a strongly irregular (interference) relief with a large number of peaks of the level $I\sim 1$.

The average length of a level line and the average length of contours depends on, besides the parameter $\beta_0(x)$, the parameter $D(x)$, i.e., they increase with decreasing microscale of the inhomogeneities. This is because a small ripple structure, due to scattering by small irregularities, is superposed on the larger-scale relief.

5. CONCLUSIONS

An attempt was made in the present paper to give a qualitative explanation of the cluster (caustic) structure of the wave field in the transverse plane for a plane light wave propagating in a turbulent medium and to make quantitative estimates of the parameters of this structure. In the general case, the problem involves many parameters. However, if the problem is analyzed in a fixed plane, then for a plane wave with a constant wave parameter the solution of the problem is described by one parameter—the variance of the intensity in the region of weak fluctuations. In the present paper two extreme asymptotic cases corresponding to weak and saturated intensity fluctuations were analyzed. It should be noted that the limits of applicability of the asymptotic formulas probably depend on the intensity level I. It is natural to think that these limits are wider for low values of the level. However, this question requires additional investigation.

To analyze the intermediate case corresponding to a region where there is a developed caustic structure, which is of greatest interest for applications, it is necessary to know the probability density of the intensity and its transverse gradient for arbitrary distances traveled by the wave. Such an analysis can be performed either by using approximating expressions for the probability density for all values of the parameters[19] or on the basis of numerical modeling, as done, for example, in Refs. 21 and 22. We also note that in the geometric-optics approximation the equations (3) and (3′) lead to a closed equation for the joint probability density of the intensity of the wave field and the second derivatives of the phase of the wave with respect to the transverse coordinates characterizing the curvature of the phase front. This equation has not been studied at all.[2]

This work was supported in part by the Russian Fund for Fundamental Research under Projects Nos. 95-05-14247, 96-05-65347, and 96-05-65354.

[1] S. M. Rytov, Yu. A. Kravtsov, and V. I. Tatarskiĭ, *Principles of Statistical Radiophysics. Random Fields*, Springer, New York, 1987.
[2] V. I. Klyatskin, *Stochastic Equations and Waves in Randomly Inhomogeneous Media* [in Russian], Nauka, Moscow, 1980.
[3] J. Ziman, *Models of Disorder: The Theoretical Physics of Homogeneously Disordered Systems*, Cambridge University Press, N. Y., 1979 [Russian translation, Mir, Moscow, 1982, p. 146].
[4] S. D. Rice, Bell Syst. Tech. J. **23**, 282 (1944); **24**, 46 (1945).
[5] M. S. Longuet-Higgins, Philos. Trans. R. Soc. London, Ser. A **249**, 321 (1957); **250**, 157 (1957).
[6] P. Swerling, IRE Trans. Inf. Theory **8**, 315 (1962).
[7] M. B. Isichenko, Rev. Mod. Phys. **64**, 961 (1992).
[8] V. I. Klyatskin, W. A. Woyczynski, and D. Gurarie, J. Stat. Phys. **84**, 797 (1996).
[9] V. I. Klyatskin and A. I. Saichev, Zh. Éksp. Teor. Fiz. **111**, 1297 (1997) [JETP , 716 (1997)].
[10] F. V. Bunkin and K. S. Gochelashvili, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **11**, 1864 (1968).
[11] F. V. Bunkin and K. S. Gochelashvili, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **12**, 875 (1969).
[12] K. S. Gochelashvili and V. I. Shishov, *Waves in Randomly Inhomogeneous Media* [in Russian], VINITI, Soviet Academy of Sciences, Moscow, 1981.
[13] V. A. Kulkarny and B. S. White, Phys. Fluids **25**, 1770 (1982).
[14] B. S. White, SIAM (Soc. Ind. Appl. Math.) J. Appl. Math. **44**, 127 (1984).
[15] D. I. Zwillinder and B. S. White, Wave Motion **7**, 207 (1985).
[16] V. I. Klyatskin, Waves Random Media **3**, 93 (1993).
[17] V. I. Klyatskin and A. I. Saichev, Usp. Fiz. Nauk **162**, 161 (1992) [Sov. Phys. Usp. **35**, 231 (1992)].
[18] R. Dashen, Opt. Lett. **9**, 110 (1984).
[19] J. H. Churnside and S. F. Clifford, J. Opt. Soc. Am. A **4**, 1923 (1987).
[20] I. G. Yakushkin and V. U. Zavorotny, Waves Random Media **2**, 165 (1992).
[21] S. M. Flatte, G.-Y. Wang, and J. Martin, J. Opt. Soc. Am. A **10**, 2363 (1993).
[22] S. M. Flatte, C. Bracher, and G.-Y. Wang, J. Opt. Soc. Am. A **11**, 2080 (1994).

Translated by M. E. Alferieff

Interaction of light with a dye-doped nematic liquid crystal

M. I. Barnik, A. S. Zolot'ko, and V. F. Kitaeva

P. N. Lebedev Physics Institute, Russian Academy of Sciences, 117924 Moscow, Russia

(Submitted 5 December 1996)
Zh. Éksp. Teor. Fiz. **111**, 2059–2073 (June 1997)

The interaction of linearly and circularly polarized beams with a nematic liquid crystal doped with light-absorbing dyes has been studied by light-diffraction and microprojection methods. It has been found that there is a threshold for the emergence of light-induced anisotropic structures, which, depending on the type of dye, can be axisymmetric or extended in a direction determined by the light field. Possible mechanisms leading to the formation of the anisotropic structures are discussed. *© 1997 American Institute of Physics.* [S1063-7761(97)01006-8]

1. INTRODUCTION

As is well known, even small traces of light-absorbing molecules can alter the properties of liquid crystals and the character of their interaction with light. For example, small (less than 1%) additions of anthraquinone and azoic dyes increase the efficiency of light-induced reorientation of the director of nematic liquid crystals (NLCs) by at least an order of magnitude, and can change the direction of the director.[1–8] Adding methyl red makes light-induced reorientation of even smectic liquid crystal possible.[9]

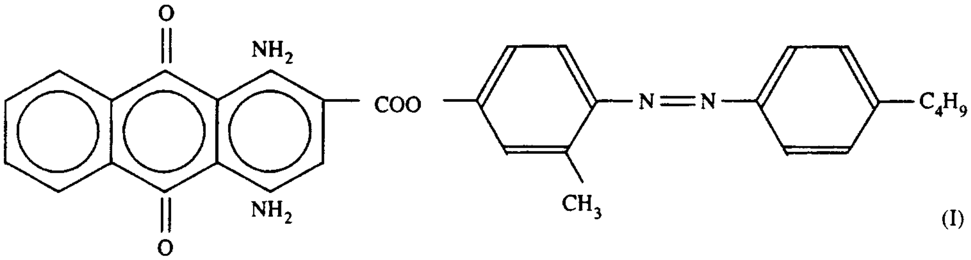
Light-induced changes in the structure of dye-doped liquid crystals also influence the light beam itself, altering its divergence and structure.

A specific property of liquid crystals — the collective nature of the response to external perturbations — is observed in the interaction of light with oriented liquid crystals. The collective nature of the response makes it possible to observe in liquid crystals a number of interesting effects due to the absorption of light, specifically, periodic distortions of the director field (square lattices) in cholesteric liquid crystals[10] and some features of the smectic–nematic phase transition.[11]

Our objective in the present work is to study the photoresponse of absorbing homeotropically-oriented multicomponent nematic liquid crystals.

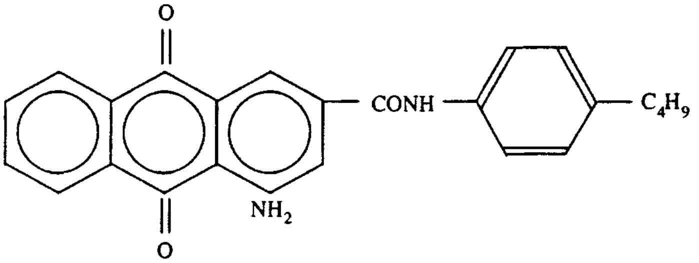
2. EXPERIMENTAL CONDITIONS AND RESULTS

The investigations were conducted with two samples of a homeotropically oriented NLC ZhKM-1277 (produced by the Main Science Center of the Russian Federation, Research Institute of Organic Semiproducts and Dyes), consisting of a five-component mixture of nematics based on biphenyls and esters. The samples contained small (1 wt.%) additions of dyes I (sample No. 1) and II (sample No. 2). The substance ZhKM-1277 possesses a nematic phase in the temperature range $-20\,°\mathrm{C} < T < 60\,°\mathrm{C}$. It is characterized by a positive low-frequency dielectric anisotropy ($\Delta\epsilon = 12.2$). The dyes



(I)

and



(II)

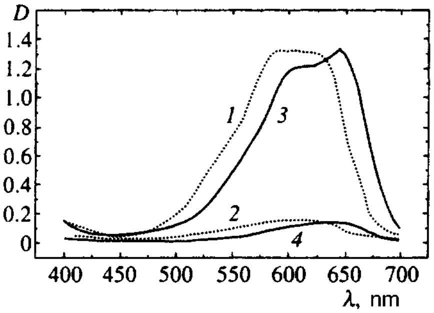


FIG. 1. Optical density D of the dyes I (curves *3* and *4*) and II (*1, 2*) versus the wavelength of the light (for $L=22$ μm and dye concentration $\sim 1\%$): *1, 3* — $D_{\parallel}$, *2, 4* — $D_{\perp}$.

possess positive dichroism. Both dyes contain anthraquinone fragments. The first dye also contains the azobridge N═N. The degree of ordering of the dye molecules in the nematic matrix is approximately 0.7. The polarization absorption spectra of the dyes are presented in Fig. 1. The absorption bands of both dyes belong to anthraquinone fragments.

The inner glass walls of the wedge-shaped cell were coated with SnO_2. This made it possible to apply an alternating electric field to the crystal. A homeotropic orientation of the liquid crystal was achieved by coating the inner surfaces of the walls with chromium stearyl chloride. Most of the investigations were carried out with crystal thickness $L=50$ μm, temperature $T=20\,^{\circ}$C, and wavelength of light $\lambda=6471$ Å.

The crystal was illuminated with light (focused in the crystal) from a continuous-wave argon–krypton laser. The linearly or circularly polarized light was normally incident on the cell, i.e., the wave vector of the incident radiation was parallel to the director. A thin-film polarizer was positioned in the path of the light behind the crystal, followed by a screen on which changes in the light beam transmitted through the crystal were observed. Changes in the crystal itself could be followed on the same screen with a microscope system (approximately $400\times$ magnification).

2.1. Changes in a linearly polarized light beam interacting with a crystal

The photographs in Figs. 2 and 3 illustrate the changes in the illuminated liquid crystal.

1. The divergence and structure of the light beam transmitted through the crystal change. The magnitude of the effect depends on the power of the light beam. The effect has a threshold.

For $P>P_{th}$ ($P_{th}\approx 9$ mW for sample No. 1 and

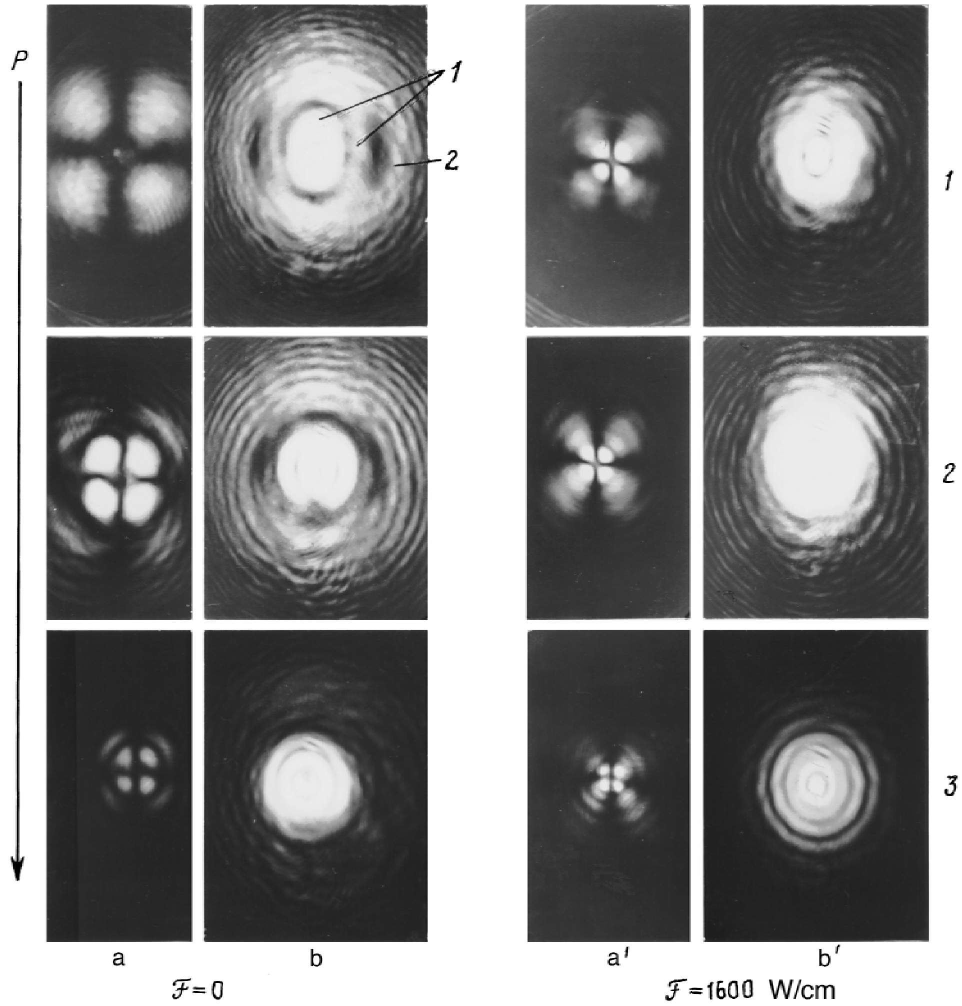


FIG. 2. Variations in the diffraction pattern with increasing radiation power P for sample No. 1: a, a′ — crossed polarizers, b, b′ — no analyzer: *1* — $P=11$ mW, *2* — $P=13$ mW, *3* — $P=17$ mW.

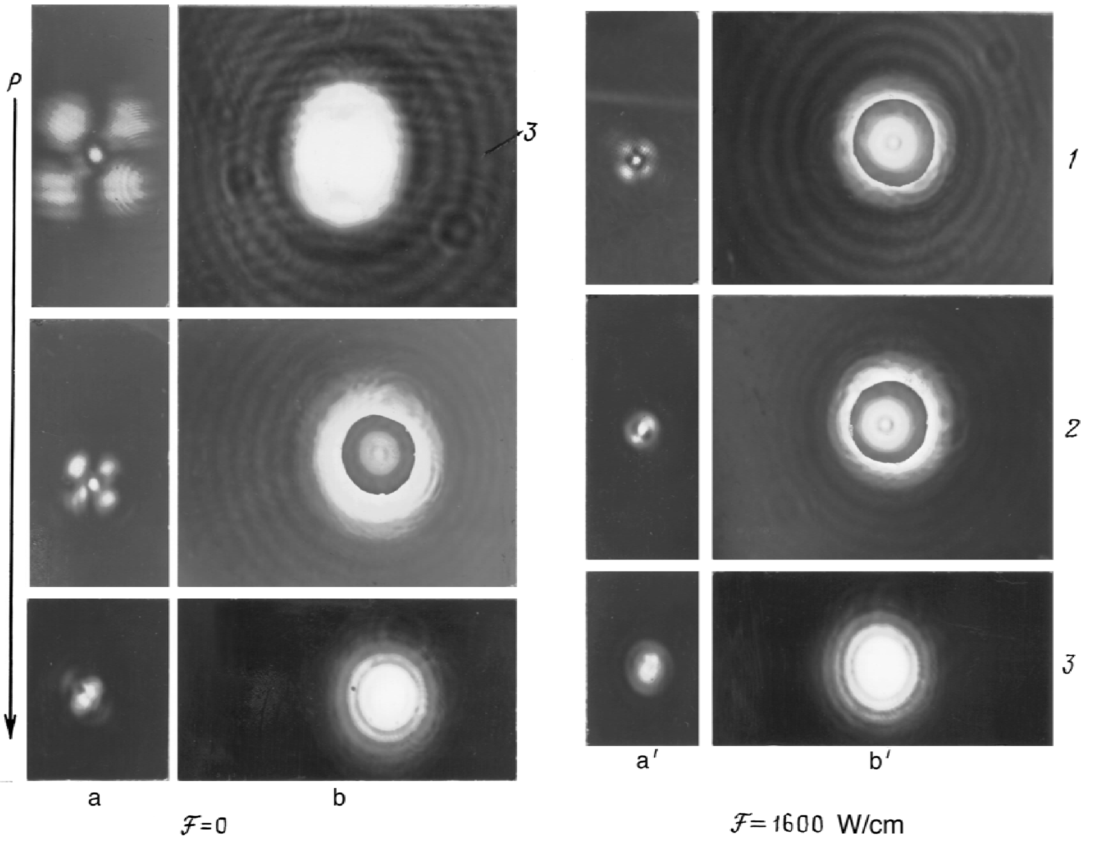


FIG. 3. Changes in the diffraction pattern with increasing radiation power P for sample No. 2: a, a′ — crossed polarizers, b, b′ — no analyzer; *1* — $P=13$ mW, *2* — $P=15$ mW, *3* — $P=25$ mW.

$P_{th}\approx 10$ W for sample No. 2), a cross appears on the screen in crossed polarizers. Without the analyzer, we observe a complicated ring pattern whose form depends strongly on the optical power and the type of dye.

2. Three ring systems can be identified for sample No. 1 close to threshold: Two central systems (*1*, *2* in Fig. 2b) exhibit low angular divergence. They are imposed on a third system (*3* in Fig. 3b), whose divergence is very large.

The first system consists of two highly prolate ovals. The ovals are extended perpendicular to the electric field (polarization) of the light. The cross observed in crossed polarizers is similary extended. The degree of prolateness (ratio of the size perpendicular to the light field to the size parallel to the field) is ≈ 1.7 (for the cross ≈ 1.2) for $P=11$ mW.

The second system consists of two very bright half-rings superposed on the outer low-intensity oval of the first system. The diameter connecting the half-ring is oriented along the electric field of the light wave.

The third system consists of a large number of equidistant rings (they are clearly seen in Fig. 3). The higher the optical power, the closer the spacing of the rings. The lower the power of the beam, the more clearly visible the rings.

The divergence of the central part is greater for sample No. 1 than for sample No. 2. Moreover, for sample No. 2 the central rings are essentially circular, and for small values of P the first and second systems cannot be distinguished. The third system for sample No. 2 shows up much more clearly.

3. As P increases, the degree of prolateness of the ovals and the angular divergence decrease. For $P=17$ mW the first two systems for sample No. 1 transform into a single system of completely circular but not very sharp rings. For crystal No. 2 and $P\approx 25$ mW the rings are clear and circular with sharp edges.

4. The action of an external low-frequency electric field $\mathcal{F}$, which should stabilize the initial homeotropic orientation of the crystal, is in a certain sense equivalent to an increase in optical power — in the field $\mathcal{F}$ the elongation of the diffraction pattern vanishes and circular rings with very sharp edges are also observed for crystal No. 1 with $P\approx 17$ mW. For sample No. 2 with $P\approx 25$ mW the diffraction pattern in the field $\mathcal{F}$ is identical to the diffraction pattern without the field.

5. The cross observed in crossed polarizers decreases in size with increasing P for crystal No. 1, and its intensity first increases and then decreases (see Fig. 2). For sample No. 2 the cross also decreases in size with increasing P, but at $P\geqslant 25$ mW, it transforms into a spot of indeterminate shape.

6. An external electric field substantially reduces the size of the cross. For crystal No. 1 two crosses are seen — a small bright cross against the background of a cross consisting of much larger arcs. For sample No. 2 The cross can still be seen at low power, but at high power it turns into a spot (see Fig. 3).

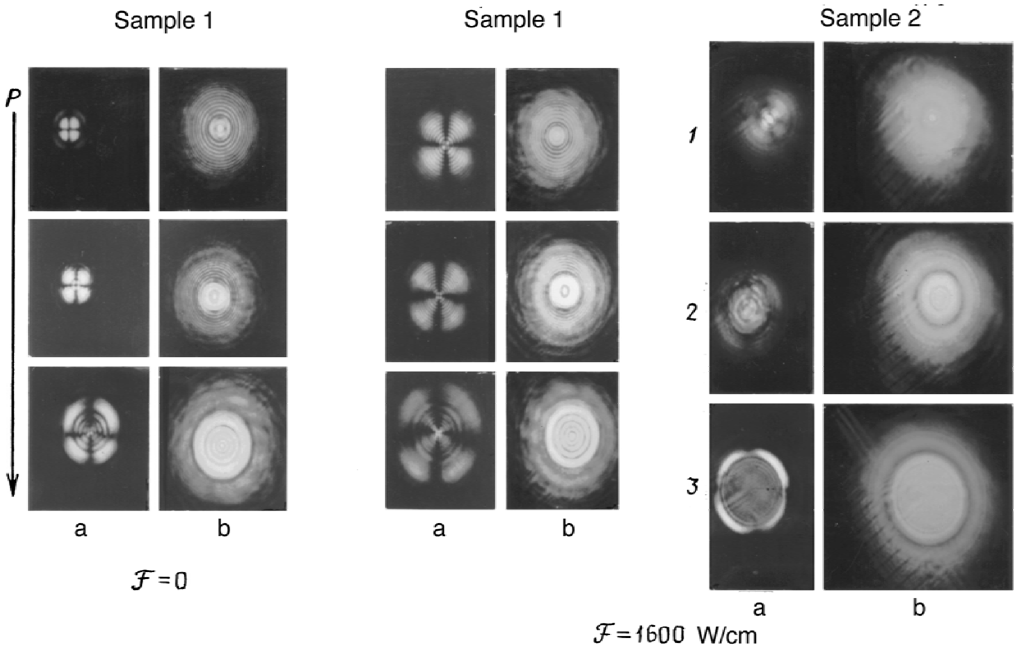


FIG. 4. Changes in the pattern observed in a microscope with increasing P for samples Nos. 1 and 2: a — crossed polarizers, b — no analyzer; *1* — $P=11$ mW, *2* — $P=13$ mW, *3* — $P=17$ mW.

2.2. Changes in a crystal illuminated by a linearly polarized light beam

These changes, observed with the aid of a microscope system, are illustrated in Fig. 4.

The following was determined for sample No. 1:

1. An anisotropic pattern, whose size increases with optical power appears in the zone of the light beam with $P>P_{th}$. Without the analyzer, the anisotropic pattern has the form of a disk. The disk is surrounded by an aureole in the form of a ring, which at low powers P has fine structure. A cross is observed in crossed polarizers. The structure of the cross at high power P becomes complicated: fine structure (arcs) appears at the center, and the intensity drops sharply.

2. An electric field has essentially no effect on the size of the disk, and the aureole turns into a ring with sharp edges.

In crossed polarizers the cross changes substantially: it increases in size, fine structure appears, and as P increases, its intensity decreases appreciably, especially at the center.

The following was determined for sample No. 2:

1. Patterns similar to those for crystal No. 2 are observed without the analyzer and in crossed polarizers (for $\mathcal{F}=0$).

2. Whereas the patterns are similar in parallel polarizers for samples Nos. 1 and 2 in an electric field, they differ substantially in crossed polarizers. For sample No. 2 a complicated intensity distribution, not at all similar to a cross, is observed at moderate optical power. At high, P, a ‘‘thin rim’’ from the cross remains.

The behavior of the diffraction and microscope patterns near threshold should be especially noted. In this case they vanish after several seconds, i.e., the light-induced distortions of the director field ‘‘dissipate’’ quite rapidly.

For P slightly above P_{th}, the distortions of the director field are oscillatory: they appear and disappear, and there is even an average period of the oscillations $\approx 5-20$ s.

One can talk about a stable distortion of the director field in the zone of the light beam only for $P \geqslant P_{stab} \geqslant P_{th}$.

The value of P_{th} depends on the thickness of the crystal. As the thickness of the crystal increases, P_{th} decreases. The dependence $P_{th}(L)$ for sample No. 1 for $\lambda=6471$ Å is shown in Fig. 5.

The relaxation time of the distortions after the light is turned off is 1–2 s.

2.3. Changes in a circularly polarized light beam interacting with a crystal

The changes are illustrated for sample No. 1 in Fig. 6.

1. The crystal changes the divergence and structure of a circularly polarized light beam. This is a threshold effect, as in the case of linear polarization. P_{th} is close to the value of P_{th} for linearly polarized light.

2. For $P \simeq P_{th}$ a vortex-like pattern is observed on the screen for light that has traversed the crystal and the analyzer — a nucleus with two ‘‘tails’’ curling clockwise or counter-

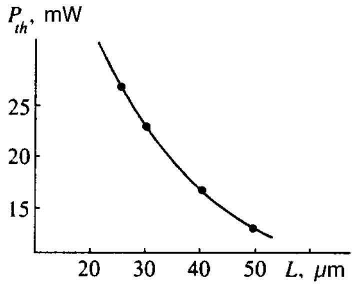


FIG. 5. P_{th} versus crystal thickness L for sample No. 1 ($\lambda=6471$ Å).

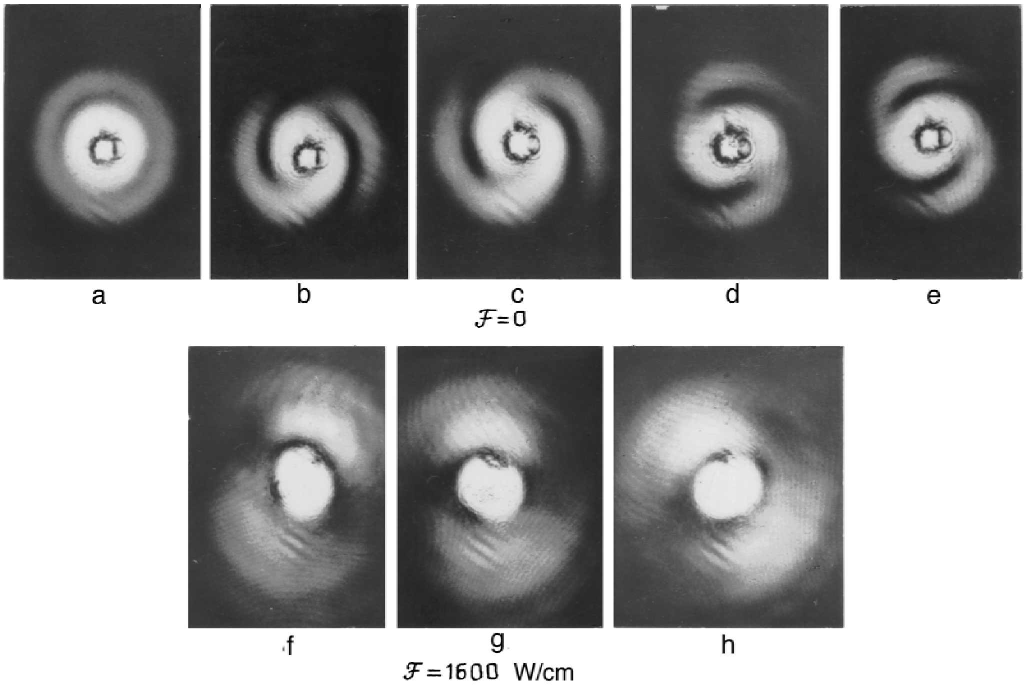


FIG. 6. Changes in the diffraction patterns for circular polarization accompanying a change in analyzer rotation angle α: a — no analyzer, b — $\alpha=0$ °C (the analyzer passes horizontal polarization), c — 20 °C, d — 77 °C, e — 90 °C (the analyzer passes vertical polarization), f — 20 °C, g — 55 °C, h — 90 °C.

clockwise, depending on the direction of circular polarization.

3. As for linear polarization, the pattern observed near threshold is oscillatory. A vortex forms over a time 1–15 s, and relaxes over a time. 0.5–5 s.

4. A stable pattern is easily achieved by increasing P. A sharp nucleus and long bright tails are observed.

5. As P increases further, the pattern remains stable but the vortex turns into a system of rings around a bright spot. The first ring is quite narrow and very intense, and the other three to five rings are wide and have low intensity.

6. A vortex and wide rings are observed against the background of a system of narrow equidistant rings whose angular dimensions are an order of magnitude greater than those of the vortex.

2.4. Changes in a crystal illuminated by a circularly polarized light beam

The changes are illustrated in Figs. 7 and 8.

1. For $P>P_{th}$, anisotropic patterns appear in the zone of the light beam. As in the case of linearly polarized radiation, without an analyzer these patterns have the form of a disk (whose size depends on P) surrounded by an aureole.

2. When an analyzer is present, the disk is ''sliced'' in half. The halves and their centers are displaced relative to one another at an angle of 45° to the direction of the analyzer. Very close to threshold ($P\cong P_{th}$), one or two half-rings, which are not displaced relative to one another, are seen instead of half disks.

3. Relatively close to threshold (at P_1), quasiperiodic oscillations of the disk size are observed. As follows from Fig. 7 (for sample No. 1), when an analyzer is present, the pattern with the half-disks is replaced by a pattern with half-rings and vice versa. For $P_2>P_1$, no changes in the pattern were observed over a period of 25 min.

4. An external low-frequency electric field changes the

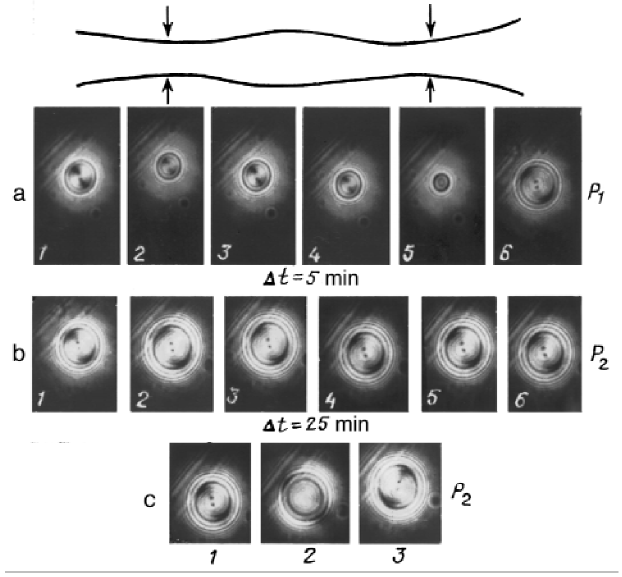


FIG. 7. Changes with time in the microscope pattern for P_1 (a) and $P_2>P_1$ ($\alpha=90°$) (b) as well as in an external electric field $\mathscr{F}$ (c): *1, 3* — for $\mathscr{F}=0$ for right- and left-hand circular polarization of the light, *2* — $\mathscr{F}\neq 0$.

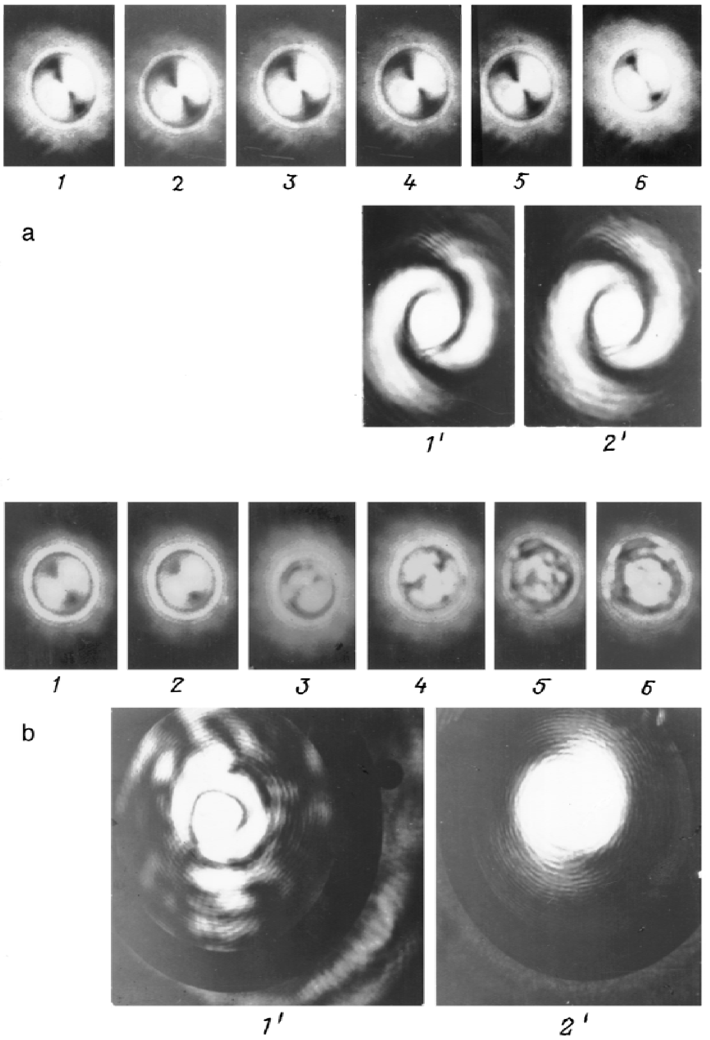


FIG. 8. Changes in microscope pcitures and diffraction patterns under the action of an external electric field ($\mathscr{F}=1600$ V/cm) for different types of dyes: a — sample No. 1, $P=15$ mW; b — sample No. 2, $P=16$ mW. Microscope patterns: *1* — $\mathscr{F}=0$, *2–6* — with continuous illumination of the crystal and short-time actions of a field with different durations. Diffraction patterns: *1′* — after the action of a field, *2′* — before the action of a field.

pattern observed with the analyzer — the half-disks merge into a single uniform disk. In this case, however, the aureole surrounding the disk turn into wide half-rings, and the diameter connecting them is perpendicular to the direction defined by the quarter-wave plate.

5. Repeated application of an external high-frequency electric field to the crystal and rotation of the quarter-wave plate do not produce any appreciable changes in the pattern seen in a microscope for sample No. 1 (Fig. 7c).

6. A different pattern is observed for sample No. 2. Repeated application of a high-frequency electric field produces distortions of the director field that persist for a long time (memory). As follows from Fig. 8, these distortions appear in the diffraction pattern as well.

3. DISCUSSION

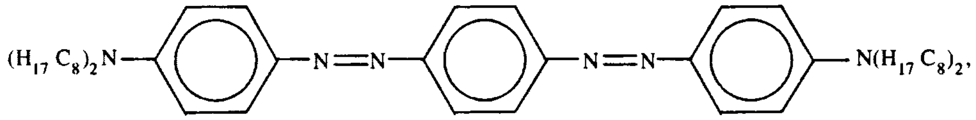
1. The changes in the divergence, profile, and polarization of a light beam that has traversed an NLC and the results of investigations of texture under a microscope attest to the appearance of light-induced optical nonuniformity in a crystal doped with dyes.

2. The light-induced nonuniformity is characterized by transverse (with respect to the direction of light propagation) anisotropy. Indeed, depolarization of the light (appearance of crosses in crossed polarizers) can give rise to either transverse anisotropy (the direction of the optical axis changes in the zone of the light beam) or nonlinear refraction of the beams, forming a conoscopic pattern.[12] However, the angular divergence of the beam for a conoscopic pattern should be many times greater than the observed divergence, and the appearance of a cross in crossed polarizers must be ascribed to transverse anisotropy.

3. The axisymmetric (sample No. 2) or nearly so (sample No. 1) character of the light-induced nonuniformity obviously directly reflects the axial symmetry of a Gaussian light beam.

The observed deviation from axial symmetry for sample No. 1 can be explained by the special properties of compounds containing an azobridge N═N. The molecules of such compounds tend to orient themselves so that their long axis is perpendicular to the electric field of the light wave.[13–16] This will naturally affect the symmetry of the light-induced anisotropy of the pattern.

To confirm the validity of this conjecture we investigated the interaction of linearly polarized light with a ZhKM-1277 crystal doped with the diazodye



the molecules of which have two azobridges. The degree of elongation of the cross has increased appreciably (for sample No. 1 it became 1.6 instead of 1.2).

4. The diffraction pattern enables one to estimate the size of the spatial optical nonuniformity. Examination under a microscope makes it possible to determine the sizes directly. The estimates agree quite well with one another. For sample No. 1 the nonuniformities are 30–60 μm in size for optical power $P \simeq 9-13$ mW, and for sample No. 2 the nonuniformities are 20–70 μm in size for $P \simeq 13-18$ mW.

The distinctness of the diffraction pattern 3 (Fig. 3b) and the large number of diffraction orders suggest that the light-induced nonuniformity has a sharp boundary.

5. The action of an external electric field on the light-induced structure suggests that it forms as a result of the reorientation of the director in the zone of the light beam. In this case, since the NLC ZhKM-1277 possesses a positive dielectric anisotropy, an external field will tend to reestablish the initial homeotropic orientation of the director. Naturally, the light-induced pattern will change in this case, as is observed in both the diffraction pattern (Figs. 2 and 3) and the microscope pattern (Fig. 4).

6. There are several possible physical mechanisms of light-induced reorientation of the director.

In the nematic phase, the director can be reoriented by electric fields[17,18] (including optical fields[19–21]), hydrodynamic flows,[17,18] a nonuniform temperature field (thermomechanical effect, predicted in Ref. 22), and surface forces acting on the nematic–isotropic interphase boundary.[23]

We now analyze the possibility that these mechanisms are at work in our experiment.

The electric field of a linearly polarized light wave normally incident on a homeotropically oriented crystal reorients the director in the polarization plane of the wave. Therefore neither reorientation of the director by the electric field of the light wave[19–21] nor orientational effects due to the absorption of light[1–8] should lead to the formation of a nonplanar structure—much less, one that is axisymmetric or almost so.

The hydrodynamic flows that can be induced by laser heating in a crystal can obviously not be symmetric under the conditions of our experiment, in which the force of gravity is perpendicular to the axis of the light beam. There remain the reorientation of the director at the boundary of the isotropic hole that a light beam can burn into the crystal (thermal heating),[24] and the reorientation of the director by a nonuniform temperature field (thermomechanical effect) produced by a narrow light beam in the crystal.

We now consider the possibility that these effects are manifested in the experiment described.

7. Let us estimate the power required to burn an isotropic hole into the crystal. According to Ref. 25, the power P needed to raise the temperature of a homeotropic crystal by ΔT is

$$P=\frac{2\pi k_{\perp} w \Delta T}{[1-\exp(-\alpha_{\perp} L)] f(w/L)}, \tag{1}$$

where $\alpha_{\perp}$ is the absorption coefficient of the ordinary light wave, $k_{\perp}$ is the thermal conductivity, w is the radius of the beam waist at $1/e$ times the maximum intensity, and f is a geometric factor that depends on the ratio w/L. Under the conditions of our experiment, $w \approx 25$ μm, $L=50$ μm, and $f=0.23$; $\alpha_{\perp}$ is determined from the optical density D (Fig. 1): $\alpha_{\perp}=D_{\perp}\ln(10/L)$. The thermal conductivity $k_{\perp}$ of the nematic phase is typically $k_{\perp} \sim 10^4$ ergs/cm·s·deg.

It follows from Eq. (1) that optical power $P \sim 4$ mW is sufficient to heat the crystal to the transition point into the isotropic phase (i.e., by 40 °C). This agrees in order of magnitude with the experimental value $P_{th} \sim 10$ W.

8. Consider now the thermomechanical effect. According to Ref. 22, the force that reorients the director in a nonuniform thermal field (the magnitude of the force is proportional to the temperature gradient) can appear only in a nonuniform director field. For a homeotropic orientation of the crystal, the director field is uniform, but the reorienting force can result from a noise-induced deviation of the director (as happens in a Frederiks transition in an electric field oriented perpendicular to the director). The magnitude of this force is proportional to the temperature gradient. Deviations of the director are opposed by the elastic force (whose magnitude depends on the Frank elastic constants), which tends to restore the initially uniform orientation. The ratio of these forces determines the state of the director field, but does it remain uniform or will it be deformed?

According to Ref. 1, the threshold ΔT_{th} of the thermomechanical effect in a narrow light beam can be obtained from

$$\xi \sim K/\Delta T_{th}, \tag{2}$$

where K is the Frank elastic constant and ξ is the thermomechanical constant. In this case $K \sim 10^{-6}$ dynes,[18] $\xi \sim 10^{-6}$ dynes/deg (Ref. 22) or, according to a different estimate, $\xi \sim 10^{-8}$ dynes/deg.[1] For these values of ξ we find $\Delta T_{th} \sim 1$ °C and $\Delta T_{th} \sim 100$ °C, respectively. It follows from Eq. (1) that a light beam with power $P \sim 10$ mW is sufficient to heat the crystal by $\Delta T = 100$ °C and 0.1 mW is sufficient to heat the crystal by 1 °C. Therefore, a thermomechanical effect is entirely possible under our experimental conditions.

9. The threshold power for heating the crystal to the nematic–isotropic phase transition point and the threshold for the thermomechanical effect should increase with decreasing crystal thickness (see Eq. (1)). This is observed experimentally (Fig. 5). But this result does not make it possible to decide between thermal heating and the thermomechanical effect. We now look at some other facts.

a) According to Ref. 25, hysteresis in power should occur during thermal heating, during which the nematic–isotropic phase transition arises and vanishes. This was not observed experimentally, suggesting a thermomechanical effect.

b) The fact that the threshold for the appearance of light-induced structure is independent of the external low-frequency field suggests a thermal effect.

c) Increasing the optical power should substantially increase the radius of the isotropic ''hole.''

A substantial increase in the magnitude of the light-induced nonuniformity resulting from the thermomechanical effect should not occur, however, since for a Gaussian beam the distance from the beam axis at which the temperature gradient is maximum cannot change much with increasing power. A substantial increase in nonuniformity was observed experimentally, which is consistent with thermal heating.

d) The characteristic formation times of light-induced nonuniformity (t_{for} is of the order of fractions of a second) and its relaxation ($t_{rel}\sim 1-2$ s) obtained for samples Nos. 1 and 2 also do not answer the question at hand. Additional experiments with high-temperature NLC ZhKM-1443 containing 1% dye I (its transition point to the isotropic phase ≈90 °C, thickness $L=40$ μm) showed that for P slightly greater than P_{th} the appearance of a cross on the screen is determined by two times — t_{ret} and t_{for} (t_{ret} is the time at which illumination of the crystal starts prior to the appearance of the cross). In this case t_{ret} and t_{for} are close to one another and equal several tens of seconds.

The long times t_{ret} can be equally responsible for the thermomechanical effect and a phase transition accompanying thermal heating.

Long times t_{for} (~ 10 s), however, can be expected only in the case of the thermomechanical effect, where, as in a Frederiks transition, for a small distance above threshold[26] a slow collective rotation of the initially identically oriented molecules occurs.

e) Oscillatory regimes accompanying a nematic–isotropic phase transition were observed in Refs. 25, 27, and 28. Diffraction losses of light[27,28] and diffusion of the components of the NLC[25] were mentioned as possible reasons for the appearance of these regimes.

The appearance of oscillatory regimes accompanying thermomechanical reorientation of the director also cannot be ruled out. These regimes can arise both because of the influence of reorientation of the director and excitation of the dye molecules on absorption and diffusion of heat, the orienting effect of the light field.

All this shows that the experimental results do not make it possible to choose unequivocally between thermal heating and the thermomechanical effect, but they do suggest that the two processes simultaneously play a role in the formation of anisotropic structure.

10. In closing, we examine the interaction of circularly polarized light with a NLC.

It is obvious that light-induced reorientation of the director should not depend on whether the polarization is linear or circular, but the diffraction patterns and the microscope patterns should be different.

The computational results for the diffraction of a circularly polarized light beam by an axisymmetric untwisted (not changing under reflection in a plane perpendicular to the symmetry axis) anisotropic structure[11] show that the form of the diffraction pattern is determined by the radial dependence of the refractive indices of the extraordinary and ordinary waves. The diffraction pattern, generally speaking, may appear to be twisted, but the sign of the twisting depends on the sign of the circular polarization of the incident light.

However, an orienting effect of a circularly polarized light field on the director is entirely possible. We shall explain this. As light propagates through a deformed crystal, its polarization changes. In the geometry of the described experiment, circular polarization turns into elliptical polarization. Elliptically polarized light can reorient the director in a direction parallel or perpendicular to the major axis of the polarization ellipse. The accompanying twisting of the director field will also be manifested in the diffraction pattern. The direction of the major axis of the polarization ellipse of the light obviously depends on the sign of the circular polarization, so that the direction of reorientation of the director and, correspondingly, the twisting also depend on this sign.

A calculation of the microscope pattern formed by an untwisted director field, performed under the assumption that the rays comprising the light beam are not deflected inside the crystal, shows that this picture should be symmetric relative to the direction corresponding to the minimum of intensity and should make an angle of 45° with the direction of the analyzer. However, as follows from Fig. 7, the half-disks can be shifted relative to one another, indicating that the initial symmetry of the light-induced structure is destroyed. Therefore the form of the microscope pattern confirms the possibility that the director field can be twisted by the light field.

It is well known that the efficiency of the reorientation of the director in absorbing liquid crystals depends on the type of dye. This could explain the difference observed in the patterns for samples Nos. 1 and 2.

4. CONCLUSIONS

1. Nonplanar reorientation of a director occurs under the action of a light beam in NLCs doped with light-absorbing dyes. This is a threshold effect. For optical power close to threshold, the reorientation of the director can be oscillatory.

2. The deformation of the director field depends on the type of dye and the polarization of the light beam. For a dye that does not contain an N=N asobridge it is axisymmetric for linear polarization and for dyes with an azobridge it extends in the direction of the light field; for circular polarization the director field becomes twisted.

3. The reorientation of the director in the light-beam zone and many of its properties are explained by the action of a temperature gradient (thermomechanical effect) and the nematic–isotropic phase interface accompanying a thermal phase transition.

We thank V. G. Rumyantsev for helpful discussions of this work.

This work was performed as part of the Interdepartmental Scientific-Technical Program ''Optics. Laser Physics.'' M. I. B. is grateful to INTAS for support under Grant No. 93-1700-EXT.

[1] I. Janossy, A. D. Lloyd, and B. S. Wherett, Mol. Cryst. Liq. Cryst. **179**, 1 (1990).

[2] I. Janossy and T. Kosa, Opt. Lett. **17**, 1183 (1992).

[3] S. T. Sun, W. M. Gibbons, and P. J. Shannon, Liq. Cryst. **12**, 869 (1992).

[4] I. C. Khoo, H. Li, and Y. Liang, IEEE J. Quantum Electron. **QE-29**, 1444 (1993).

[5] L. M. Blinov, M. I. Barnik, A. Mazzula, and C. Umeton, Mol. Mater. **5**, 237 (1995).

[6] D. Paparo, P. Maddalena, G. Abbate, E. Santamato, and I. Janossy, Mol. Cryst. Liq. Cryst. **251**, 73 (1994).

[7] M. I. Barnik, A. S. Zolot'ko, V. G. Rumyantsev, and D. B. Terskov, Kristallografiya **40**, 746 (1995).

[8] T. Ya. Marusii, V. Ya. Reshetnyak, Yu. A. Reznikov, and S. S. Slussarenko, in *Abstracts of the 6th International Topcal Meeting on Optics of Liquid Crystals,* Le Touget, France (1995).

[9] A. S. Zolot'ko and V. F. Kitaeva, JETP Lett. **59**, 33 (1994).

[10] V. F. Kitaeva and A. S. Zolot'ko, Mol. Cryst. Liq. Cryst. Sci. Technol. B: Nonlinear Opt. **2**, 261 (1992).

[11] A. S. Zolot'ko, V. F. Kitaeva, and D. B. Terskov, Zh. Éksp. Teor. Fiz. **101**, 1827 (1992) [Sov. Phys. JETP **74**, 974 (1992)].

[12] M. Born and E. Wolf, *Principles of Optics*, 4th ed., Pergamon Press, New York (1969).

[13] A. M. Makushenko, B. S. Neporent, and O. V. Stobova, Opt. Spektrosk. **31**, 557 (1971) [Opt. Spectrosc. **31**, 295 (1971)].

[14] T. Todorov, L. Nikolova, N. Tomova, and V. Dragostinova, IEEE J. Quantum Electron. **QE-22**, 1262 (1986).

[15] S. Ivanov, I. Yakovlev, S. Kostromin, and V. Shibaev, Macromol. Chem. Rep. Comm. **12**, 709 (1991).

[16] K. Anderle, R. Birenheide, M. J. A. Werner, and J. H. Wendorff, Liq. Cryst. **9**, 691 (1991).

[17] P. de Gennes, *The Physics of Liquid Crystals*, Clarendon Press, Oxford (1974).

[18] L. M. Blinov, *Electro- and Magneto-Optics of Liquid Crystals* [in Russian], Nauka, Moscow (1978).

[19] I. C. Khoo and Y. R. Shen, Opt. Eng. (Bellingham) **24**, 579 (1985).

[20] A. V. Sukhov, N. V. Tabiryan, and B. Ya. Zel'dovich, Mol. Cryst. Liq. Cryst. **136**, 1 (1986).

[21] V. F. Kitaeva and A. S. Zolot'ko, Laser Res. **10**, 275 (1985).

[22] R. S. Akopyan and B. Ya. Zel'dovich, Zh. Éksp. Teor. Fiz. **87**, 1660 (1984) [Sov. Phys. JETP **60**, 953 (1984)].

[23] W. Bragg, Nature **133**, 445 (1934).

[24] V. F. Kitaeva, N. N. Sobolev, A. S. Zolot'ko, L. Csillag, and N. Kroo, Mol. Cryst. Liq. Cryst. **91**, 137 (1983).

[25] I. Janossy and T. Kosa, Mol. Cryst. Liq. Cryst. **207**, 189 (1991).

[26] A. S. Zolot'ko, V. F. Kitaeva, V. A. Kuyumchyan, N. N. Sobolev, and A. P. Sukhorukov, JETP Lett. **36**, 80 (1982).

[27] F. Simoni, G. Cipparone, C. Umeton, and I. C. Khoo, Opt. Lett. **13**, 886 (1988).

[28] V. G. Kamensky, G. Cipparone, C. Umeton, and F. Simoni, Mol. Cryst. Liq. Cryst. **207**, 331 (1991).

Translated by M. E. Alferieff

Characteristic features of the π-electron states of carbon nanotubes

S. S. Savinskiĭ and N. V. Khokhryakov

Udmurt State University, 426034 Izhevsk, Russia

(Submitted 14 February 1996)
Zh. Éksp. Teor. Fiz. **111**, 2074–2085 (June 1997)

A simple tight-binding model of the π-electron states of carbon nanotubes is analyzed. The symmetry group of nanotubes and its relation to electronic structure are discussed. The applicability of the simple model is analyzed on the basis of numerical calculations of the electronic states of energy-optimized nanotubes, performed in a parametric tight-binding model that takes account of the carbon valence electrons. Numerical data on the gap widths of optimized nanotubes and data obtained from the zone-folding model employed in the literature are presented. © *1997 American Institute of Physics.* [S1063-7761(97)01106-2]

1. INTRODUCTION

Carbon nanotubes—tubulenes—are molecular cylindrical surfaces which are close-packed with atomic carbon hexagons[1] and are obtained by the thermal decomposition of graphite. They can be represented geometrically as the result of gluing a strip cut from a single graphite plane. Depending on the method of gluing, a set of atomic carbon structures with a wide spectrum of conducting properties is obtained (see, for example, Refs. 2 and 3)—from semiconductors with gap widths of 0–2 eV to semimetals, of which graphite is a typical representative. The unique conducting and capillary properties of tubulenes make them promising materials for nanoelectronic devices. Tubulenes are also of interest theoretically as a new class of quasi-one-dimensional structures.

A characteristic feature of the energy structure of the valence electrons of a single graphite plane is the existence of π electrons, whose states can be described by a simple analytical model, at the Fermi level; σ electrons at the Fermi level have an energy gap of the order of 10 eV.[4] When an ideal tubulene is glued from a strip cut from a graphite plane, the perturbation of the π and σ electrons depends on the dimensionless parameter a/R, where a is the distance between the closest atoms in the graphite plane and R is the radius of the tubulene. For tubulenes with a large radius this perturbation is weak, and accordingly the electronic spectrum of a tubulene can be obtained from that of an isolated graphite plane; this approximation is widely employed in the literature and is termed zone folding (see, for example, Refs. 2, 3, and 5–7). In the present paper, we discuss the zone-folding method in detail and we propose an illustrative geometric analysis of the characteristic features of the π-electron spectrum. This analysis could be helpful in studying transport phenomena and the characteristic features of electron–phonon interactions in tubulenes. For tubulenes with a small radius, the finiteness of the perturbation, which can result in hybridization of the π and σ electrons, must be taken into account. We assess the applicability of the zone-folding method for π electrons based on a direct numerical optimization of the energy of a tubulene in a parametric tight-binding method that we employed previously for calculations of fullerenes[8] and the phonon spectra of tubulenes.[9]

In our parametric tight-binding method, the energy spectrum of the valence electrons is calculated by diagonalizing the matrix of the Hamiltonian operator. In so doing, the matrix elements are assumed to depend on the distances between the carbon atoms and the numbers of the quantum states of electrons in the chosen basis of s and p wave functions of the isolated atoms. The inner-shell electrons and many-body effects are taken into account via a parametric interatomic interaction potential. The parameters of the method considered here were fit in Ref. 10 for carbon atoms using data on the electronic structure and elastic moduli of diamond and graphite.

2. CHARACTERISTIC FEATURES OF THE π-ELECTRON STATES OF AN ISOLATED ATOMIC GRAPHITE PLANE

Consider an isolated, infinite graphite plane close-packed with regular atomic hexagons. In the tight-binding approximation, the wave functions of the π electrons have the form

$$|\pi,\mathbf{k},\gamma\rangle=\frac{1}{\sqrt{N}}\sum_{\mathbf{T}} e^{i\mathbf{kT}}|\pi,\mathbf{T},\tau_\gamma\rangle, \tag{1}$$

where $\gamma=0,1$ labels the atoms in a unit cell; the vector $\mathbf{T}=n_1\mathbf{a}_1+n_2\mathbf{a}_2$, where $\mathbf{a}_1$ and $\mathbf{a}_2$ are the elementary vectors of the Bravais cell and n_1 and n_2 are integers, determines the position of the unit cell; $\mathbf{k}$ is a quasiwave vector; $|\pi,\mathbf{T},\tau_\gamma\rangle$ is the atomic π-electron wave function in a $\mathbf{T}$ cell at the γth atom and consists of a p orbital oriented perpendicular to the graphic plane; $\tau_0=0$ and τ_1 are basis vectors (see Fig. 1); and, N is a normalization number. In what follows, we neglect the overlap of the wave functions of neighboring atoms

$$\langle\pi,\mathbf{T},\tau_\gamma|\tau_{\gamma'},\gamma',\pi\rangle=\delta_{\mathbf{TT}'}\delta_{\tau_\gamma\tau_{\gamma'}}.$$

where $\mathbf{T}'$ is an arbitrary lattice vector and δ is the Kronecker delta.

We form the linear combination

$$|\pi,\mathbf{k}\rangle=A|\pi,\mathbf{k},0\rangle+B|\pi,\mathbf{k},1\rangle \tag{2}$$

from the functions (1), where A and B are coefficients that we determine from the equations for the eigenfunctions and eigenvalues of the Hamiltonian operator of an electron

$$H|\pi,\mathbf{k}\rangle=E_{\pi,\mathbf{k}}|\pi,\mathbf{k}\rangle. \tag{3}$$

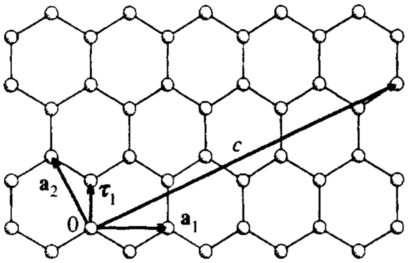


FIG. 1. Fragment of a graphite plane.

To solve Eq. (3) we set the diagonal matrix elements of the Hamiltonian operator equal to

$$\langle \pi,\mathbf{k},\gamma|H|\gamma,\mathbf{k},\pi\rangle=E_{\pi}^{0},$$

where E_{π}^{0} is the binding energy of a π electron with a free carbon atom. In calculating the matrix element

$$H_{01,\mathbf{k}}=\langle \pi,\mathbf{k},0|H|1,\mathbf{k},\pi\rangle=\sum_{\mathbf{T}}\langle \pi,\mathbf{T},\tau_0|H|\tau_1,0,\pi\rangle e^{-i\mathbf{kT}} \tag{4}$$

in the sum (4) we take account of only the overlap integrals $\langle \pi,\mathbf{T},\tau_0|H|\tau_1,0,\pi\rangle$ involving nearest neighbors (see Fig. 1). As a result, we obtained for the matrix element (4)

$$H_{01,\mathbf{k}}=\beta_0+e^{-i\mathbf{ka}_2}(\beta_1+\beta_2 e^{-i\mathbf{ka}_1}), \tag{5}$$

where $\beta_0=\langle \pi,0,\tau_0|H|\tau_1,0,\pi\rangle$, $\beta_1=\langle \pi,\mathbf{a}_2,\tau_0|H|\tau_1,0,\pi\rangle$, $\beta_2=\langle \pi,\mathbf{a}_1+\mathbf{a}_2,\tau_0|H|\tau_1,0,\pi\rangle$. The nearest-neighbor positions in an atomic graphite plane are equivalent, so that $\beta_0=\beta_1=\beta_2$.

Solving Eq. (3), taking account of Eqs. (2) and (5), we obtain two solutions for each fixed vector $\mathbf{k}$ in the graphite plane:

$$E_{\pi,\mathbf{k}}^{\pm}=E_{\pi}^{0}\pm|H_{01,\mathbf{k}}|,$$

$$|\pi,\mathbf{k}\rangle^{\pm}=(|\pi,\mathbf{k},0\rangle\pm e^{i\vartheta}|\pi,\mathbf{k},1\rangle)/\sqrt{2}, \tag{6}$$

where ϑ is the phase of the complex number $H_{01,\mathbf{k}}=e^{i\vartheta}|H_{01,\mathbf{k}}|$.

Figure 2 displays level curves of the function $E_{\pi,\mathbf{k}}^{+}$ in the first Brillouin zone. The function $E_{\pi,\mathbf{k}}^{+}$ clearly peaks at the central point Γ, the minimum is reached at the point K and equals zero, and M is a saddle point. We note that the surface $E_{\pi,\mathbf{k}}^{+}$ (see Fig. 2) can be approximated over most of the Brillouin zone by a paraboloid of revolution; near K in the extended band scheme, it is a conical surface.

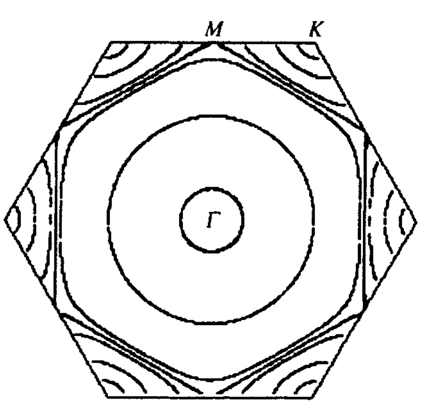


FIG. 2. Typical level curves of π-electron energies in the Brillouin zone.

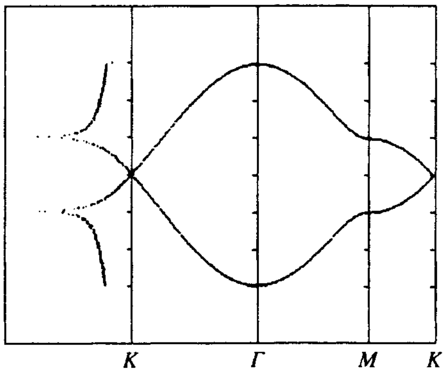


FIG. 3. Numerical calculation of the density of states (left-hand figure) and energy bands of π electrons. The unit on the energy scale is the modulus of the matrix element β_0.

Figure 3 displays the computational results for the bands and densities of the π-electron states of a graphite plane. The bands were calculated according to Eq. (6) for vectors $\mathbf{k}$ in the intervals ΓK, ΓM, and MK. Logarithmic van Hove singularities, associated with the saddle points M on the surfaces $E_{\pi,\mathbf{k}}^{+}$, are present in the density of states (see Fig. 3); a van Hove singularity associated with the discontinuity of the density of states is also present at the point Γ (a more detailed discussion of van Hove singularities in two-dimensional systems is given, for example, in Ref. 11). The Fermi energy of the π electrons is E_{π}^{0}; the electrons completely fill all states of the lower branch of the spectrum (6), and accordingly fill the entire Brillouin zone.

3. π-ELECTRON STATES ON A GRAPHITE STRIP AND AN IDEAL TUBULENE

Let us excise from the graphite plane a strip defined by the vector $\mathbf{c}$, $\mathbf{c}=i_1\mathbf{a}_1+i_2\mathbf{a}_2$ (see Fig. 1) and introduce in the strip the unit vector $\mathbf{n}_z$, directed perpendicular to the vector $\mathbf{c}$ and the unit vector $\mathbf{n}_x$ directed along the vector $\mathbf{c}$. We identify the points on one edge of the strip with the points on the other edge. This gives for the wave functions $|\pi,\mathbf{k}\rangle^{\pm}$ (6) the conditions $e^{i\mathbf{k}\cdot\mathbf{c}}=1$; writing these conditions for the wave vector $\mathbf{k}=k_z\mathbf{n}_z+k_x\mathbf{n}_x$, we obtain $k_x=(2\pi/c)m$, $m=0,\pm1,\pm2,\ldots$, and k_z is a continuous number. The discreteness of the wave vector $\mathbf{k}$ means that inside the Brillouin zone the allowed values of $\mathbf{k}$ lie on equidistant lines with spacing $2\pi/c$, and perpendicular to $\mathbf{c}$ (see Fig. 4). In calculating the density of the π electron states of the graphite strip, van Hove singularities arise at energies for which the corresponding level curves are tangent to the equidistant lines inside the Brillouin zone (see Figs. 2 and 4).

Gluing a graphite strip onto a cylindrical surface with radius $R=c/2\pi$, we identify the Cartesian coordinates (x,z) of the points on the strip with the coordinates (φ,z) of the points on the surface of the cylinder according to the rule $2\pi x/c=\varphi$. We associate the elementary translational vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ on the graphite strip with the elementary screw

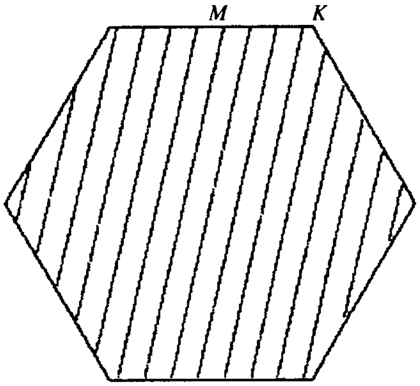


FIG. 4. Allowed states in the Brillouin zone for a (13,3) graphite strip.

rotations $D_1(\varphi_1,z_1)$ and $D_2(\varphi_2,z_2)$ on the surface of the cylinder, and we identify the screw rotation $D_3(\varphi_3,z_3)$ with the basis vector τ_1.

It is easy to obtain from Eq. (6) the following formula for the energy spectrum of electrons on the cylindrical surface of an ideal tubulene:

$$\begin{aligned} E^{\pm}_{\pi,m,k_z} &= E^0_\pi \pm |H_{01,m,k_z}|, \\ H_{01,m,k_z} &= \beta_0 + \exp(-im\varphi_2 - ik_z z_2) + (\beta_1 + \beta_2 \\ &\quad \times \exp(-im\varphi_1 - ik_z z_1)), \end{aligned} \tag{7}$$

where m is the band number, $\hbar m$ is the angular momentum of an electron on the surface of the cylinder, $\hbar k_z$ is the momentum along the generatrix of the cylinder, $\hbar$ is Planck's constant, and the numbers β_0, β_1, and β_2 are the matrix elements of the electron Hamiltonian, calculated with the atomic wave functions of the nearest atoms on the cylindrical surface. Note that the distance between atoms on the cylindrical surface differs from their separation on the graphite strip, and that the angle between π states of neighboring atoms that are locally normal to the surface varies; this is why β_0, β_1, and β_2 are generally different.

In a static magnetic field $\mathbf{B}\|\mathbf{z}$, we must make the substitution $m \to m + \Phi/\Phi_0$ in Eq. (7), where $\Phi = \pi R^2 B$ is the magnetic flux threading the transverse cross section of the tubulene and $\Phi_0 = 2\pi c\hbar/e = 4.1\times 10^{-7}$ G·cm^2 is the elementary quantum of magnetic flux. As the magnetic field changes, the equidistant lines of the allowed values of the wave vector **k** move in the Brillouin zone and, in consequence, all physical characteristics of the nanotube oscillate. The step size in the oscillations with respect to the magnetic field is determined from the relation $\delta B = \Phi_0/\pi R^2$ (Bohm–Aharonov effect).[12] As the magnetic field varies, the type of conductivity of the tube changes periodically from metallic to semiconductor. Thus, oscillations of the physical characteristics of an ideal tubulene in a magnetic field are related to the Bohm–Aharonov effect, and to the fact that the equidistant lines of allowed values of the wave vector periodically coincide with K points of the Brillouin zone, where the gap width vanishes.

We now discuss the relationship between the parameters of our model of the electron energy spectrum and an experimentally observed physical quantity—the resistance of a conducting tube. Let two contacts with potentials equal to zero and V be connected to the ends of a tube of length L. To estimate the current I flowing along the tube, we employ the relaxation time approximation, taking account of the quasi-one-dimensional motion of the electrons and the weakness of the electric field. We also assume that electron transfer processes between bands with different numbers m are forbidden.

As a result, the current assumes the form

$$I = \frac{e^2}{\pi}\frac{V}{L}\sum_m \int_{k_z} v_z^2 \tau \left(-\frac{\partial f^0_{m,k_z}}{\partial E}\right)\bigg|_{E=\mu} dk_z, \tag{8}$$

where $v_z = \hbar^{-1}\partial E_{\pi,m,k_z}/\partial k_z$ is the electron velocity along the tube and f^0_{m,k_z} is the equilibrium Fermi–Dirac function. The summation over m and the integration over k_z in Eq. (8) extend along the equidistant lines inside the Brillouin zone (see Fig. 4); μ is the chemical potential, equal to E^0_π at zero temperature; τ is the electron relaxation time. It follows from Eq. (8) that the numerical value of the current in a conducting tube at zero temperature depends on the dynamical characteristics of the electronic states at the corner points of the Brillouin zone. Mathematically this is related to the fact that on switching to integration over the energy variable in Eq. (8), a Dirac δ function with argument equal to the difference of the energy and the chemical potential arises in the integrand, and therefore only electronic states at the Fermi level that have the same value as the chemical potential contribute to the integral.

In Ref. 13 it is shown that the characteristic electron relaxation time in tubes is of the order of 10^{-12} s; this means that for conducting tubes with characteristic size less than 10^{-4} cm, a collisionless approximation can be used to calculate the resistance. In this approximation electrons travel along the nanotube without scattering, from one contact with potential 0 to the other contact with potential V, and as a result, making the formula substitution $L = |v_z|\tau$, Eq. (8) for the current can be written in the form[12]

$$I = \frac{e^2}{\pi} V \sum_m \int_{k_z} |v_z| \left(-\frac{\partial f^0_{m,k_z}}{\partial E}\right)\bigg|_{E=\mu} dk_z. \tag{8a}$$

If at zero temperature the integration in Eq. (8a) is replaced by integration over the energy variable, the current in the conducting tube will not depend on the type of tube.

Note that in the formulas presented above, we have neglected singularities in electron scattering by the contacts through which voltage is applied to the nanotube.

4. SYMMETRY GROUP OF A TUBULENE

Let us discuss the symmetry group of a tubulene. The group consists of the screw translation operators D_1 and D_2. The method for gluing the tubulene imposes on the operators D_1 and D_2 the condition

$$D_1^{i_1} D_2^{i_2} = E, \tag{9}$$

where E is the identity transformation. It follows from (9) that

$$i_1 z_1 + i_2 z_2 = 0, \quad i_1 \varphi_1 + i_2 \varphi_2 = 2\pi.$$

As a result, the geometry of an infinite nanotube is completely defined by five parameters: φ_1 or φ_2, z_1 or z_2, and φ_3, z_3, and R.

This symmetry group is abelian, its irreducible representations are one-dimensional, and the characters of the irreducible representations can be obtained from the following considerations. Let the characters of the operators D_1 and D_2 be denoted by $\exp(ik_1)$ and $\exp(ik_2)$. Then the character of an arbitrary screw translation operator $D^{l_1 l_2} = D_1^{l_1} D_2^{l_2}$ is $\exp(i(l_1 k_1 + l_2 k_2))$, where l_1 and l_2 are arbitrary integers, $-\pi \leqslant k_1 \leqslant \pi$, and $-\pi \leqslant k_2 \leqslant \pi$. According to Eq. (9) $\exp(i(i_1 k_1 + i_2 k_2)) = 1$ or $i_1 k_1 + i_2 k_2 = 2\pi J$, where J is an arbitrary integer. Then all irreducible representations of the group of screw translations of an ideal graphite nanotube can be labeled by one continuous and one discrete number, and either k_1 or k_2 can be regarded as continuous. The case $i_2 = 0$ is an exception. Nanotubes defined in this manner are known in the literature as zigzags.

Note that for relatively prime numbers i_1 and i_2, the group of screw translations of a tubulene is cyclic, i.e., there exist integers m_1 and m_2 such that the powers of the operator $D^{m_1 m_2}$ fill the entire symmetry group.

In the tight-binding approximation, the wave functions of the π electrons of a tubulene have the form

$$|\pi, k_1, k_2, \gamma\rangle = \frac{1}{\sqrt{N}} \sum_{l_1, l_2} \exp(i(k_1 l_1 + k_2 l_2)) \times D^{l_1 l_2} |\pi, 0, \gamma\rangle, \tag{10}$$

where $|\pi, 0, \gamma\rangle$ is the atomic wave function of a π electron, centered in the zeroth cell on the γth atom, and the numbers l_1 and l_2 determine the number of the cell on the surface of the nanotube. The screw translation operators in Eq. (10) operate on the atomic function of the zeroth cell, translating the π orbitals of the electron along the screw lines and orienting them along a local normal to the cylindrical surface of the nanotube. The choice of functions in the form (10) makes it possible to use the screw symmetry of the tubulene.

By analogy with Eq. (2), by constructing a linear combination of the functions (10) and finding the spectrum of the Hamiltonian, we obtain a formula analogous to Eq. (7) for the energy of the π electrons of an ideal tubulene, in which the numbers m and k_z are replaced by the numbers k_1 and k_2, which are related by $k_1 = m\varphi_1 + k_z z_1$ and $k_2 = m\varphi_2 + k_z z_2$.

5. ELECTRONIC SPECTRUM OF OPTIMIZED NANOTUBES IN A PARAMETRIC TIGHT-BINDING MODEL

To estimate the applicability of the simple π-electron model of a tubulene studied in the preceding section, we performed numerical calculations of the electronic spectra of optimized tubulenes in a parametric tight-binding model. The numerical calculations were performed with a finite tubulene fragment for which boundary conditions of a special type were imposed so as to be able to simulate an infinite tubulene. Our boundary conditions for a tubulene can be explained as follows: powers of the operator D_1 (or D_2) on the surface of a cylinder define an infinite atomic screw line, which when cut lengthwise forms an infinite ribbon packed with carbon atoms along the screw lines. This ribbon covers the surface of the cylinder, touching along the cut.

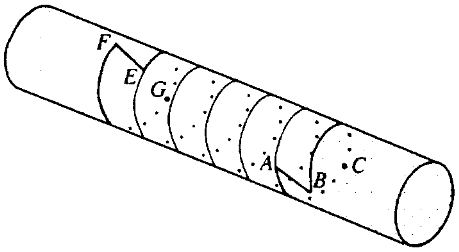


FIG. 5. Ideal-tubulene fragment employed in the numerical calculations.

We introduce generalized Born–Karman conditions on the surface of the cylinder, identifying atoms separated by a fixed number n of screw translations along the screw lines with one another. The number n is chosen from numerical calculations so as to obtain results that have only a weak dependence on n. Figure 5 displays a fragment of a ribbon and elucidates the special periodic boundary conditions (the closed curves ABC are identified with FEG with the required additional rotation). We employed boundary conditions of this type previously to calculate the phonon spectrum of tubulenes.[9]

Note that an ideal nanotube formed by identifying the points of the graphite cavity with points on the surface of the tubulene exhibits periodicity along the axis of the tube, and the ordinary Born–Karman conditions can be used.[14] An advantage of our special boundary conditions is that the dimensions of the computational region can be reduced, and the conditions of invariance with respect to translations along the axis of the tube can be removed. In this manner, in the process of optimization we assume only that the symmetry of the nanotube under the operators D_1 and D_2 is preserved.

In the numerical calculations, we optimized the energy of a tubulene, which in our parametric tight-binding method can be written in the form

$$E = 2 \sum_{\lambda}^{\text{occ}} \varepsilon_\lambda + \sum_{i>j} U(|\mathbf{r}_1 - \mathbf{r}_j|), \tag{11}$$

where ε_λ are the eigenvalues of the Hamiltonian, the summation in the first sum extends over the filled electron states, and the factor 2 in front of the first sum takes account of the spin. The second sum in (11) takes account of many-body effects and repulsion of the inner electron shells, the functions $U(r)$ depend parametrically on the interatomic distances, and $\mathbf{r}_i$ and $\mathbf{r}_j$ are the radius vectors of the atoms in the tubulene. The energy of the nanotube was optimized with respect to the five parameters listed above, which fixed the geometry of the nanotube.

The wave functions of the valence electrons of the tubulene were chosen in the form

TABLE I.

Type of nanotube	Ideal radius, Å	E_g of an ideal nanotube, eV	Optimized radius, Å	E_g of an optimized nanotube, eV	Binding energy, eV/atom
(5,0)	1.99	1.69	2.07	1.14	−6.84
(6,0)	2.39	0.00	2.46	0.49	−6.93
(8,0)	3.18	1.04	3.23	1.14	−7.03
(20,0)	7.96	0.41	7.96	0.44	−7.14
(5,1)	1.82	0.01	1.91	0.49	−6.77
(9,1)	3.40	0.90	3.44	0.61	−7.05
(14,1)	5.39	0.01	5.41	0.07	−7.11
(19,1)	7.37	0.44	7.37	0.48	−7.13
(6,2)	2.11	1.53	2.17	1.44	−6.88
(18,2)	6.80	0.48	6.81	0.52	−7.13
(7,3)	2.42	1.28	2.47	1.15	−6.94
(17,3)	6.26	0.52	6.26	0.57	−7.13
(20,4)	7.30	0.01	7.30	0.06	−7.13
(15,5)	5.27	0.61	5.28	0.64	−7.11
(12,6)	4.14	0.02	4.16	0.00	−7.08

$$|\alpha,k_1,k_2,\gamma\rangle=\frac{1}{\sqrt{N}}\sum_{l_1,l_2}\exp(i(k_1l_1+k_2l_2))\times D^{l_1l_2}|\alpha,0,\gamma\rangle, \qquad (12)$$

where α is the number of the carbon atom orbital (s and p orbitals) and $|\alpha,0,\gamma\rangle$ is the orbital in the zeroth cell on the atom with number $\gamma=0,1$. The screw-translation operators in Eq. (12) rotate and translate the local system of coordinates together with the atomic orbitals specified in this system and determined in the zeroth cell of the tubulene.

Constructing from Eq. (12) a linear combination over the number of atoms and atomic orbitals in a unit cell, we solve the eigenfunction and eigenvalue problem for the electron Hamiltonian, yielding eight values of the electron energy for each fixed set of numbers k_1 and k_2. We used the formulas from Ref. 10 to calculate the matrix elements of the Hamiltonian, which were determined in terms of the coordinates of the atoms and the local orientations of the orbitals calculated according to the action of the screw-translation operators.

The equilibrium structure was determined using Eq. (11), and the electronic structure of the nanotubes with different indices i_1 and i_2 were investigated. Minimizing the energy (11) of a nanotube changes the interatomic distances (1.39–1.45 Å depending on the type of tube) somewhat. As the tubulene radius increases, the interatomic distances approach 1.44 Å, which corresponds to the equilibrium interatomic distance obtained in the graphite plane with the chosen parametric model.

The electron densities of states of optimized nanotubes with indices (7.3) and (14.1) (see Table I below) are presented, as an example, in the right-hand sides of Figs. 6a and 6b. The electronic spectra of the corresponding graphite strips are presented for comparison on the left-hand sides of Figs. 6a and b. The differing electronic densities of states near the Fermi level of an optimized tubulene and a corresponding graphite strip are due both to the partial hybridization of the electronic σ and π states, and a shift in the σ- and π-electron energy as a result of the deviation of the tubulene geometry from a flat geometry. In the parametric tight-binding method which we are examining, the corrections due to nonplanarity of the structure are proportional to a/R for the $\pi-\sigma$ matrix elements, and $(a/R)^2$ for the $\pi-\pi$ and $\sigma-\sigma$ matrix elements of the Hamiltonian; in the nondegenerate case, the corrections to the electron energy will be proportional to the squared matrix element of the Hamiltonian i.e., $(a/R)^2$.[15] It also follows from Fig. 6 that as the indices of the nanotube increase, the number of van Hove singularities increases, and an average envelope coinciding with the π-electron density of states for the graphite plane is formed.

Our computed values of the radii and gap widths for a number of optimized nanotubes, the radii of ideal tubes, and the gap widths obtained by the zone-folding method are presented in Table I. As follows from the table, for nanotubes with small radii ($R\sim2-3$ Å), the numerical values of the gaps and radii of the ideal and optimized tubulenes differ, and for nanotubes with radii $R>5$ Å, the differences are small. The gap for an ideal nanotube was estimated according to Eq. (6) for the numerical value of the matrix element $\beta_0=\beta_1=\beta_2=2.02$ eV, taken from Ref. 10.

Other methods of parameterizing the matrix elements of the Hamiltonian in the tight-binding approximation are also discussed in the literature. Additional parameters can be introduced,[16] and nonorthogonality of the atomic orbitals can be taken into account (see, for example, Ref. 17).

Note that the published data on the numerical values of the gap widths in nanotubes with a small radius exhibit a large variance. For example, in Ref. 18 gap widths obtained by different methods are presented for an (8,0) tube: 1.22 eV, 1.19 eV, and 0.62 eV. There are discrepancies in the results of semi-empirical and first-principles methods for the metallic conductivity of a (6,0) tube.[2,18] In the literature, an interpolation formula of the form $\beta_0 a/R$ is used to determine the gap width of a nonconducting carbon nanotube with a large radius. The values of the parameter β_0 vary over the range 2.0–2.8 eV, depending on the parameterization of the tight-binding model.

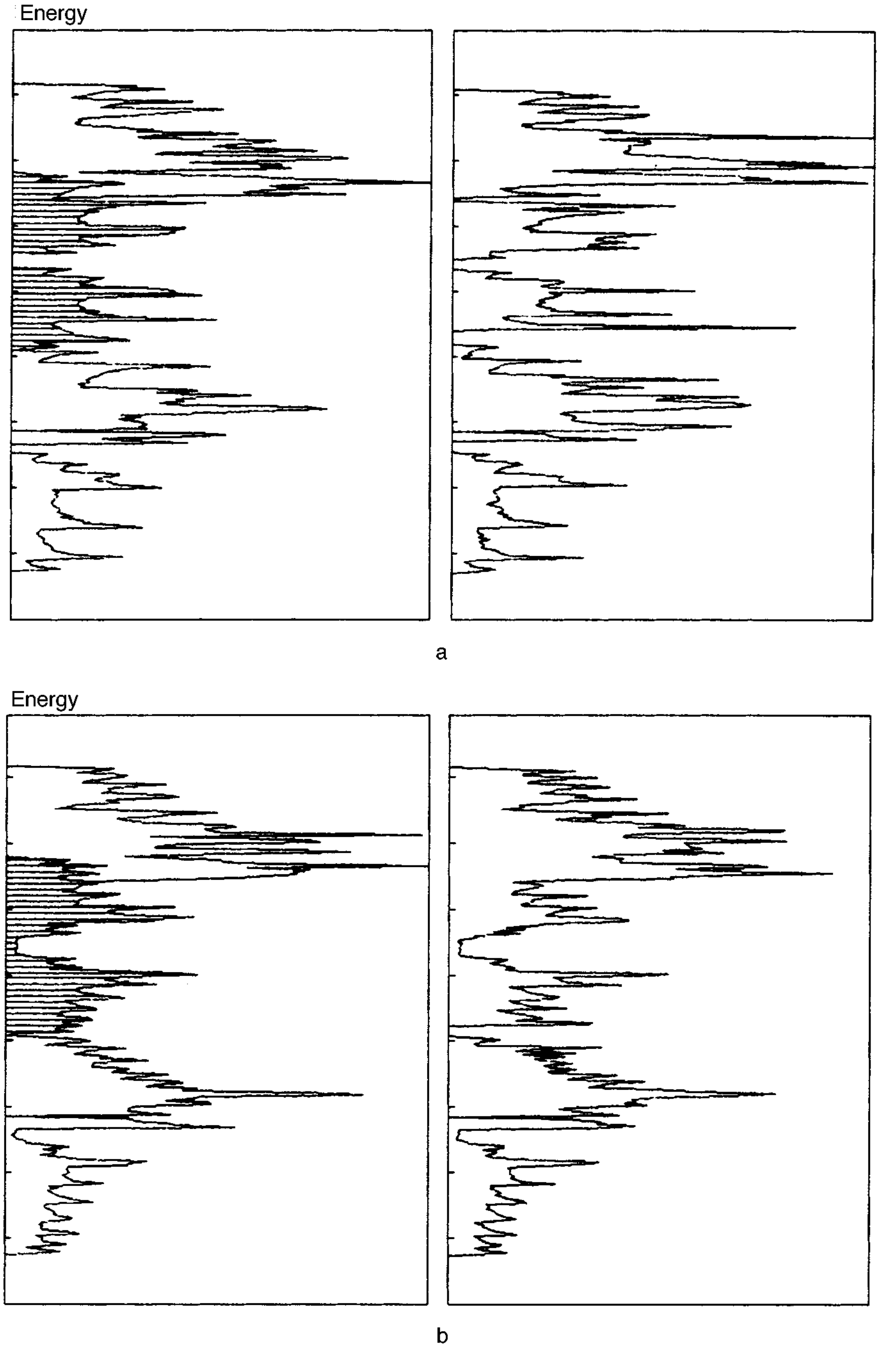


FIG. 6. Electron density of states of a graphite strip (left-hand figures) and optimized tubulenes (right-hand figures). Hatching signifies π-electron partial densities. The unit on the energy scale is 5 eV.

Note also that the use of a non-self-consistent tight-binding model to investigate the electronic properties of small-radius nanotubes can yield incorrect results. In a non-self-consistent model, electron–electron interactions, which push electrons out of the inner region of the nanotube, are not described accurately.[18] Nonetheless, comparing our results with the first-principles calculations in Refs. 3 and 18, we see that the model yields a qualitatively correct description of the electronic structure of the tubes.

6. CONCLUSIONS

In this paper, we analyzed a theoretical zone-folding model for ideal nanotubes, which enables one to study transport phenomena and electronic properties in a ±5 eV energy band around the level of the chemical potential, and we provided a transparent geometric interpretation of the electronic states in the first Brillouin zone, making it possible to understand qualitatively the formation of an energy gap in the electron spectrum and the transformation of the spectrum in a static magnetic field. In the parametric tight-binding method, we directly assessed the applicability of the zone-folding model by means of a numerical calculation of the atomic and electronic structure of an optimized tubulene, taking account of the *s* and *p* orbitals of the valence electrons of the carbon atoms. Comparing the gap widths obtained from our numerical calculations of optimized tubulenes and the data obtained from the zone-folding model for the corresponding graphite strip, we find that the models disagree for small-radius tubulenes; for large-radius tubulenes, however, they agree.

[1] S. Iijima, Nature **354**, 56 (1991).

[2] N. Hamada, S. Sawada, and A. Oshiyama, Phys. Rev. Lett. **68**, 1579 (1992).

[3] J. W. Mintmire, D. H. Robertson, and C. T. White, J. Phys. Chem. Solids **54**, 1835 (1993).

[4] S. B. Trickey, F. Muller-Plathe, G. H. F. Diercksen, and J. C. Boettger, Phys. Rev. B **45**, 4460 (1992).

[5] D. J. Klein, A. Seitz, and T. G. Schmalz, J. Phys. Chem. **97**, 1232 (1993).

[6] C. T. White, D. H. Robertson, and J. W. Mintmire, Phys. Rev. B **47**, 5485 (1993).

[7] R. Saito, M. Fujita, and G. Dresselhaus, Phys. Rev. B **46**, 1804 (1992).

[8] N. V. Khokhryakov and S. S. Savinskiĭ, Fiz. Tverd. Tela (St. Petersburg) **36**, 3524 1994 [Sov. Phys. Solid State **36**, 1872 (1994)].

[9] N. V. Khokhryakov, S. S. Savinskiĭ, and J. M. Molina, JETP Lett. **62**, 617 (1995).

[10] L. Goodwin, J. Phys.: Condens. Matter **3**, 3869 (1991).

[11] A. A. Abrikosov, *Fundamentals of the Theory of Metals*, North-Holland, New York (1988).

[12] M. F. Lin, W. Kenneth, and K. Shung, Phys. Rev. B **51**, 7592 (1995).

[13] R. A. Jishi, M. S. Dresselhaus, and G. Dresselhaus, Phys. Rev. B **48**, 11385 (1993).

[14] J. Y. Yi and J. Bernholc, Phys. Rev. B **47**, 1708 (1993); D. H. Robertson, D. W. Brenner, and J. W. Mintmire, Phys. Rev. B **45**, 12592 (1992).

[15] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics*, Pergamon Press, New York (1977).

[16] C. H. Xu, C. Z. Wang, C. T. Chang, and K. M. Ho, J. Phys.: Condens. Matter **4**, 6047 (1992).

[17] M. van Schilfgaarde and W. A. Harrison, J. Phys. Chem. Solids **46**, 1093 (1985).

[18] X. Blase, L. X. Benedict, E. L. Shirley, and S. G. Louie, Phys. Rev. Lett. **72**, 1878 (1994).

Translated by M. E. Alferieff

Relaxation and 1/*f* noise in phonon systems

Yu. E. Kuzovlev

Donetsk Physicotechnical Institute, Ukrainian National Academy of Sciences, 340144 Donetsk, Ukraine

(Submitted 17 June 1996; resubmitted 19 December 1996)
Zh. Éksp. Teor. Fiz. **111**, 2086–2098 (June 1997)

This paper examines the relationships that exist between low-frequency fluctuations of the rate of dissipation in nonequilibrium thermodynamic systems and higher-order multitime statistical moments of equilibrium noise. In particular, it studies the relationships between internal friction fluctuations in the phonon system being excited and low-frequency fluctuations of Raman scattering of light in an equilibrium phonon system. We show that both processes are related to strong fluctuations in the phase diffusion rate and the relaxation of phonon modes generated, in turn, by the exponential instability of the dynamical paths of the system. © *1997 American Institute of Physics.* [S1063-7761(97)01206-7]

1. INTRODUCTION

Low-frequency fluctuations with a $1/f$ spectrum observed in various systems (in particular, fluctuations of electrical conductivity in solids and frequency fluctuations in quartz frequency stabilizers[1,2]) constitute one of the most intriguing problems of physics. The tendency today is to relate $1/f$ noise to metastable states, such as defects in the atomic structure and electron traps, characterized by a broad range of lifetimes. Typically, however, specifying the physical nature of such a set of ''fluctuants'' proves impossible.

A fundamentally different explanation of $1/f$ noise is given in Ref. 3. Suppose that a random dynamical parameter of elementary kinetic processes influences the effectiveness of these processes in the sense of their contribution to relaxation and dissipation, but does not affect the thermodynamic state of the system. Then the corresponding random variations in the rate of relaxation and dissipation are not balanced by a ''restoring thermodynamic reaction'' of the system. As a result they do not have a characteristic upper lifetime, and at low frequencies their spectrum must be of the scale-invariant power type.

The effect of a random impact parameter for particle collisions in an equilibrium gas discussed in Ref. 3 may serve as an example. The lack of a dynamical or thermodynamic mechanism that would ensure a specified distribution of collisions over the impact parameter, i.e., a certain ''weighted-mean'' (in the impact parameter) scattering cross section, leads[3] to $1/f$ fluctuations of the rate (or coefficient) of self-diffusion of the gas particles. Thus, $1/f$ noise appears in a natural way as a constituent of thermal noise, despite the fact that there are sure to be no ''slow'' processes. Other examples of this kind are discussed in Refs. 4 and 5.

The possibility of such a fundamental source of $1/f$ noise applies as well to vibrations in quartz and, in general, dielectric crystals. As is known,[2] the observed dependence of the level of relative $1/f$ frequency fluctuations in quartz oscillators and frequency stabilizers on their Q-factors indicates that such fluctuations are a reflection of fluctuations of internal friction in crystals. Here it must also be said[2] that friction fluctuations are enormous: the relative variations in the internal friction in the active mode of a crystal are of order unity, and can be much larger.

Additional information about thermodynamic equilibrium $1/f$ noise in quartz can be found in Ref. 6. Musha *et al.*[6] studied spontaneous Raman scattering of laser light in quartz and discovered $1/f$ fluctuations in the number of scattered photons (the photocount flux density) at frequencies below 0.01 Hz. According to the researchers, the level of relative fluctuations in the photocount density is inversely proportional to the number of phonon modes participating in the scattering onto the photon counter.

From the standpoint of $1/f$ noise theory, a purely phonon system serving as a model of a weakly anharmonic dielectric crystal is an extremely interesting object, since it is one of the basic models of statistical physics (a weakly nonideal gas is another). Both approximate analytic methods and computer simulations show[7,8] that as the number of nonlinearly interacting oscillators grows, the fraction of regions of regular motion in phase space diminishes, and the behavior of the system is governed by the exponential instability of the paths under a perturbation. It is also known that the characteristic instability development times and the characteristic relaxation times (determined from the decay of two-time correlators) are actually the same thing. However, the statistical characteristics of instability responsible for the fluctuations in the ''rate'' of relaxation and dissipation, described by multitime correlators (cumulants) of fluctuations in the equilibrium ensemble, and, as shown in Refs. 3 and 4, related to the problem of $1/f$ noise, have yet to be thoroughly studied.

The problem of discriminating fundamental $1/f$ noise in a phonon system appears to be more complicated than in the case of a gas, since here we are dealing with a completely different interaction between the quasiparticles (weak and continuous instead of strong and sporadic). It is therefore advisable to consider the general properties of multitime correlators and their relation to exponential instability in order to clarify the problems that a future constructive theory is sure to encounter.

The idea of $1/f$ noise being caused by fluctuations in the dissipation rate was first proposed in general form by Handel.[9–11] Handel's pioneering quantum theory of $1/f$ noise[9–13] relates these fluctuations (and, in the final analysis,

$1/f$ noise) to the scattering cross section fluctuations in elementary kinetic processes. The cross section fluctuations are related in turn to the emission of ultrasoft photons in electrodynamic processes and of other quanta of massless fields or excitations, including gravitational fields. The resulting formulas of this theory often demonstrate remarkable quantitative agreement with the experimental data.[12–14] In particular, according to Refs. 2, 12, and 13, it provides a quantitative interpretation for $1/f$ noise in quartz.

As for phonon systems, Handel's theory[9–13] states that $1/f$ noise is closely related to the electric (more precisely, piezoelectric) properties of quartz; not only do these properties provide an easy way of observing $1/f$ noise, they also generate this noise, in accordance with the dominant role that electromagnetic interactions and ultrasoft electromagnetic field quanta play in Handel's theory. According to the viewpoint expressed in this paper, however, lattice dynamics is capable of generating $1/f$ noise by itself, irrespective of the type of interaction forces between the lattice particles.

The present study has two goals. First, it introduces the general relationships that exist between excess noise in a nonequilibrium system excited by extraneous forces and equilibrium correlators. We will see that in the event of Raman scattering of light in a phonon system, the fluctuations in dissipation (internal friction) and $1/f$ noise are actually described by the same multitime (i.e., referring to more than two moments in time) higher-order correlators (cumulants, to be precise). Generally such correlators carry information about low-frequency fluctuations of the kinetic and transport characteristics of the system. Second, we will provide arguments in favor of the fact that dissipation fluctuations, as well as scattering fluctuations, are related to the large and long-lived friction fluctuations of each phonon mode, the friction being caused by mode interaction. The same fluctuations are in turn generated by the exponential instability of small perturbation of the dynamic path of the system. Their long-lived nature follows from the fact that in a system dynamically becoming stochastic, the random (i.e., not averaged over the ensemble) response of the variables of a given mode to their past values never decays. As a result, the rapid decay of ordinary quadratic equilibrium correlations, which is a reflection of dissipation and irreversibility, is inevitably accompanied by the emergence of infinitely long-lived higher-order correlations.

2. THE MODEL

We consider a classical system whose Hamiltonian is

$$H'(q,p,t)=H(q,p)-\sum q_k f_k(t),$$

$$H(q,p)=\frac{1}{2}\sum p_k^2+U(q),$$

where $\{q_k,p_k\}\equiv X$ are canonically conjugate coordinates and momenta, and the force $f(t)$ specifies an extraneous dynamic perturbation. To simplify matters, we assume all masses equal to unity, so that the velocities $v_k\equiv dq_k/dt$ coincide with the momenta, $v_k=p_k$. The label k in the phonon system numbers either the positions of atoms or the vibrational modes defined in the harmonic approximation. In the latter case, and in the presence of weak cubic anharmonicity and spatial uniformity (translational invariance), $U(q)$ typically has the form

$$U(q)=\frac{1}{2}\sum \omega_k^2 q_k^2+\frac{\lambda}{6}N^{-1/2}\sum u_{ijk}\omega_i\omega_j\omega_k q_i q_j q_k+\dots .$$

Here the coefficients u_{ijk} are nonzero only if a combination of wave vectors $i\pm j\pm k$ is zero or an integral multiple of the reciprocal lattice vector, N is the total number of particles or modes, and the anharmonicity parameter λ is of order $\vartheta^{-1/2}$, with ϑ the "melting point" of the lattice.

3. GENERAL RELATIONSHIPS FOR RELAXATION AND DISSIPATION FLUCTUATIONS

1. Suppose that initially the system is in a thermodynamically stable state described by the canonical ensemble. Energy dissipation due to the perturbation, i.e., the work A done by the perturbing forces,

$$A\equiv\int dH(q(t),p(t))=\int v(t)f(t)dt,$$

is determined by the statistical characteristics of the velocities conjugate to the forces. An arbitrary nonequilibrium-velocity cumulant can be reduced to what is known as "quasiequilibrium" correlators (cumulants), which we denote by $\langle\cdots\rangle_q$. By definition, the cumulant $\langle v(t_1),\dots,v(t_n)\rangle_q$ corresponds to the modified perturbation $f(t)\eta(t-\min(t_1,...,t_n))$, where $\eta(x)$ is the Heaviside step function, which "turns on" at time $t_0\equiv\min(t_1,...,t_n)$ and coinciding with $f(t)$ for $t>t_0$. As shown in Refs. 15 and 16,

$$\langle v(1),\dots,v(n)\rangle$$
$$=\langle v(1),\dots,v(n)\rangle_q$$
$$+\frac{1}{T}\int_{-\infty}^{\min(1,\dots,n)}\langle v(1),\dots,v(n),v(t)\rangle_q f(t)dt,$$

where T is the initial temperature of the system, and where for the sake of brevity we denote the time variables by integers. The angle brackets containing m commas (Malakhov's cumulant bracket) denote the combined $(m+1)$st-order cumulant of the $m+1$ expressions separated by commas.

This formula yields the following expressions for the average work $\langle A\rangle$ and the corresponding variance $\langle A,A\rangle$:

$$\langle A\rangle=\int f(1)\langle v(1)\rangle d1$$
$$=\frac{1}{T}\int f(1)\langle v(1),v(2)\rangle_q f(2)d1\,d2,$$
$$\langle A,A\rangle=\int f(1)\langle v(1),v(2)\rangle f(2)d1\,d2$$
$$=2T\langle A\rangle+\frac{2}{T}\int f(1)f(2)$$
$$\times\langle v(1),v(2),v(3)\rangle_q f(3)d1\,d2\,d3, \qquad (1)$$

where all integrals over the time variables meet the condition $1>2>3$. Thus, the fluctuations of the work of perturbing forces consist of two contributions. The first, $2T\langle A\rangle$, can be interpreted as shot noise corresponding to the average energy influx $\langle A\rangle$, and the second as excess noise. We see that the excess contribution is determined by higher-order cumulants, with the result that it vanishes in the Gaussian approximation.

Let us suppose that the perturbation is quasiperiodic and ''infinitely weak.'' Then if the equilibrium quadratic velocity correlator decays rapidly enough, both the average value of the absorbed energy and the shot contribution to the fluctuations of this energy are proportional to $f^2\vartheta$, where ϑ is the time interval within which the perturbation acts.

Next we allow for the fact that the equilibrium ternary velocity cumulant is zero because velocities are odd under time reversal. The ''quasiequilibrium'' ternary cumulant in the expression for $\langle A,A\rangle$ is therefore proportional to f, and on the whole the excess component of $\langle A,A\rangle$ is of order f^4. We isolate the fourth power of the perturbation explicitly by introducing a function characterizing the dynamic linear differential response of velocities to the weak perturbation:

$$\Gamma_{jk}(t-t_0)\equiv \left.\frac{\delta v_j(t)}{\delta f_k(t_0)}\right|_{f=0}.$$

Then from (1) we get

$$\langle A,A\rangle = 2T(A)+\frac{2}{T}\int_{(1>4,2>3,2>4)} f(1)f(2) \times\langle v(1),\Gamma(2,3),v(4)\rangle_0 f(3)f(4)\,d1\ldots d4, \tag{2}$$

where the subscript ''0'' indicates that the correlator is an equilibrium one, higher-order terms are dropped, and the ''lower limit of integration'' indicates the rule by which the variables of integration are ordered.

If the integral in (2) increases in proportion to the duration of the perturbation, i.e., $\sim\vartheta f^4$, then qualitatively the excess part of the noise differs in no way from shot noise, introducing only a small correction. But if the integral grows rapidly, say $\sim\vartheta^{s+1}f^4$ $(s>0)$, it can be interpreted only as a contribution due to fluctuations in the system's dissipation parameters, and we can easily show that it is equivalent to low-frequency fluctuations with a spectrum of type ω^{-s}.

Recall now that for the equilibrium canonical Gibbs ensemble we have the following relationship (which can easily be obtained via integration by parts):

$$\langle v(t_0)\Phi\{X(t)\}\rangle = T\left\langle \frac{\partial\Phi\{X(t)\}}{\partial v(t_0)}\right\rangle, \tag{3}$$

where $\Phi\{\ \}$ denotes an arbitrary functional of the path $X(t)$ interpreted as a function of the ''initial'' values $X(t_0)$ at a fixed time t_0. Selecting the appropriate functionals and using the equalities

$$\left.\frac{\partial X(1)}{\partial f(0)}\right|_{f=0}=\frac{\partial X(1)}{\partial v(0)},\qquad \Gamma_{jk}(t,t_0)\equiv\frac{\partial v_j(t)}{\partial v_k(t_0)},$$

which follow from the Hamiltonian equations of motion, we find that (3) yields

$$\langle v(1),v(0)\rangle = T\langle\Gamma(1,0)\rangle,$$

$$\langle v(1),v(2),v(3),v(0)\rangle = T\langle\Gamma(1,0),v(2),v(3)\rangle + \text{permutations}. \tag{4}$$

As Eqs. (2) and (4) imply, the excess fluctuations of the work (dissipation fluctuations) are related to fourth-order equilibrium cumulants and therefore cannot be described consistently by the Gaussian approximation.[1]

2. The same cumulants or, to be more precise, the cumulants of the complex-valued phonon-mode amplitudes $a(t)$ (which can be reduced to the former) are responsible for fluctuations in the intensity of light scattering.[2] Indeed, the photon scattering cross section (or the scattering probability per unit time) for a given mode incorporates $|\int a(t)\exp(i\Omega t)dt|^2$, where Ω is the combination frequency, and the integral is taken over a time interval of the order of the reciprocal Brillouin linewidth. Accordingly, the observed low-frequency fluctuations are related to four-point correlators of the type

$$\langle a^+(t_1)a(t_1'),a^+(t_2)a(t_2')\rangle = \langle a^+(t_1)a(t_2')\rangle\langle a(t_1')a^+(t_2)\rangle + \langle a^+(t_1),a(t_1'), a^+(t_2),a(t_2')\rangle, \tag{5}$$

where the pair t_1,t_1' is far from the pair t_2,t_2'. In view of this, the last (cumulant) term on the right-hand side of Eq. (5), which describes four-time correlations, plays a decisive part.

Note that fluctuations in the occupancy of the crystal's phonon modes, $a^+(t)a(t)$, are described only by two-time cumulants, although these cumulants are fourth-order. Musha *et al.*[6] proposed that the occupancy fluctuation is the reason for $1/f$ fluctuations in the scattering intensity. But variations in the number of photons in a mode will inevitably generate a restoring thermodynamic reaction. Hence the occupancy fluctuations have a characteristic lifetime and cannot be the reason for the $1/f$ spectrum (but, of course, they can acquire an induced low-frequency component in a nonequilibrium state).

We see that $1/f$ noise in light scattering must be interpreted from the standpoint of four-time cumulants, which incorporate not only fluctuations of the number of phonons but also, as we will shortly see, fluctuations of the phonon-mode relaxation times (and therefore make it possible to reduce fluctuations of the scattering probability to the latter fluctuations).

Here one must bear mind that the ''relaxation time'' (for example, the time $t-t_0$ it takes the correlator $\langle a^+(t)a(t_0)\rangle$ to decay) is determined by the rate of diffusion of the photon-mode phase, since it is phase diffusion that is responsible for the broadening of the vibration spectral line and the continuous spectrum. Therefore, it would be correct to say that four-time cumulants (in contrast to four-point two-time cumulants) determine the fluctuations in the phase diffusion coefficient.

4. MULTIPOINT CUMULANTS AND FRICTION FLUCTUATIONS

Let us select a specific phonon mode. In relation to it the other modes act as a heat bath, a source of friction and, at the same time, a random ''Langevin force.'' A nonrandom linear law of friction would correspond to a force with Gaussian statistics. However, the fact that the intermode interaction is nonlinear leads to a non-Gaussian force, as well as to nonlinearity in friction and to friction fluctuations. The latter means that it is impossible to unambiguously separate the Langevin force and friction. But suppose that friction fluctuations are slower than occupancy relaxation. Then it is natural to interpret the variables of the selected mode as random processes ''modulated'' by the low-frequency variations in friction. Such processes would be Gaussian under deterministic modulation, but in view of the random nature of the modulation, they become non-Gaussian.[3)]

From the standpoint of the given phenomenological picture, each statistical moment is a result of double averaging: first the Gaussian noise at fixed friction, and then the friction fluctuations. After the first averaging, the fourth statistical moment of the amplitude has the form

$$\langle a^+(1)a(3)a^+(2)a(4)\rangle = K(13)K(24)+K(14)K(23),$$

where $K(12)\equiv\langle a^+(1)a(2)\rangle$ characterizes the relaxation on the interval between 1 and 2. After the second averaging (of the friction fluctuations), which we denote by double angle brackets, for the correlator (5) we have the following chain of equalities:

$$\begin{aligned}
&\langle\langle a^+(1)a(3),a^+(2)a(4)\rangle\rangle\\
&\quad\equiv\langle\langle a^+(1)a(3),a^+(2)a(4)\rangle\rangle\\
&\qquad-\langle\langle a^+(1)a(3)\rangle\rangle\langle\langle a^+(2)a(4)\rangle\rangle\\
&\quad=\langle\langle a^+(1)a(3),a^+(2)a(4)\rangle\rangle-\langle K(13)\rangle\langle K(24)\rangle\\
&\quad=\langle K(13),K(24)\rangle+\langle K(14),K(23)\rangle.
\end{aligned}$$

Comparing this with (5) yields

$$\langle a^+(1),a(3),a^+(2),a(4)\rangle=\langle K(13),K(24)\rangle +\langle K(14),K(23)\rangle. \tag{6}$$

We see that a four-time cumulant can be interpreted as the characteristic of dissipation fluctuations. In view of (4), the same can be said of the combined cumulant of velocities and differential response in (2).

Let us estimate the order of magnitude of the correlators (6) and (4). First we note that the anharmonicity of the potential energy $U(q)$ is sure to generate single-time multipoint correlations between the amplitudes q_k or a_k. In particular, when the anharmonicity parameter λ is small, each quadruplet of modes for which one of the combinations $j\pm k\pm l\pm m$ coincides with a reciprocal lattice site is coupled by the cumulant $\langle q_j,q_k,q_l,q_m\rangle\sim\lambda^2T^3N^{-1}(\omega_j\omega_k\omega_l\omega_m)^{-1}$. Hence, single-time quadruple correlations are of order N^{-1}. The equations of motion clearly show that the four-time cumulants of q_k and velocities v_k must be at least of the same order.

Even such smallness of the higher-order correlators does not mean that they can be ignored: while being ''smeared'' in the momentum space, they are localized in the coordinate representation and do not depend on the size of the system, N. But actually the equations of motion supply even more information. When there is mixing in the phase space, i.e., when small perturbations are exponentially unstable, even in the momentum representation at least some of the four-time cumulants of q_k and v_k do not contain the small parameter $1/N$.

We assign the label ''0'' to the specified phonon mode. Putting $Q\equiv q_0$, in the potential energy, we ''isolate'' the interaction of this mode with other modes:

$$U(q,Q)=\tfrac{1}{2}\omega_0^2Q^2+U_0(q)+\mu Qu(q),\quad \mu\equiv\lambda N^{-1/2},$$

where we have limited ourselves to the cubic approximation. The equations of motion

$$\frac{d^2q_k}{dt^2}=F_k+f_k,\quad F_k\equiv-\frac{\partial U_0}{\partial q_k},\quad f_k=-Q\mu\,\frac{\partial u(q)}{\partial q_k},$$

$$\frac{d^2Q}{dt^2}=-\omega_0^2Q-\mu u(q)$$

show that the contribution due to the interaction of the selected mode with the other modes, f_k, to the force acting on any of them is fairly small, in view of the factor $N^{-1/2}$. Indeed, with allowance for the above-mentioned property of the coefficients u_{ijk} and the fact that $\langle q_j^2\rangle=T\omega_j^{-2}$, we arrive at the following estimates:

$$\langle f_k^2\rangle\simeq(\mu\omega_0\omega_k)^2\langle Q^2\rangle\left\langle\left(\sum_j u_{0jk}\omega_jq_j\right)^2\right\rangle\sim T^2\mu^2\omega_k^2,$$

$$\langle F_k^2\rangle\sim T\omega_k^2,\quad \frac{\langle f_k^2\rangle}{\langle F_k^2\rangle}\sim\frac{T}{\vartheta}\,N^{-1}.$$

Assuming, on the basis of this, that f_k is a small perturbation, we can express its effects in terms of the linear differential response

$$R_{kj}(tt')\equiv\frac{\delta q_k(t)}{\delta f_j(t')}=\frac{\partial q_k(t)}{\partial v_j(t')}.$$

We have

$$q(t)\simeq q^0(t)+\int R(tt')f(t')dt'\equiv q^0(t)+q^1(t),$$

where q^0 describes the evolution of the heat bath, which is not perturbed by the selected mode. Plugging this expansion into the equation for the latter and linearizing $u(q)$ in $q^1(t)$ on the same grounds, we arrive at the approximate Langevin equation

$$\frac{d^2Q}{dt^2}+\omega_0^2Q=\xi(t)+\int_{-\infty}^{t}\gamma(t,t')Q(t')dt'. \tag{7}$$

Here $\xi(t)\equiv-\mu u(q)$ acts as the Langevin force, and the integral with the kernel

$$\gamma(t,t')\equiv\mu^2\sum_{jk}\,\frac{\partial u}{\partial q_j}\,(t)R_{jk}(tt')\,\frac{\partial u}{\partial q_k}\,(t') \tag{8}$$

determines the frequency renormalization (which, generally speaking, depends on the heat bath variables) and, more importantly, the nonlocal random friction, with unperturbed heat bath variables ($q=q^0$).

The linearization procedure would make sense and would be justified if the resulting friction fluctuations were small (say, of the order $1/N$). Fluctuations of the integral in (7) would be of the same order if we were to replace the differential response $R_{kj}(tt')$ and estimate the remaining factors in the Gaussian approximation. However, with allowance for the random nature of $R_{kj}(tt')$, the variance of the integral in (7) contains, among other things, terms like

$$\sum_{jk} L_{jj}(tt)\int\int L_{kk}(t_1t_2)\langle R_{jk}(tt_1)R_{jk}(tt_2)\rangle$$
$$\times\langle Q(t_1)Q(t_2)\rangle dt_1dt_2,$$

where L_{jk} denotes the correlator of the variables $\mu\partial u/\partial q$. But exponential instability means[7,8] that the variance of the elements of the linear differential response matrix $G(tt')\equiv\partial X(t)/\partial X(t')$, and, in particular of the submatrix R, grow exponentially as the temporal arguments move apart. Hence the correlator $\langle R_{jk}(tt_1)R_{jk}(tt_2)\rangle$ is an exponentially growing function of the difference argument $t-\max(t_1,t_2)$ (also, obviously, containing a nonoscillatory component), which means that the integral is divergent in one of the two integration variables.

Thus, in the linear approximation, the term with friction in the Langevin equation has infinite variance. Assuming, on the basis of the weakness of mode interaction, that this equation is linear, we arrive at an inconsistency: an infinite total reaction of the lattice to the selected mode.

Two interrelated conclusions can be drawn here. First, frictional fluctuations are not really small,[4)] i.e., Eq. (6) can become comparable in magnitude with $\langle K(11')\rangle\langle K(22')\rangle$. Second, a consistent Langevin equation must incorporate nonlinear friction.[5)] Since in view of its anharmonicity nonlinear friction is small and at the same time must suppress the infinity of the linear approximation, the resulting frictional fluctuations remain large.

5. RELAXATION AND LONG-LIVED MULTITIME CORRELATIONS

1. Let us discuss the nature of the decay of higher-order cumulants. First we show that these cumulants exhibit infinitely long-lived behavior, at least in principle. Let us consider the $2N$-by-$2N$ differential response matrix $G(tt_0)\equiv\partial X(t)/\partial X(t_0)$. The $N\times N$ matrices Γ and R defined above coincide, respectively, with the lower right and lower left quadrants of G. The matrix G obeys the linear equation

$$\frac{dG}{dt}=\begin{pmatrix}0 & I\\ \partial^2U/\partial q^2 & 0\end{pmatrix}G\equiv B(t)G.$$

Since in the absence of external forces (i.e., $f=0$) the column vector $dX/dt\equiv\dot X=\{v,dp/dt\}\equiv\{v,F\}$, which is tangential to the path, satisfies a similar equation $d\dot X/dt=B\dot X$, it is a linear combination of the columns of matrix G. This yields the relationship

$$\dot X(t)=G(tt_0)\dot X(t_0), \tag{9}$$

which expresses the invariance of the autonomous motion of the system under shifts in time. A similar relationship follows from energy conservation in autonomous evolution. Because in the absence of external forces the system is conservative, we have $\partial H(t)/\partial X(t_0)=\partial H(t_0)/\partial X(t_0)$, with the result that

$$\nabla H(t)G(t,t_0)=\nabla H(t_0), \tag{10}$$

where $\nabla H\equiv\partial H/\partial X=\{-F,v\}$ must be interpreted as a row vector. The two relationships are linked by the identity

$$G^{-1}=-gG^{+}g,\quad g\equiv\begin{pmatrix}0 & I\\ -I & 0\end{pmatrix},$$

in which I is the $N\times N$ identity matrix, and the "+" on G denotes Hermitian conjugation. Incidentally, Eqs. (9) and (10) show that the vectors $\dot X(t)$ and $\nabla H(t)$, being bounded everywhere, lie in the (two-dimensional[7)]) subspace of perturbations that neither expand nor contract in time.

By multiplying (9) or (10) into any functional of $X(t)$ and averaging over an arbitrary equilibrium path ensemble (an invariant measure on the set of initial conditions), we arrive at formally exact relationships between the correlators. In particular, from Eq. (9) we obtain

$$\sum_k [\langle v_k(t)\Gamma_{kj}(tt_0)v_j(t')\rangle-\langle \dot p_k(t)R_{kj}(tt_0)v_j(t')\rangle]$$
$$=\langle v_j(t_0)v_j(t')\rangle=T\langle\Gamma_{jj}(t_0t')\rangle, \tag{11}$$

where the second equality refers to a canonical ensemble.

In each equilibrium correlator on the left-hand side of Eq. (11), we isolate the connected cumulant part, bearing in mind that $\langle\dot X(t)\rangle=0$, so that, for instance,

$$\langle v(t)\Gamma(tt_0)v(t')\rangle=\langle v(t)v(t')\rangle\langle\Gamma(tt_0)\rangle$$
$$+\langle v(t),\Gamma(tt_0),v(t')\rangle.$$

Let us suppose that the quadratic correlators (and hence $\langle\Gamma(t,t_0)\rangle$) tend to zero as $t-t_0$ and $t-t'$ increase. Then, as the moment t moves still farther away, the corresponding terms vanish, and all the dynamic correlations reflected by (9) become shifted to higher-order cumulants. Hence at least some of these do not disappear for arbitrarily large values of $t-t_0$ and $t-t'$.

In the momentum representation the different degrees of freedom must have equal status in (11). If we also allow for the fact that the two terms in the square brackets in (11) have the same limit, thanks to the oscillatory nature of the evolution of phonon variables, we arrive at the following asymptotic behavior:

$$\langle v_k(t),\Gamma_{kj}(tt_0),v_j(t')\rangle\simeq\frac{\langle v_j(t_0)v_j(t')\rangle}{2N},\quad (t-t_0,t'\to\infty). \tag{12}$$

In particular, at $t_0=t'$ we can use (3) and (4) and easily transform (12) into

$$\langle n_k(t),n_j(t_0)\rangle\sim\frac{1}{N}\ \langle N_k\rangle\langle n_j\rangle,\quad (t-t_0\to\infty),$$

where n_k are the occupation numbers. The given asymptotic behavior expresses (in accordance with the physical meaning of the initial equations (9) and (10)) a weak but never decaying correlation between the energies of two arbitrary modes.

Equation (12) is a formally exact example of the long-lived behavior of higher-order correlators (and at the same time demonstrates in an obvious manner the undamped nature of the fluctuations of the matrix Γ and its correlation with phase diffusion). However, to describe the correlation of the phase diffusion ''rates'' for two time intervals that are far apart, it is not enough to use the three-time cumulants in Eqs. (11) and (12)—we need four-time cumulants like (6).

2. Let us describe the structure of such cumulants. We examine a cumulant (6) for an individual mode, assuming that the time intervals $r\equiv 1-3>0$ and $s\equiv 2-4>0$ are fixed and comparable to the relaxation time of this mode. Suppose that fluctuating friction on the interval ''r'' can be described in the approximation $K(1,3)\simeq n_1\,\exp[(i\omega-\gamma_1)r]$, where ω is the renormalized frequency of the mode, n_1 is the random occupancy of the mode on the interval ''r,'' and γ_1 is the random friction ''smoothed'' over this interval, with the same being true for the interval ''s.'' If the difference $1-2>0$ is much larger than the correlation time, the second term on the right-hand side of Eq. (6) can be neglected, while the first can be written as

$$\begin{aligned}\langle K(13),K(24)\rangle\simeq&\exp[i\omega(r+s)]\{\langle n_1,n_2\rangle\langle\exp(-r\gamma_1)\rangle\\&\times\langle\exp(-s\gamma_2)\rangle+\langle n_1\rangle\langle n_2\rangle\\&\times\langle\exp(-r\gamma_1),\exp(-s\gamma_2)\rangle\}.\end{aligned}\tag{13}$$

Let $r=0$. Then, with allowance for Eqs. (3) and (4), the cumulant (6) reduces to a three-time cumulant similar to (12) (at $j=k$). On the other hand, the only term that remains in braces in (13) is the first, so that we obtain

$$\langle K(13),K(24)\rangle\simeq\langle n_1,n_2\rangle\langle K(24)\rangle\sim\frac{1}{N}\,\langle K(24)\rangle,$$

which by its very meaning coincides with the right-hand side of Eq. (12). However, when the intervals r and s are finite, the order of (13) (and hence of (6)) is determined not by the factor $1/N$ but by the second term in (13), i.e., the correlation (not small, as we have seen) of the fluctuations of the friction coefficients γ_1 and γ_2:

$$\langle K(13),K(24)\rangle\sim\langle\exp(-r\gamma_1),\exp(-s\gamma_2)\rangle.$$

Since the infinity of the variance of linear friction fluctuations is due to the contribution to the integral in (7) of the infinitely remote past, and the same is true of intervals 1–3 and 2–4, Eq. (7) naturally leads (if divergences are ignored) to undamped correlations between γ_1 and γ_2 and the long-term dependence of the correlators (6) and (12) on interval 1–2. This statement can be verified, for example, by estimating the four-time cumulant of the friction integral. Under the sign of quadruple integration with respect to primed variables, this cumulant contains, among other things, terms whose structure is

$$\begin{aligned}&L(13)L(24)\langle R(11')R(33'),R(22')R(44')\rangle L(1'3')\\&\quad\times L(2'4')\langle Q(1')Q(3')\rangle\langle Q(2')Q(4')\rangle.\end{aligned}$$

We allow for the multiplicative nature of the fluctuations of matrix G and its constituents, including R. For $1',3'<2,4$, in the quadruple correlator of matrix R the most important variant of factor ''pairing'' under averaging is the one in which the contributions to $R(11')$ and $R(33')$ from time intervals between $1',3'$ and 2, 4 become paired with $R(22')$ and $R(44')$ (correspondingly or in a ''crisscross'' manner), while the multiplicative contributions to $R(11')$ and $R(33')$ from intervals 1,3–2,4 become paired with each other. In view of exponential instability, the latter leads to a result that remains undamped for any distance 1–2.

The long-lived nature of four-time cumulants must be preserved, even if we allow for the nonlinearity in the Langevin equation. Since the coefficient functions of each of the friction terms nonlinear in $Q(t)$ are also sensitive to the infinitely remote past, their sum can remain finite only due to the long-lived dynamical correlations of $Q(t)$ that are characteristic of the coefficient functions.

We emphasize that the concomitant of instability emerges, as does instability proper, already in the ''initial'' Gaussian approximation, where the random component of the quadrant $\partial U^2/\partial q^2$ of the matrix $B(t)$, which governs the evolution of the differential response, is assumed Gaussian. The point (and this is well-known fact) is that a linear system with randomly varying parameters is exponentially unstable in the sense of the second- and higher-order statistical moments. Hence, as shown by the above reasoning, the statistics of the fluctuations experiences a kind of ''phase transition'' from small (of order $1/N$) to large (N-independent) long-lived multitime non-Gaussian correlations. These correlations, which affect $B(t)$ and the differential response, must also affect low-frequency fluctuations of the degree of instability (the ''rate'' of mixing in the phase space).

6. CONCLUSION

To conclude, we now discuss how the ideas in this paper relate to other theories. From the standpoint of principles, the long-lived scale-invariant behavior of the phase-diffusion-rate fluctuations appears to be the only possible one: since the characteristic time scales for mode interaction are fixed by phase diffusion, there is no characteristic time scale remaining for the diffusion rate fluctuations. The presence of corresponding low-frequency noise is therefore consistent with the fact that the phase is random, which is required by kinetic models of energy exchange between modes.[8] Nor is there an inconsistency with ergodicity. Long ago Krylov demonstrated[22] that despite mixing, the frequencies of events are not necessarily averaged in time along a specific path of the system. Analysis[4] of measurements of, for example, four-time equilibrium cumulants shows how this can be the case in kinetic events. The ''long-lived nature'' is an internal property of a measured complex event and characterizes its duration, i.e., the difference between the extreme time arguments in the cumulant. Nothing in this property is inconsistent with the fact that for a fixed time difference, a real measurement carried out by time averaging (the same time shift in all the arguments) coincides with the result of ensemble averaging, as ergodicity requires.

Let us now touch on the differences and similarities that exist in relation to the concept of fundamental $1/f$ noise[9–13] without discussing the internal problems of Handel's theory (see, e.g. Ref. 22) and our approach (see above). The non-Gaussian statistics of noise and of the higher-order correlations play in important role in Handel's theory.[10,11] In our case, however, long-lived correlations are exclusively a property of the statistical ensemble. They are not backed by any long-term causal link accessible to observation along a specific phase path and there are no large characteristic ''coherence'' times.

On the contrary, under the slightest coarsening of the description, the exponential instability leads to loss of memory about the past and causality. But Handel's theory has long characteristic times, which are the reciprocal frequencies of the ultrasoft photons. Also, in that theory the scattering cross section fluctuations act as a source of noise (which, incidentally, is related to the surroundings). In our case the fluctuations of the diffusion coefficient,[4,19,20] the scattering cross section,[3] the mobility,[6] and, in general, the kinetic quantities are in an obvious and purely linguistic sense fictitious, since they are ascribed not to dynamical variables but to quantities that have no definite value outside the scope of ensemble averaging.[6)] They appear at the phenomenological level[19,20] or (as scattering cross section fluctuations) at the statistical-mechanics level when the description is coarsened by employing the collision integral technique.[3] In any case, they have a formally rigorous equivalent in the form of assertions concerning correlators of purely dynamical variables.

A final remark is in order. Generally, the fundamental nature of $1/f$ noise is not inconsistent with the possible important role of structural and surface defects in real phonon systems, since these defects affect the spectra, and the interaction of phonon modes and introduce additional relaxation mechanisms. What is important here is whether the defects are the primary source of $1/f$ noise or only one possible contributory channel, which may also be present in an ideal system.

[1)] A detailed discussion of the universal fluctuation–dissipation relations linking quaternary equilibrium cumulants and pairwise nonequilibrium cumulants and of higher-order relations can be found in Refs. 15, 16, and 17.

[2)] From the standpoint of fluctuation–dissipation relations, scattering fluctuations constitute a special case of dissipation fluctuations, since in an (initially equilibrium) system one must incorporate not only phonons but also the thermal and scattered fields.[13]

[3)] A similar statistical pattern emerges in the analysis of Brownian motion, whose statistics is non-Gaussian, and diffusion.

[4)] Note that the assumption that frictional fluctuations of a separate mode are large and highly irregular was first suggested by Handel.[2]

[5)] This statement follows from fluctuation–dispersion relationships, according to which linear frictional fluctuations must be accompanied by cubic friction,[17,21] and cubic frictional fluctuations must be accompanied by higher-order nonlinearities.

[6)] Incidentally, this is also true of scattering cross sections and probabilities in quantum mechanics and, all the more so, in quantum statistical mechanics.

[1] Y. Nogushi, Y. Teramachi, and T. Musha, Appl. Phys. Lett. **40**, 872 (1982).

[2] P. H. Handel, Solid-State Electron. **22**, 875 (1979).

[3] Yu. E. Kuzovlev, Zh. Éksp. Teor. Fiz. **94**, No. 12, 140 (1988) [Sov. Phys. JETP **67**, 2469 (1988)].

[4] G. N. Bochkov and Yu. E. Kuzovlev, Usp. Fiz. Nauk **141**, 151 (1983) [Sov. Phys. Usp. **26**, 829 (1983)].

[5] Yu. E. Kuzovlev, Phys. Lett. A **194**, 285 (1994).

[6] T. Musha, B. Gabor, and M. Shoji, Phys. Rev. Lett. **64**, 2394 (1990).

[7] A. J. Lichtenberg and M. A. Lieberman, *Regular and Stochastic Motion*, Springer, Berlin (1982).

[8] G. M. Zaslavskiĭ, *Chaos in Dynamical Systems*, Harwood Academic, New York (1985).

[9] P. H. Handel, Phys. Rev. Lett. **34**, 1492 (1975); **34**, 1495 (1975).

[10] P. H. Handel, Phys. Rev. A **22**, 745 (1980).

[11] P. H. Handel, Phys. Rev. A **26**, 3727 (1982).

[12] P. H. Handel, IEEE Trans. Electron Devices **41**, 2023 (1994).

[13] P. H. Handel, Phys. Status Solidi B **194**, 393 (1996).

[14] A. van der Ziel, Proc. IEEE (Russian translation) **76**, 29 (1988) [Mir Publishers, Moscow].

[15] G. N. Bochkov and Yu. E. Kuzovlev, Zh. Éksp. Teor. Fiz. **76**, 1071 (1979) [Sov. Phys. JETP **49**, 543 (1979)].

[16] G. N. Bochkov and Yu. E. Kuzovlev, Physica A **106**, 238 (1977).

[17] G. N. Bochkov and Yu. E. Kuzovlev, Zh. Éksp. Teor. Fiz. **72**, 238 (1977) [Sov. Phys. JETP **45**, 125 (1977)].

[18] G. N. Bochkov and Yu. E. Kuzovlev, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **24**, 855 (1981).

[19] Yu. E. Kuzovlev and G. N. Bochkov, *On the Nature and Statistical Characteristics of 1/f noise* [in Russian], Preprint NIRFI No. 157, Gorkiĭ (1982).

[20] Yu. E. Kuzovlev and G. N. Bochkov, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **26**, 310 (1983).

[21] G. N. Bochkov and Yu. E. Kuzovlev, Izv. Vyssh. Uchebn. Zaved. Radiofiz. **21**, 1467 (1978).

[22] N. S. Krylov, *Papers on the Foundations of Statistical Physics* [in Russian], Izd. Akad. Nauk SSSR, Moscow–Leningrad (1950).

[23] Th. M. Nieuwenhuizen, D. Frenkel, N. G. van Kampen, Phys. Rev. A **35**, 2750 (1987).

Translated by Eugene Yankovsky

Quasi-isentropic compression of liquid argon up to 500 GPa

V. D. Urlin, M. A. Mochalov, and O. L. Mikhaĭlova

Russian Federal Nuclear Center, All-Russia Research Institute of Experimental Physics,
607190 Sarov, Russia

(Submitted 8 July 1996)
Zh. Éksp. Teor. Fiz. **111**, 2099–2105 (June 1997)

The compressibility of liquid argon up to pressures ~500 GPa has been investigated experimentally. The argon was compressed by a cylindrical shell accelerated by the detonation products of an explosive. The density was recorded by the gamma-graphic method and the pressure was determined from the gas-dynamic calculations. Comparing the experimental and computational results showed that the compression process studied is isentropic to a quite high degree. The compression of liquid argon up to a density of 7.3 g/cm^3 did not show any clear anomalies associated with a structural transition or metallization. © *1997 American Institute of Physics.* [S1063-7761(97)01306-1]

1. INTRODUCTION

It is of interest to study the compressibility of argon at very high pressures (above 100 GPa) from the standpoint of observing phase transitions as well because there are absolutely no experimental data in this region. Argon has been studied under static conditions at $T=298$ K up to ~80 GPa.[1] The shock adiabat of liquid argon has been investigated up to ~90 GPa.[2,3]

A quasi-isentropic compression process is slower and more even than a shock process. The transformation of shock-wave into isentropic compression in an experimental apparatus substantially decreases the thermal heating of the material, which in turn increases the compression of the material. Furthermore, under these conditions the experimental material can be maintained at a high pressure for a longer time. This greatly expands the possibilities for investigating the states in which phase transitions are thought to occur.

In Ref. 4 no anomalies associated with a phase transition were found with quasi-isentropic compression of liquid Ar by a pressure of 60 GPa up to a density of 3.9 g/cm^3. Data indicating a phase transition have been obtained for xenon,[5] which was investigated by the same method up to a density of 13 g/cm^3. According to the results of Refs. 6 and 7, this transition can be attributed to a restructuring of a face-centered cubic (fcc) lattice into a hexagonal close-packed (hcp) structure. This transition is estimated, on the basis of the experimental data on isothermal compression,[8,9] to occur at a density of 8.37 g/cm^3.

Solid argon at atmospheric pressure possesses a fcc structure. According to calculations performed by McMahan,[10] the initial structure of solid argon can transform at pressure ~230 GPa into a hcp structure and metallization can occur at ~430 GPa. The same calculations show that metallization of the initial fcc Ar lattice is possible at ~550 GPa.

In the present work we obtained new data on the quasi-isentropic compressibility of liquid argon at pressures from 90 to 500 GPa and densities from 4 to 8 g/cm^3. No clear anomalies associated with the structural transitions predicted in Ref. 10 were observed in this region.

2. MEASUREMENT PROCEDURE

The experimental arrangement is similar to that described in Ref. 4. A high pressure produced by the detonation products of a powerful explosive surrounding the experimental apparatus pushes a cylindrical copper or tungsten alloy shell toward the axis, initially forming a relatively weak converging shock wave in the experimental material (argon) located inside the shell. Quasi-isentropic compression is produced at subsequent times by the circulation of shock waves and the continuous compression of the argon by the shell converging toward the axis. The shell is decelerated by the increasing pressure in the experimental material. At some time the shell stops and then moves in the opposite direction. Maximum compression occurs at the moment the shell stops.

The trajectory of the shell, according to which the size of the cavity containing liquid Ar at the moment of the shell stops is estimated, was recorded in a series of subsequent experiments by means of a powerful gamma-graphic system with a short exposure time.[11] Analysis of the gas-dynamic calculations showed that, even at the highest pressure in the range investigated, unloading of the experimental volume does not occur during the compression of the gas. In this case the average density of the compressed argon was estimated from the expression

$$\rho=\rho_0\left(\frac{R_0}{R_{\min}}\right)^2.$$

Here $\rho_0=1.4$ g/cm^3 is the initial Ar density at $T=87$ K and R_0 and $R_{\min}$ are, respectively, the initial radius of the shell and the radius of the shell at the moment the x-ray picture of the stopped shell is taken. The accuracy of the density estimate is determined by the accuracy with which the shell radius is measured from the x-ray photograph. The higher the quality of the x-ray photograph and the higher the contrast of the boundary between the shell and the liquid argon, the more reliable the measurement. To this end and to obtain high degrees of compression as well, in the experiments copper and tungsten alloy shells were used for argon densities $\lesssim 5$ g/cm^3 and >5 g/cm^3, respectively.

1063-7761/97/061145-04$10.00

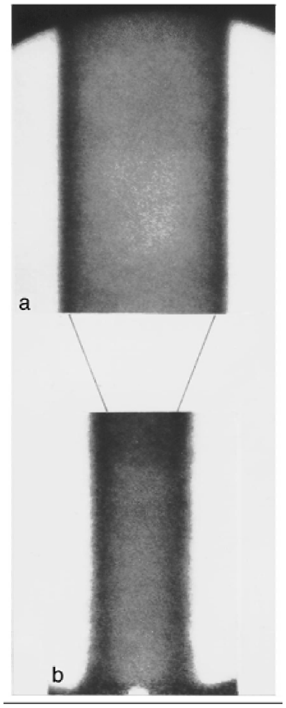


FIG. 1. Shadow x-ray image of a tungsten shell with compressed argon density ~6 g/cm^3: a — shell in the initial state; b — shell at the moment of maximum compression.

The quality of x-ray pictures is determined by different factors, one of which is the blurriness of the image. Blurriness depends on the geometric size of the focus of the radiation source. It is caused by the motion of the shell at the time when the x-ray photograph is taken and by the effect of the amplifying screens which convert x-rays into visible-range radiation. The total blurriness produced by distorting factors of all types will be minimum at the moment of maximum compression of the gas, when the shell has stopped and there is no dynamical blurriness. Distortion effects can not only degrade image quality. As a result of their presence, the position of the shell boundary recorded on the film may not be the true position.

The total effect of the distorting factors was estimated in special experiments in which shell images simulating real experimental structures in a compressed state were recorded. The main experiment differs from the model experiment in that the mass distribution over the cross section of the compressed cavity is more complicated. Analysis of the results obtained shows that the displacements of the visible boundaries of the recorded shells relative to their true sizes does not occur. Estimates of the sizes of the model images showed that the methodological accuracy of the measurements equals $\pm 4\%$ for diameters $\geqslant 8$ mm. This figure was taken as the accuracy in our experiments.

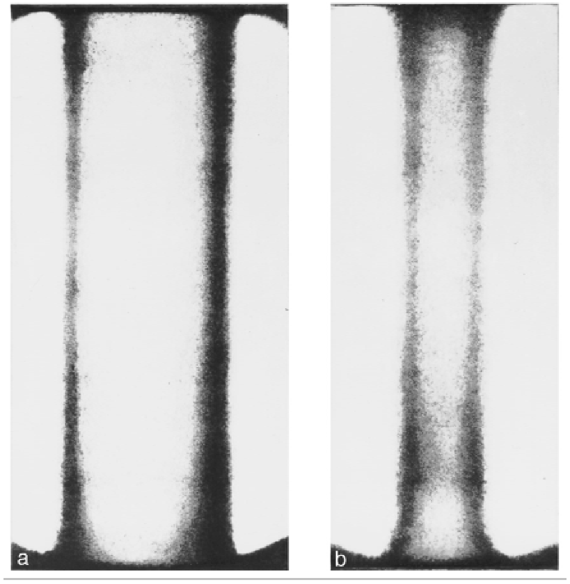


FIG. 2. Same as in Fig. 1 but with the maximum achieved argon density ~7.3 g/cm^3: a — shell in the initial state; b — shell in the compressed state.

Figure 1 displays a photograph of the shadow image from a tungsten alloy in an experiment with compressed argon density ~6 g/cm^3. Figure 2 shows a similar photograph but for a maximum density ~7.3 g/cm^3. One can see from the figures that the compression process in the structures used is characterized by good symmetry and satisfactory contrast of the inner boundary of the shell with liquid argon. The boundary remains rectilinear and symmetric over the entire length of the compressed part of the cylinder (~10 cm) even with the highest degrees of compression achieved in the present work. This also shows, qualitatively, that no unloading occurs from the ends of the cylinder.

The experimental values for the trajectory of the inner cavity of a tungsten shell are presented in Fig. 3. The trajectory computed taking account of the equations of state of all materials present in the structure are also shown there. The results of the density measurements performed in three series of experiments at pressures of 93, 247, and 480 GPa are presented in Table I.

3. EQUATIONS OF STATE OF SOLID AND LIQUID PHASES OF ARGON

The form of the multiphase equation of state and the method for finding the parameters in this equation are substantiated in Refs. 12, 3, and 4.

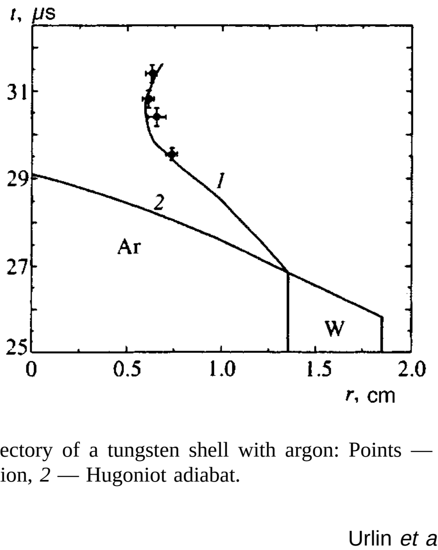


FIG. 3. Trajectory of a tungsten shell with argon: Points — experiment; *1* — calculation, *2* — Hugoniot adiabat.

TABLE I.

		Experiment		Calculation		
No.	R_0, cm	$R_{\min}$, cm	ρ, g/cm^3	$R_{\min}$, cm	ρ, r/cm^3	P, GPa
1	1.5	0.81±0.03	4.8±0.3	0.82	4.68	93
2	1.5	0.72±0.02	6.1±0.3	0.716	6.15	247
3	1.35	0.59±0.02	7.3±0.5	0.588	7.39	480

TABLE II. Parameters in the equation of state of argon.

ρ_k,	g/cm^3	1.77	l	3
n		2	b	0.694
a_1	GPa	-105.73	δ_0	0.8
a_2,	GPa	346.72	r_1	1
a_3	GPa	-383.72	C_1	1.7925
a_4	GPa	142.73	C_2	-0.0589
Θ_0,	K·s/km	78.65	W_0, eV	14.4
T_0	K	83.8	Γ	1

The expressions for the free energy in the solid and liquid phases are[1)]

$$F_S=E_x(\rho)+RT[3\ln(1-e^{-\Theta/T})-D(\Theta/T)], \tag{1}$$

$$F_L=E_x(\rho)+3RT\ln(\Theta/T)-3RT\ln\alpha(\rho,T), \tag{2}$$

where T and ρ are the temperature and density, R is the gas constant, and

$$D(x)=3x^{-3}\int_0^x\frac{x^3dx}{e^x-1}$$

is the Debye function. The elastic interaction curve is approximated by the function

$$E_x=\frac{3}{\rho_k}\sum_i a_i(\delta^{i/3}-1)/i,$$

where $\delta=\rho/\rho_k$, ρ_k is the density at $P_x=\rho^2dE_x/d\rho=0$, and a_i are empirical constants. The Debye temperature is given by the relation

$$\Theta=\Theta_0\delta^{1/3}\sqrt{C_x^2-2nP_x/3\rho},\quad C_x^2=dP_x/d\rho,$$

where Θ_0 and n are empirical parameters. The parameter

$$\alpha=(1+z)^{-1/2}\exp\{b-f(\delta)T_0/T\}$$

is a measure of the deviation of the thermal and elastic properties of the liquid from the solid phase. Here $z=lRT/(C_x^2-2nP_x/3\rho)$, b and l are constants, T_0 is the melting temperature at $P=0$, and $f(\delta)$ is an empirical function of the density. For definiteness, the latter function was taken in the form

$$f(\delta)=C_1\frac{(\delta/\delta_0)^{r_1}-1}{r_1}+C_2,$$

where C_1, C_2, and r_1 are empirical constants, $\delta_0=\rho_0/\rho_k$, and ρ_0 is the density of the liquid phase at T_0.

For $T>1$ eV in Ar the contribution of the electronic component must be taken into account. For insulators this component has the form

$$F_{el}=-\frac{4kT}{\rho}\sqrt{n_pn_n}\left(\frac{2\pi m^*kT}{h^2}\right)^{1.5}\exp\left(-\frac{W}{2kT}\right), \tag{3}$$

where m^* is the geometric mean of the electron and hole effective masses, and n_p and n_n are the orbital degeneracy in the bands. Just as in Ref. 3, in Eq. (3) the density dependence of the energy gap was taken to be of the form

$$W=W_0\delta^{-\Gamma},\quad m^*=m_0\delta^{2/3-\Gamma},$$

where W_0 is the gap width at normal density, m_0 is the free-electron mass, and Γ is an empirical constant whose value is found so as to obtain the best description of the shock adiabat at high temperatures.

The numerical values of the parameters are presented in Table II.

4. COMPARISON OF THE COMPUTATIONAL AND EXPERIMENTAL RESULTS

To determine the pressure corresponding to the experimentally measured argon density, gas-dynamic calculations were performed using the equations of state of the materials present in the structure and the equation of state given in Ref. 3 for Ar. The density in the cavity of the compressed material and the trajectory of the inner boundary of the shell were also determined from the calculations. The average density of the compressed argon was estimated if the experimental and computed trajectories agreed. The computational results for a series of experiments with the maximum achievable pressure are displayed in Figs. 3 and 4. As one can see from Fig. 3, the computed trajectory of the inner shell agrees satisfactorily with experiment. Figure 4 shows the argon pressure and density distributions in the cavity at the moment of maximum compression. The dot-and-dash lines in the figure mark the average density and the corresponding pressure. It is evident from this figure that the deviation of the density and pressure from their average values equal ~5% and ~11%, respectively.

Figure 5 shows in the coordinates $P(\rho)$ the experimental results for the quasi-isentropic compression of argon to ~500 GPa. These results are compared with the computed isentropes with the entropy S/R corresponding to the experiment. The computed isentropes agree well with the experimental data. The 300 K isotherm and the shock adiabat of

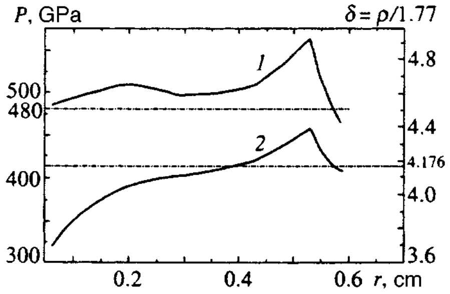


FIG. 4. Radial pressure (*1*) and density (*2*) distributions at the moment of maximum compression of argon.

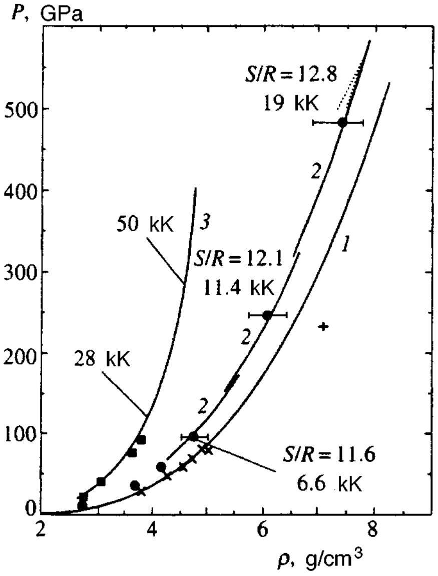


FIG. 5. Isentropic compression of liquid argon. Experiment: ● — isentropic compression, this work; × — isotherm;[1] ■ — Hugoniot adiabat.[2] Calculation: *1* — $T=300$ K isotherm; *2* — isentropes; *3* — Hugoniot adiabat; + — fcc–hcp.[10]

liquid Ar are also presented here for comparison. One can see that the isentropes are substantially ''softer'' than the shock adiabat and approach the elastic curve P_x. Figure 5 (dotted curve) shows the radial distribution of the state of the matter inside the shell at the moment the shell stops. One can see that more than two thirds of the mass of the material has the minimum entropy, whose value is indicated in the figure. The figure also displays for comparison the temperatures on the isentrope and on the shock adiabat, showing the difference in the degree of heating in these two processes.

TABLE III. Computed values at the moment the shell stops.

No.	S/R	T, K	P_x, GPa	P, GPa	C, km/s
1	11.6	6600	64	93	8.4
2	12.1	11400	185	247	11.6
3	12.8	19000	360	480	14.4

The entropy, temperature, vapor and total pressures, and the velocity of sound for the computed densities at three experimental points are summarized in Table III. One can see from the table that under isentropic compression the thermal pressure equals 23–31% of the total pressure. The values presented in this table for the sound velocity in argon at the moment of maximum compression in each experiment make it possible to estimate the upper limit of the length of the end region encompassed in the compressed cylinder by the unloading. Estimates performed for a specific structure, where a pressure ~500 GPa was achieved, showed that with a ~10 cm long compression tube unloading encompasses not more than 1.5 cm of the compressed volume from each end. Therefore the region of unloaded argon is not less than 7 cm long.

In summary, in the present work the isentrope of argon was continued up to pressures ~500 GPa, where there are no other experimental data. These investigations, just as in Ref. 4, did not show any visible anomalies in the compressibility of liquid argon.

---

1)In the present work the contribution of the zero-point vibrations is taken into account in the term $E_x(\rho)$ in the equation of state.

[1] M. Ross, H. K. Mao, P. M. Bell, and J. A. Xu, J. Chem. Phys. **85**, 1028 (1986).
[2] M. Ross, W. Nellis, and A. Mitchell, Chem. Phys. Lett. **68**, 532 (1979).
[3] F. V. Grigor'ev, S. B. Kormer, O. L. Mikhaĭlova *et al.*, Zh. Éksp. Teor. Fiz. **88**, 1271 (1985) [Sov. Phys. JETP **61**, 751 (1985).]
[4] I. A. Adamskaya, F. V. Grigor'ev, O. L. Mikhaĭlova *et al.*, Zh. Éksp. Teor. Fiz. **93**, 647 (1987) [Sov. Phys. JETP **66**, 366 (1987).
[5] V. D. Urlin, M. A. Mochalov, and O. L. Mikhaĭlova, High Press. Res. **8**, 595 (1992).
[6] A. P. Jephcoat, H. K. Mao, L. Finger *et al.*, Phys. Rev. Lett. **59**, 2670 (1987).
[7] R. Reichlin, K. Brister, A. McMahan *et al.*, Phys. Rev. Lett. **62**, 669 (1989).
[8] A. N. Zisman, I. V. Aleksandrov, and S. M. Stishov, Phys. Rev. **32**, 484 (1985).
[9] K. A. Goettel, J. H. Eggert, and I. F. Silvera, Phys. Rev. Lett. **62**, 665 (1989).
[10] A. K. McMahan, Phys. Rev. B **33**, 5344 (1986).
[11] A. I. Pavlovskiĭ, G. D. Kuleshov, G. V. Sklizkov *et al.*, Dokl. Akad. Nauk SSSR **160**, 68 (1965) [Sov. Phys. Dokl. **10**, 30 (1965)].
[12] V. D. Urlin, Zh. Éksp. Teor. Fiz. **49**, 485 (1965) [Sov. Phys. JETP **22**, 341 (1966)].

Translated by M. E. Alferieff

The kinetics of low-temperature electron–phonon relaxation in a metallic film following instantaneous heating of the electrons

A. I. Bezuglyĭ and V. A. Shklovskiĭ

Khar'kov Physicotechnical Institute, National Scientific Center, 310108 Khar'kov, Ukraine

(Submitted 11 September 1996; resubmitted 15 December 1996)
Zh. Éksp. Teor. Fiz. **111**, 2106–2133 (June 1997)

The theoretical analysis of experiments on pulsed laser irradiation of metallic films sputtered on insulating supports is usually based on semiphenomenological dynamical equations for the electron and phonon temperatures, an approach that ignores the nonuniformity and the nonthermal nature of the phonon distribution function. In this paper we discuss a microscopic model that describes the dynamics of the electron–phonon system in terms of kinetic equations for the electron and phonon distribution functions. Such a model provides a microscopic picture of the nonlinear energy relaxation of the electron–phonon system of a rapidly heated film. We find that in a relatively thick film the energy relaxation of electrons consists of three stages: the emission of nonequilibrium phonons by ''hot'' electrons, the thermalization of electrons and phonons due to phonon reabsorption, and finally the cooling of the thermalized electron–phonon system as a result of phonon exchange between film and substrate. In thin films, where there is no reabsorption of nonequilibrium phonons, the energy relaxation consists of only one stage, the first. The relaxation dynamics of an experimentally observable quantity, the phonon contribution to the electrical conductivity of the cooling film, is directly related to the dynamics of the electron temperature, which makes it possible to use the data of experiments on the relaxation of voltage across films to establish the electron–phonon and phonon–electron collision times and the average time of phonon escape from film to substrate. © *1997 American Institute of Physics.* [S1063-7761(97)01406-6]

1. INTRODUCTION

Lately there has been an upsurge of interest in experimental studies of relaxation processes that emerge in the electron–phonon system of a metallic film subjected to an ultrashort ($\sim$100 fs) laser pulse (see, e.g., Refs. 1, 2 and 3). Usually, in describing experiments involving the heating and relaxation of the electron subsystem, theoreticians use the so-called two-temperature approximation suggested by Kaganov *et al.*[4] in 1956. In this approximation it is assumed that the electrons and phonons are thermalized and can be characterized by temperatures T_e and T_p, respectively. The difference in these temperatures ($T_e > T_p$) is fixed by the heat flux from electrons to phonons. What is important is that the magnitude of this flux is determined only by the parameters of the electron–phonon interaction in the metal and does not depend on the real conditions of heat removal in which the sample is observed.

In the theoretical work that followed, Shklovskiĭ[5,6] and Maslov and Shklovskiĭ[7] gave a detailed analysis at the kinetic level of the various aspects of the theory of hot electrons in thin metallic films at low temperatures not discussed in Refs. 4 and 8. The main goal of their work was to establish the physical conditions that heat removal from DC-heated films must satisfy and in which the experimental observation of the nonlinear corrections to Ohm's law predicted earlier[8,6] becomes possible. It was found that the realization of the two-temperature approximation[4] in a stationary experiment imposes severe restrictions on the film thickness, the acoustic transparency of the film/substrate interface,[5] and the width of the film being studied (see the concluding section in Ref. 6).

Note that later these conditions were met, we believe, most accurately in the experimental work of Bergmann *et al.*[9] and Gershenzon *et al.*[10] Most experiments, however, have been carried not in the two-temperature regime (which in Ref. 5 was called the electron overheating regime) but in the Joule heating regime, where in view of the insufficient acoustic matching of a fairly thick film and the substrate the electrons and phonons in the film may be assumed, to a high degree of accuracy, to have equal temperatures. The value of this temperature depends on the acoustic characteristics of the metal and the substrate, according to Little's well-known theory of acoustic mismatch.[11] In Ref. 5 a strict theoretical criterion was derived for observing the size effect in heat removal from metallic films at low temperatures (see below). These results were generalized in Ref. 6 to the case of an arbitrary electron spectrum and an arbitrary frequency dependence of the Eliashberg function $\alpha^2F(\omega)$. A similar generalization allowed Allen[12] to express, in the high-temperature limit, the rate of electron cooling in terms of the electron–phonon coupling constant λ.

Note, however, that in our case of low temperatures and low-intensity heating (in which the typical phonon frequency is low compared to the Debye frequency ω_D), the two-temperature approximation is not quite justified, since in this temperature range the phonon subsystem becomes thermalized primarily because of collisions with electrons, and under thermalized conditions the two subsystems must have the same temperature.

In another semiphenomenological model, suggested by

Perrin and Budd[13] to describe experiments on the generation of heat pulses, the phonon distribution function is assumed to differ from the Bose distribution function and is found by requiring that the emission of phonons by the heated electrons balance the escape of phonons to the substrate. However, in this model the phonon distribution function is assumed uniform over the film's thickness, since phonon escape can be described phenomenologically in terms of the average time of phonon escape from film to substrate. As a result, this approach requires microscopic substantiation.

A microscopic kinetic study carried out by one of the authors of the present paper for the case of stationary heating[5] makes it possible (with a well-founded assumption about the uniformity of the electron temperature over the film thickness) to examine both the nonthermal nature and the spatial nonuniformity of the phonon distribution function in a metallic film. The present paper generalizes this microscopic approach to the case of nonstationary film heating (Sec. 2).

The concept of a time-dependent electron temperature drastically simplifies the analysis of the two coupled kinetic equations for the electron and phonon distribution functions with a very general assumption concerning the power-law dependence of the Eliashberg function on frequency. The latter makes it possible to account phenomenologically for the possible renormalization of the electron–phonon interaction by impurities (see, e.g., Ref. 14), and thus to compare our results with experimental data even when such renormalization is substantial.

In Sec. 3 we discuss the nonlinear kinetics of the energy relaxation of the electron subsystem after instantaneous heating by, say, an ultrashort laser pulse. It turns out that the dynamics of electron cooling differs drastically for effectively thick and effectively thin films, i.e., when nonequilibrium phonons are reabsorbed in the film, and when such phonons escape to the substrate without being reabsorbed, respectively. We show that electrons in effectively thick films cool in three stages: the emission of nonequilibrium phonons by ''hot'' electrons, the thermalization of the interacting electrons and phonons, and the subsequent cooling of the thermalized electron–phonon system due to phonon exchange between film and substrate. The durations of the first two stages coincide, in order of magnitude, with the electron–phonon collision times at the initial electron temperature and at the thermalization temperature of these electrons, respectively. The duration of the last stage is determined by the characteristic time of phonon escape to the substrate. In thin films there is no reabsorption of nonequilibrium phonons, so that the characteristic electron cooling time in a thin film is the electron–phonon collision time at the lowest temperature, i.e., the temperature of the heat bath.

Since direct measurements of the electron temperature in high-speed processes is clearly fraught with difficulty, the phonon contribution to the electrical resistance of the film can serve as a possible ''indicator'' of electron temperature. In Sec. 4, we therefore examine the dynamics of film conductivity under pulsed heating. Note that comparison of the results of Sec. 4 with the appropriate experimental data can provide useful information about the characteristic times of relaxation of the electron–phonon system in contaminated films (see Sec. 5). Section 6 provides a brief description of the special features of our approach.

2. KINETIC APPROACH TO THE RELAXATION OF ELECTRON ENERGY IN A METALLIC FILM. MODEL AND GENERAL RELATIONSHIPS

We consider a simple microscopic model, in which the electron dispersion law is isotropic and quadratic, i.e., the energy of an electron with quasimomentum $\mathbf{p}$ (or wave vector $\mathbf{k}=\mathbf{p}/\hbar$) is $\varepsilon_{\mathbf{p}}=\mathbf{p}^2/2m$, where m is the effective mass. We also assume that phonons have a single acoustic branch with longitudinal polarization. Since we are interested in a range of temperatures that are low compared to the Debye temperature Θ_D, we assume the dispersion law for longitudinal phonons to be linear, $\omega_q=sq$, where s is the velocity of longitudinal sound, and $q=|\mathbf{q}|$ is the absolute value of the phonon wave vector. Using the standard notation,[15] we can write the kinetic equations for the electron and phonon distribution functions in a thin metallic film as follows:

$$\frac{\partial f_{\mathbf{p}}}{\partial t}+e\mathbf{E}\,\frac{\partial f_{\mathbf{p}}}{\partial \mathbf{p}}+\frac{p_z}{m}\,\frac{\partial f_{\mathbf{p}}}{\partial z}=I_i(f_{\mathbf{p}})+I_{ep}(f_{\mathbf{p}},N_{\mathbf{q}}), \tag{2.1}$$

$$\frac{\partial N_{\mathbf{q}}}{\partial t}+s_z\,\frac{\partial N_{\mathbf{q}}}{\partial z}=I_{pe}(N_{\mathbf{q}},f_{\mathbf{p}}), \tag{2.2}$$

where it is assumed that the z axis is perpendicular to the film's surface. The electron–phonon and phonon–electron collision integrals have the well-known form (without allowing for transfer processes)

$$\begin{aligned} I_{ep}(f_{\mathbf{k}},N_{\mathbf{q}})=\int \frac{d^3q}{(2\pi)^3}\, w(q)\{ & f_{\mathbf{k}+\mathbf{q}}(1-f_{\mathbf{k}})[(N_{\mathbf{q}}+1) \\ & \times\delta(\varepsilon_{\mathbf{k}+\mathbf{q}}-\varepsilon_{\mathbf{k}}-\Omega)+N_{-\mathbf{q}}\delta(\varepsilon_{\mathbf{k}+\mathbf{q}}-\varepsilon_{\mathbf{k}}+\Omega)] \\ & -f_{\mathbf{k}}(1-f_{\mathbf{k}+\mathbf{q}})[N_{\mathbf{q}}\delta(\varepsilon_{\mathbf{k}+\mathbf{q}}-\varepsilon_{\mathbf{k}}-\Omega) \\ & +(N_{-\mathbf{q}}+1)\,\delta(\varepsilon_{\mathbf{k}+\mathbf{q}}-\varepsilon_{\mathbf{k}}+\Omega)]\}, \end{aligned} \tag{2.3}$$

$$\begin{aligned} I_{pe}(N_{\mathbf{q}},f_{\mathbf{k}})=\int \frac{2\,d^3k}{(2\pi)^3}\, w(q)[& f_{\mathbf{k}+\mathbf{q}}(1-f_{\mathbf{k}})(N_{\mathbf{q}}+1) \\ & -f_{\mathbf{k}}(1-f_{\mathbf{k}+\mathbf{q}})N_{\mathbf{q}}]\delta(\varepsilon_{\mathbf{k}+\mathbf{q}}-\varepsilon_{\mathbf{k}}-\Omega). \end{aligned} \tag{2.4}$$

In the above formulas, e is the electron charge, $\mathbf{E}(t)$ is the strength of the electric field generated in the film when a weak measuring current of average density j_0 is sent through the film, and $\Omega=\hbar\omega_q$. The function $w(q)$ determines the intensity of the electron–phonon interaction (see below). Depending on what is convenient, in Eqs. (2.3) and (2.4) and in what follows electron states are characterized either by a wave vector $\mathbf{k}$ or by a quasimomentum $\mathbf{p}$ ($\mathbf{p}=\hbar\mathbf{k}$).

Elastic electron–impurity collisions are described by the collision integral

$$I_i(f_{\mathbf{p}})=n_i\int \frac{d^3p'}{(2\pi\hbar)^3}\, u(\mathbf{p}-\mathbf{p}')(f_{\mathbf{p}}-f_{\mathbf{p}'})\,\delta(\varepsilon_{\mathbf{p}}-\varepsilon_{\mathbf{p}'}), \tag{2.5}$$

where n_i is the impurity concentration, and the function $u(\mathbf{p}-\mathbf{p}')$ is proportional to the square of the absolute value of the Fourier transform of the impurity potential. When electron–impurity collisions dominate, the distribution is weakly anisotropic and can be approximately written as

$$f_{\mathbf{p}}=f_0(\varepsilon_{\mathbf{p}})+\mathbf{p}\cdot\mathbf{f}_1(\varepsilon_{\mathbf{p}}), \tag{2.6}$$

where it is assumed that the second term is small compared to the first. In this approximation, the collision integral (25) can be reduced to what is known as the τ-approximation:

$$I_i(f_{\mathbf{p}})=-\tau_i^{-1}\mathbf{p}\cdot\mathbf{f}_1(\varepsilon_{\mathbf{p}}), \tag{2.7}$$

where the electron–impurity collision rate is given by

$$\tau_i^{-1}=\frac{1}{4}\ n_i N(0)\int_0^\pi u(\varphi)(1-\cos\varphi)\sin\varphi\ d\varphi. \tag{2.8}$$

The electron density of states $N(0)$ is $mp_F/\pi^2\hbar^3$, and φ denotes the angle between the vectors $\mathbf{p}$ and $\mathbf{p}'$.

In the deformation potential model, the function $w(q)$, which is proportional to the square of the matrix element of the electron–phonon interaction, is a linear function of the phonon wave vector:

$$w(q)=w_0(q)=\frac{\pi\mu^2\omega_q}{\rho s^2}, \tag{2.9}$$

where the deformation potential constant is of order of the Fermi energy, $\mu\sim\varepsilon_F=p_F^2/2m$, and ρ is the film's density. To be able to phenomenologically take into account the renormalization of the electron–phonon interaction in contaminated metals and compare our results with experimental data, throughout the paper we consider the more general case of an arbitrary power-law dependence of the matrix element of the electron–phonon interaction on the phonon wave vector and write (as is done in the experimental work of Bergmann *et al.*[9] and Gershenzon *et al.*[16])

$$w(q)=\frac{\pi\mu_1^2\Omega^{1+r}}{\rho s^2\hbar}=C_r q^{1+r}, \tag{2.9$'$}$$

where r is not necessarily an integer. The representation (2.9′) for the function $w(q)$ allows going over to the deformation potential approximation in the formulas below by a simple substitution: $r=0$ and $\mu_1=\mu$. It is well known that Eq. (2.9) corresponds to a quadratic dependence of the Eliashberg function on the phonon frequency. It can be shown that the expression (2.9′) leads to a power-law frequency dependence of the Eliashberg function of the form

$$\alpha^2F(\omega_q)=\frac{N(0)\pi\mu_1^2\Omega^{2+r}}{4\rho s^2(sp_F)^2}\propto\omega_q^{2+r}. \tag{2.10}$$

Here the characteristic electron–phonon collision time as a function of temperature has the form $\tau_{ep}\propto T^{-(3+r)}$.

Note that tunneling experiments often reveal a linear behavior of the Eliashberg function $\alpha^2F(\omega_q)$ at low phonon frequencies,[17] while measurements involving thin films[9,16] and massive samples[18] quite often reveal a quadratic temperature dependence of τ_{ep}^{-1}, which corresponds to $r=-1$. At the same time, since other values of r were also obtained in experiments and since theoretical calculations yield values of r ranging from -1 to 1, we assume that r is an arbitrary number (not smaller than -1).

The feasibility of a quantitative description of the dynamics of the electron–phonon system of a metallic film is to a large extent related to the adopted simplifying assumption about the Fermi behavior of the isotropic part of the electron–phonon distribution function, i.e.,

$$f_0(\varepsilon_p)=\left\{1+\exp\left[\frac{\varepsilon_p-\varepsilon_F}{k_BT_e(t)}\right]\right\}^{-1}, \tag{2.11}$$

where $T_e(t)$ is the time-dependent electron temperature.

Note that the concept of an electron temperature, which is equivalent to the common assumption of instantaneous thermalization of the electron subsystem, cannot always be rigorously justified. For instance, at extremely low temperatures $T_e<T^*$ $(T^*\sim\Theta_D^2/\varepsilon_F)$, where electron–electron collisions predominate over electron–phonon collisions, the electron distribution function becomes thermalized in the course of the characteristic electron–electron interaction time τ_{ee}. In ordinary relatively pure metals, $T^*\sim1$ K, while in specially prepared impure films, where the electron–electron interaction is enhanced due to effects of weak localization, T^* may be of order 10 K. At $T_e>T^*$ (but $T_e\ll\Theta_D$), electron thermalization in relatively thick films occurs not because of direct electron–electron interaction but because of indirect interaction, via phonon exchange.

In Ref. 7 it was shown that an electron distribution function closely resembling the Fermi function can also be formed in relatively thin films (nonequilibrium phonons leave the film for the substrate without reabsorbing electrons) as a result of phonon emission by ''hot'' electrons alone. In both cases, the electron–phonon interaction time τ_{ep} is the characteristic electron thermalization time. Thus, at relatively low frequencies of external perturbations ($\omega\ll\tau_{ee}^{-1}$ for $T_e<T^*$ and $\omega\ll\tau_{ep}^{-1}$ at $T_e>T^*$), the isotropic part of the electron distribution function can be assumed to be the Fermi function. In examining the case of high-frequency perturbations acting on the electron subsystem, Eq. (2.11) should be considered an approximation at the first stage, i.e., an approximation that must be refined, possibly, by the approach suggested by Fann *et al.*,[19] who examined the dynamics of both $T_e(t)$ and the thermal correction to the Fermi electron distribution function.

We also note that in optically thick films, the uniformity of the electron temperature over the film thickness is ensured by the rapid departure of electrons from the skin layer[20] and the high electron thermal conductivity compared to the phonon thermal conductivity.[5] Due to the additional diffusion decrease in the density of ''hot'' electrons, the rate of thermalization of the electron subsystem grows considerably, with the result that in optically thick films the approximation of instantaneous thermalization provides good agreement between theory and experiment.[20]

We now calculate the part of the electron distribution function that is anisotropic in the momenta. To derive an equation for $\mathbf{f}_1(\varepsilon_{\mathbf{p}})$, we must plug (2.6) into the kinetic equation (2.1), multiply the result by $\mathbf{p}$, and integrate the product over the angles of $\mathbf{p}$. If we allow for the fact that

$$\int \frac{dO_{\mathbf{p}}}{4\pi}\, p_i p_k = \frac{1}{3}\, p^2 \delta_{ik}, \tag{2.12}$$

where p_i and p_k are the Cartesian components of $\mathbf{p}$, $dO_{\mathbf{p}}$ is an element of solid angle in momentum space, and δ_{ik} is the Kronecker delta, we arrive at the following equation for $\mathbf{f}_1$:

$$\frac{\partial \mathbf{f}_1}{\partial t} + \frac{e}{m}\,\mathbf{E}(t)\,\frac{\partial f_0}{\partial \varepsilon_{\mathbf{p}}} = -\frac{\mathbf{f}_1}{\tau_i} + \frac{3}{p^2}\int \frac{dO_{\mathbf{p}}}{4\pi}\,\mathbf{p} I_{ep}^{(1)} \times\{f_0,\mathbf{f}_1,N_{\mathbf{q}}\}. \tag{2.13}$$

Here $I_{ep}^{(1)}$ is the electron–phonon collision integral linearized in $\mathbf{f}_1$:

$$I_{ep}^{(1)}\{f_0,\mathbf{f}_1 N_{\mathbf{q}}\} = \int \frac{d^3k'}{(2\pi)^3}\,[\mathbf{k}'\cdot\mathbf{f}_1(\varepsilon_{\mathbf{k}'})B_{\mathbf{k}',\mathbf{k}} - \mathbf{k}\cdot\mathbf{f}_1(\varepsilon_{\mathbf{k}})B_{\mathbf{k},\mathbf{k}'}], \tag{2.14}$$

where $\mathbf{k}'=\mathbf{k}+\mathbf{q}$, and

$$\begin{aligned} B_{\mathbf{k},\mathbf{k}'} = \hbar w(|\mathbf{k}-\mathbf{k}'|)\{&[1-f_0(\varepsilon_{\mathbf{k}'})+N_{\mathbf{k}-\mathbf{k}'}] \\ &\times \delta(\varepsilon_{\mathbf{k}'}-\varepsilon_{\mathbf{k}}+\Omega) \\ &+[f_0(\varepsilon_{\mathbf{k}'})+N_{\mathbf{k}'-\mathbf{k}}] \\ &\times \delta(\varepsilon_{\mathbf{k}'}-\varepsilon_{\mathbf{k}}-\Omega)\}. \end{aligned} \tag{2.15}$$

We assume that the characteristic frequencies ω of the external parameters (the current flowing through the film, the laser beam intensity, etc.) are low compared to the elastic electron scattering rate τ_i^{-1}, which in turn is lower than $\varepsilon_F/\hbar$. (In the event of pulsed heating we consider only the cool-off stages, i.e., when external heating has ceased.) In this case, the time derivative in (2.13) and the contribution due to scattering by phonons can be taken into account perturbatively by writing

$$\mathbf{f}_1 = \mathbf{f}_1^{(0)} + \mathbf{f}_1^{(1)}. \tag{2.16}$$

Plugging this into (2.13) yields

$$\mathbf{f}_1^{(0)} = -\frac{e\tau_i}{m}\,\mathbf{E}(t)\,\frac{\partial f_0}{\partial \varepsilon_{\mathbf{p}}}, \tag{2.17}$$

$$\mathbf{f}_1^{(1)} = \frac{3\tau_i}{p^2}\int \frac{dO_{\mathbf{p}}}{4\pi}\,\mathbf{p} I_{ep}^{(1)}\{f_0,\mathbf{f}_1^{(0)},N_{\mathbf{q}}\} - \tau_i\,\frac{\partial \mathbf{f}_1^{(0)}}{\partial t}. \tag{2.18}$$

Equation (2.18) leads to the following relationship between $f_1^{(0)}$ and $f_1^{(1)}$:

$$f_1^{(1)} \sim \max\left(\omega\tau_i,\, \frac{\tau_i}{\tau_{ep}}\right) f_1^{(0)} \ll f_1^{(0)}. \tag{2.19}$$

Equation (2.17) can be used to estimate $\mathbf{p}\cdot\mathbf{f}_1$:

$$p_F f_1^{(0)} \sim \frac{\tau_i e E p_F}{m k_B T_e} \sim \frac{l_i e E}{k_B T_e} \ll 1, \tag{2.20}$$

where $l_i = v_F \tau_i$. The relationships (2.19) and (2.20) can be interpreted as the conditions of applicability of perturbation theory in calculating $\mathbf{f}_1^{(0)}$ and $\mathbf{f}_1^{(1)}$.

We now derive an equation for the electron temperature $T_e(t)$, which enters through f_0 into the expressions (2.17) and (2.18) for the anisotropic part of the electron distribution function. We multiply the kinetic equation (2.1) by $\varepsilon_{\mathbf{p}}$ and integrate the product over the angles of vector $\mathbf{p}$. Allowing for (2.6) and (2.12) and the fact that T_e is uniform over the film thickness, we arrive at the thermal balance equation

$$c_e(T_e)\,\frac{dT_e}{dt} = \langle \mathbf{j}(t)\rangle\cdot\mathbf{E}(t) - \langle P_{ep}(t)\rangle + W(t). \tag{2.21}$$

Here the electron specific heat $c_e(T_e)=\gamma T_e$, with the constant $\gamma = \frac{1}{3}\pi^2 N(0) k_B^2$. The current density averaged over the film thickness, which enters into the expression for the Joule heat release, is given by

$$\langle \mathbf{j}(t)\rangle = \frac{1}{d}\int_0^d dz \int \frac{2\,d^3p}{(2\pi\hbar)^3}\, e\mathbf{v}(\mathbf{f}_1\cdot\mathbf{p}). \tag{2.22}$$

However, below we ignore the first term on the right-hand side of Eq. (2.21), since it is the contribution of the power of a weak measuring current. For the power (averaged over the film thickness) transferred per unit volume from the electron subsystem to the phonon subsystem, we can write

$$\langle P_{ep}\rangle = \frac{1}{d}\int_0^d dz \int \frac{2\,d^3p}{(2\pi\hbar)^3}\,\varepsilon_{\mathbf{p}} I_{ep}(f_0,N_{\mathbf{q}}). \tag{2.23}$$

When the film is heated not only by a current but also, for example, by a laser beam, we must add to the right-hand side of Eq. (2.21) the specific power of a uniform external heating source, $W(t)$. Note that $W(t)$ does not affect the anisotropic part of the electron distribution function.

What is important is that when the total current flowing through the film is fixed, the electron temperature T_e and the electric field strength E can be found from Eqs. (2.21) and (2.22), since these equations constitute a complete set with respect to T_e and E. To express the quantities that enter into (2.21) and (2.22) in terms of T_e and E, we must first find the phonon distribution function. If we ignore phonon drag effects and replace the function $f_{\mathbf{k}}$ in the phonon–electron collision integral (2.4) by $f^{(0)}(\varepsilon_{\mathbf{k}})$, integration over the wave vectors $\mathbf{k}$ becomes easy. As a result we obtain

$$I_{pe}(N_{\mathbf{q}}) = \nu[n_q(T_e(t)) - N_{\mathbf{q}}(z,t)]. \tag{2.24}$$

Here $n_q(T) = [\exp(\hbar\omega_q/k_B T) - 1]^{-1}$ is the Bose occupation-number function for phonon states with wave vector $\mathbf{q}$, and

$$\nu = \frac{m^2\mu_1^2}{2\pi\hbar^4\rho s}\,\Omega^{1+r} \tag{2.25}$$

is the phonon–electron collision rate. Equation (2.2) with the phonon–electron collision integral (2.24) is a linear differential equation in $N_{\mathbf{q}}(z,t)$. The boundary conditions for this equation can be obtained in the following manner (see also Ref. 5).

Let $\alpha(\vartheta)$ be the probability that a photon incident upon the film/substrate interface $z=0$ at angle ϑ enters the substrate (which occupies the region $z<0$). (For different film and substrate materials the probability $\alpha(\vartheta)$ can be calculated by employing the well-known acoustic mismatch model.[11,21])

If we assume that photons escaping from the film into the substrate propagate ballistically and do not return to the film (because of the film's small width), the boundary condition at $z=0$ has the form

$$N_{\mathbf{q}}^{>}(0,t) = \alpha n_q(T_s) + \beta N_{\mathbf{q}'}^{<}(0,t). \tag{2.26}$$

Here T_s is the substrate temperature, $\beta=1-\alpha$, and $N_{\mathbf{q}'}^{<}$ and $N_{\mathbf{q}'}^{>}$ are the distribution functions of incident and specularly reflected phonons, with $\mathbf{q}=(q_x,q_y,q_z)$ and $\mathbf{q}'=(q_x,q_y,-q_z)$ $(q_z>0)$. At the film's free interface $(z=d)$, specular phonon reflection yields

$$N_{\mathbf{q}'}^{<}(d,t)=N_{\mathbf{q}}^{>}(d,t). \tag{2.27}$$

It is convenient to seek the solution of the kinetic equation (2.2) with the collision integral (2.24) and the boundary conditions (2.26) and (2.27) by first performing a Fourier transformation in time. For the Fourier transform

$$N_{\mathbf{q}}(z,\omega)=\int_{-\infty}^{\infty} N_{\mathbf{q}}(z,t)e^{-i\omega t}dt, \tag{2.28}$$

the kinetic equation becomes a linear ordinary differential equation with constant coefficients:

$$s_z\,\frac{dN_{\mathbf{q}}(z,\omega)}{dz}+(\nu-i\omega)N_{\mathbf{q}}(z,\omega)=\nu n_q(\omega), \tag{2.29}$$

where

$$n_q(\omega)=\int_{-\infty}^{\infty} n_q[T_e(t)]e^{-i\omega t}dt. \tag{2.30}$$

The solution of Eq. (2.29) is

$$N_{\mathbf{q}}(z,\omega)=\frac{\nu}{\nu-i\omega}\,n_q(\omega)+C_{\mathbf{q}}(\omega)\exp\left[-\frac{z(\nu-i\omega)}{s_z}\right]. \tag{2.31}$$

The coefficients $C_q^{>}$ and $C_q^{<}$ (of the vectors $\mathbf{q}$ with $q_2>0$ and $q_2<0$) are determined by the boundary conditions. Plugging (2.31) into the boundary conditions (2.26) and (2.27) yields

$$\begin{aligned}C_{\mathbf{q}}^{>}=(\omega)&\left\{1-\beta\,\exp\left[-\frac{2d(\nu-i\omega)}{|s_z|}\right]\right\}^{-1}\\ &\times\alpha(\vartheta)\left[2\pi\delta(\omega)n_q(T_s)-\frac{\nu}{\nu-i\omega}\,n_q(\omega)\right],\end{aligned} \tag{2.32}$$

$$C_{\mathbf{q}}^{<}(\omega)=C_{\mathbf{q}}^{>}(\omega)\exp\left[-\frac{2d(\nu-i\omega)}{|s_z|}\right]. \tag{2.33}$$

Implementing the rather cumbersome calculations of $N_{\mathbf{q}}(z,t)$ via the inverse Fourier transform, we arrive at the following expressions for the phonon distribution function:

$$\begin{aligned}N_{\mathbf{q}}^{>}(z,t)=&\ \alpha[1-\beta x(0)]^{-1}\,\exp(-z\nu/s_z)n_q(T_s)\\ &+\int_{-\infty}^{t}dt' n_q(T(t'))\nu\\ &\times\exp[-\nu(t-t')]\beta^{[\tau+1-z/2d]},\end{aligned} \tag{2.34}$$

$$\begin{aligned}N_{\mathbf{q}'}^{<}(z,t)=&\ \alpha[1-\beta x(0)]^{-1}\,\exp[-(2d-z)\nu/|s_z|]n_q(T_s)\\ &+\int_{-\infty}^{t}dt' n_q(T(t'))\nu\\ &\times\exp[-\nu(t-t')]\beta^{[\tau+z/2d]},\end{aligned} \tag{2.35}$$

Here $x(0)=\exp(-2d\nu/|s_z|)$, $\tau=|s_z|(t-t')/2d$, and the square brackets in the exponent of β denote the integer part of the quantity inside the brackets.

The specific power transferred from electrons to phonons, averaged over the film thickness, i.e., $\langle P_{ep}\rangle$, can be expressed in terms of the phonon–electron collision integral (2.24):

$$\begin{aligned}\langle P_{ep}(t)\rangle=&\frac{1}{d}\int_0^d dz\int\frac{d^3q}{(2\pi)^3}\,\Omega I_{pe}(N_{\mathbf{q}})\\ =&\frac{1}{d}\int_0^d dz\int_{q_z>0}\frac{d^3q}{(2\pi)^3}\\ &\times\Omega\nu[n_q(T_e(t))-N_q^{>}(z,t)]\\ &+\frac{1}{d}\int_0^d dz\int_{q_z<0}\frac{d^3q}{(2\pi)^3}\\ &\times\Omega\nu[n_q(T_e(t))-N_q^{<}(z,t)].\end{aligned} \tag{2.36}$$

Averaging the phonon distribution function over z explicitly and plugging the result into (2.21), we arrive at the final form of the heat balance equation:

$$\begin{aligned}c_e(T_e)\,\frac{dT_e}{dt}=&W(t)+\langle\mathbf{j}(t)\rangle\cdot\mathbf{E}(t)\\ &-2\int_{q_z>0}\frac{d^3q}{(2\pi)^3}\,\Omega\nu\biggl\{n_q(T_e(t))-n_q(T_s)\\ &-\int_{-\infty}^{t}dt'[n_q(T_e(t'))-n_q(T_s)]\\ &\times\nu\,\exp[-\nu(t-t')]\beta^{[\tau]}(1-\alpha\{\tau\})\biggr\}.\end{aligned} \tag{2.37}$$

In deriving Eq. (2.37), we assumed that the power $W(t)$ absorbed by the electrons is an arbitrary function of time. Likewise, $[\tau]$ and $\{\tau\}$ denote the integer and fractional parts of τ, respectively.

The phonon contribution to Eq. (2.27) for the electron temperature consists of two terms. The first (local in time) describes the emission of nonequilibrium phonons at time t, and the second (integrated over time) is related to the reabsorption of nonequilibrium phonons emitted at earlier times $t'<t$. Clearly, the sum of the last two terms in braces is the phonon distribution function averaged over the film thickness and multiplied by -1.

The phonon distribution function found by solving Eq. (2.2) makes it possible to write the following expression for the time-dependent thermal flux from film to substrate:

$$\begin{aligned}Q(t)=&\int_{q_z>0}\frac{d^3q}{(2\pi)^3}\,\Omega s_z[N_{\mathbf{q}}^{<}(0,t)-N_{\mathbf{q}}^{>}(0,t)]\\ =&\int_{q_z>0}\frac{d^3q}{(2\pi)^3}\,\Omega s_z\alpha\nu\int_{-\infty}^{t}dt'[n_q(T_e(t'))\\ &-n_q(T_s)]\exp[-\nu(t-t')]\beta^{[\tau]}.\end{aligned} \tag{2.38}$$

Here the dependence of $Q(t)$ on the intensity of the heating source, $W(t)$, is implicitly expressed in terms of $T_e(t)$.

Under time-independent heating, for which $W(t)=W_0$, the temperature of electrons in the film, T_0, is determined by the equation

$$W_0=\int_{q_z>0}\frac{d^3q}{(2\pi)^3}\ \Omega\ \frac{\alpha s_z}{2d}\,\frac{1-x}{1-\beta x}\ [n_q(T_0)-n_q(T_s)]. \tag{2.39}$$

The physical picture of time-independent heat removal that follows from Eq. (2.39) has been thoroughly analyzed by one of the present authors (see Ref. 6). (Heat removal from superconducting films in the flux-flow regime near T_c was examined in Ref. 22.) It turns out that there are two limiting cases, depending on the relationship between $l_{pe}(T_0)$ (the inelastic mean free path for phonons with energy $\hbar\omega_q\simeq k_BT_0$) and the so-called effective film thickness $d_{\text{eff}}=d/\langle\alpha\rangle$, where $\langle\alpha\rangle$ is the angle-averaged probability of a phonon's traversing the film/substrate interface. (The explicit expressions for $l_{pe}(T_0)$ and $\langle\alpha\rangle$ are given below.) In effectively thick films ($d_{\text{eff}}\gg l_{pe}(T_0)$), nonequilibrium phonons become thermalized because they are reabsorbed by electrons, and the temperature of both subsystems becomes equal to $T_0>T_s$ (the Joule heating regime). In this regime T_0 is determined by the acoustic transparency of the film/substrate interface. In the opposite limiting case of effectively thin films, almost all nonequilibrium phonons escape to the substrate without reabsorption, as a result of which electrons become overheated in relation to the lattice (electron overheating), with the magnitude of this overheating determined by the electron–phonon collision rate.

Equation (2.38) implies that in the Joule heating regime, the average thermal flux from the film is

$$Q_0=\frac{\pi^2\langle\alpha\rangle k_B^4}{120\hbar^3s^2}\ (T_0^4-T_s^4), \tag{2.40}$$

where the angle-averaged probability that a phonon traverses the film/substrate interface is

$$\langle\alpha\rangle=\int_0^{\pi/2}d\vartheta\ \alpha(\vartheta)\sin 2\vartheta. \tag{2.41}$$

In the electron overheating regime we have

$$Q_0=\frac{D_{5+r}}{4\pi^3}\ d\ \frac{m^2\mu_1^2k_B^{5+r}}{\hbar^7\rho s^4}\ (T_0^{5+r}-T_s^{5+r}), \tag{2.42}$$

where

$$D_k=\int_0^\infty\frac{x^{k-1}}{e^x-1}\ dx. \tag{2.43}$$

When $W_0=Q_0/d$ is small, temperature variations in the Joule heating regime can be expressed as follows:

$$T_0-T_s=\frac{W_0\tau_{es}}{c_p(T_0)}, \tag{2.44}$$

where $c_p=(2\pi^2/15)(k_B^4T_0^3/\hbar^3s^3)$ is the phonon specific heat in the model with one acoustic branch of vibrations, and $\tau_{es}=4d/\langle\alpha\rangle s$ is the average escape time of phonons crossing to the substrate.

In the electron overheating regime, the small increase in temperature can be conveniently written as

$$T_0-T_s=\frac{W_0\tau_e(T_0)}{c_e(T_0)}. \tag{2.45}$$

Linearizing (2.42) in the small quantity W_0 and comparing the result with (2.45), we arrive at the following expression for the cooling time of electrons in the thin film:

$$\tau_e(T_0)=\frac{4\pi^5N(0)\hbar^7\rho s^4}{3(5+r)D_{5+r}m^2\mu_1^2(k_BT_0)^{3+r}}. \tag{2.46}$$

If we now introduce the average phonon–electron collision rate

$$\nu_{pe}(T_0)=\tau_{pe}^{-1}(T_0)=\frac{1}{c_p(T_0)}\int\frac{d^3q}{(2\pi)^3}\ \Omega\nu\ \frac{dn_q(T_0)}{dT_0}, \tag{2.47}$$

we obtain for the average phonon–electron collision time

$$\begin{aligned}\tau_{pe}(T_0)&=\frac{8\pi^5}{15(5+r)D_{5+r}}\,\frac{\hbar^4\rho s}{m^2\mu_1^2}\ (k_BT_0)^{-(1+r)}\\&=\frac{c_p(T_0)}{c_e(T_0)}\ \tau_e(T_0).\end{aligned} \tag{2.48}$$

Here for the inelastic mean free path of phonons with an energy of order k_BT_0 we have $l_{pe}(T_0)=s\tau_{pe}(T_0)$.

3. NONLINEAR RELAXATION OF ELECTRON TEMPERATURE IN A METALLIC FILM

The nonlinear integrodifferential equation (2.37) derived in Sec. 2 determines the dynamics of electron temperature in the general case, i.e., for an arbitrary time dependence of the heat release power $W(t)$ and for all values of the probability that a phonon crosses the film/substrate interface, $0<\alpha<1$. In this section we are primarily interested in cooling of the electron–phonon system of a film after the film is illuminated at time t_0 by an ultrashort laser pulse that is narrow compared to the shortest characteristic time of inelastic electron–phonon and electron–electron collisions (with the initial electron temperature being $T_e(0)$). We also assume that the laser pulse intensity is relatively high and that the initial electron temperature $T_e(0)$ is much higher than T_s, although it is low compared to Θ_D. Under these assumptions, nonlinear relaxation of the electron temperature can be analyzed analytically in two limiting cases: electron overheating ($\tau_{es}\ll\tau_{pe}$), and Joule heating ($\tau_{es}\gg\tau_{pe}$).

In the electron overheating regime, nonlinear phonons leave the film without being reabsorbed by electrons, with the result that the phonon distribution function in the film is approximately $n_q(T_s)$. A formal consequence is that the time-integral term in Eq. (2.37) is small. Ignoring the integral term in the parameter $\tau_{es}/\tau_{pe}\ll 1$, we arrive at a nonlinear differential equation for the electron temperature:

$$\tau_e(T_e(0))\Theta\ \frac{d\Theta}{dt}=-\frac{\Theta^{5+r}-\Theta_B^{5+r}}{5+r}, \tag{3.1}$$

where $\Theta = T_e(t)/T_e(0)$ and $\Theta_B = T_s/T_e(0)$ are the normalized temperatures. Equation (3.1) shows that initially, when $T_e(t) \gg T_s$, electron cooling follows a power law:

$$\Theta(t) = \left[1 + \frac{3+r}{5+r}\,\frac{t}{\tau_e(T_e(0))}\right]^{-1/(3+r)}. \tag{3.2}$$

The formula generalizes the result of Kaganov *et al.*[4] to the case of an arbitrary r. The initial temperature $T_e(0)$ can be found from the equation

$$\frac{\gamma}{2}\,[T_e^2(0) - T_s^2] = \mathscr{E}_0, \tag{3.3}$$

where $\mathscr{E}_0$ is the laser pulse energy absorbed by electrons, and $\gamma = \pi^2/3N(0)k_B^2$. As the excess energy diminishes, the electron temperature approaches the thermostat temperature, and for $T_e(t) - T_s \ll T_s$ the relaxation of $T_e(t)$ becomes exponential:

$$T_e(t) - T_s \propto \exp\left[-\frac{t}{\tau_e(T_s)}\right]. \tag{3.4}$$

Note that in the case under consideration the characteristic times at the initial and final stages of relaxation differ considerably. Their ratio is of order $\tau_e(T_e(0))/\tau_e(T_s) = [T_s/T_e(0)]^{3+r} \ll 1$.

In the limit of effectively thick films ($\tau_{es} \gg \tau_{pe}$), the process of nonlinear relaxation of the electron temperature is more complicated and can be separated into three stages. At the first, initial, stage nonequilibrium phonons are emitted by ''hot'' electrons, and the number of these phonons in the film increases, since they have no time to be reabsorbed by electrons or to escape to the substrate. The first stage lasts for about $\tau_e(T_e(0))$, with the time dependence of the electron temperature described by Eq. (3.2). At the second, intermediate, stage the electron–phonon system becomes thermalized, i.e., the electron and phonon subsystems have the same temperatures. Thermalization of relatively thick films is described by Eq. (2.37), where we must ignore the escape of phonons to the substrate and put $\alpha = 0$ and $\beta = 1$:

$$c_e(T_e)\,\frac{T_e}{dt} = -\int \frac{d^3q}{(2\pi)^3}\,\hbar\omega_q\nu\Bigg\{n_q(T_e(t)) - n_q(T_s) - \int_0^t dt'[n_q(T_e(t') - n_q(T_s))]\nu e^{-\nu(t-t')}\Bigg\}. \tag{3.5}$$

Here integration is over $t' > 0$, since for $t' < 0$ we have $T_e = T_s$; the initial value of the electron temperature, $T_e(0)$, is specified by Eq. (3.3). The sum of the last two terms in braces is the phonon distribution function at time t multiplied by -1. With the passage of time t, the occupation numbers of the phonon states increase, and $n_q(T_e(t))$ decreases, which makes the right-hand side of Eq. (3.5) vanish, i.e., the electron temperature acquires its stationary value. We also note that Eq. (3.5) describes the relaxation of electron temperature in a relatively thick film without the familiar simplifying assumption concerning the phonon temperature,[4,13] since such an assumption is not justified in the temperature range of interest.

Turning to the relaxation dynamics of $T_e(t)$, we consider, first and foremost, the case in which $r = -1$, since it allows for a rigorous solution; the corresponding temperature dependence of the inelastic electron–phonon collision rate is quadratic: $\tau_e^{-1}(T_e) \propto T_e^2$. What is important is that such a dependence was observed in a number of low-temperature experiments (at $T < 20$ K), including studies of localized corrections to the resistance of thin films,[9,23–25] and of the frequency dependence of the response of films in the resistive state to oscillations in the power of the heat source.[10,16,26] In accordance with Eq. (2.25), at $r = -1$ the phonon–electron collision rate is independent of the phonon wave vector $\mathbf{q}$. This makes it possible to integrate over the vectors $\mathbf{q}$ and in this way to proceed from (3.5) to an equation for $T_e(t)$, which in dimensionless variables has the form

$$4\zeta\Theta\,\frac{d\Theta}{dx} = -\Theta^4 + \int_0^x \Theta^4(x')e^{-(x-x')}dx'. \tag{3.6}$$

Here $x = \nu_{pe}(T_e(0))t$, $\zeta = c_e(T_e(0))/c_p(T_e(0))$, and, in addition, it is assumed that $\Theta \gg \Theta_B$, as in an effectively thin film. By differentiating with respect to x we can reduce the resulting integrodifferential equation to an ordinary differential equation, which allows for reduction of order. The final equation

$$4\zeta\Theta\,\frac{d\Theta}{dx} = -\Theta^4 + 2\zeta(1-\Theta^2). \tag{3.7}$$

Clearly, by setting the right-hand side of Eq. (3.7) to zero, we arrive at an equation for the dimensionless temperature $\Theta_{th} = T_{th}/T_e(0)$, which shows up in the electron–phonon system as a result of the system's thermalization. Indeed, in the process of thermalization the laser pulse energy $\mathscr{E}_0$ absorbed by the film is redistributed among electrons and phonons, so that

$$\gamma[T_e^2(0) - T_s^2] = \gamma(T_{th}^2 - T_s^2) + \tfrac{1}{2}[c_p(T_{th})T_{th} - c_p(T_s)T_s]. \tag{3.8}$$

In the approximation $T_{th}^2 \gg T_s^2$, this yields

$$\Theta_{th}^4 - 2\zeta(1 - \Theta_{th}^2) = 0 \tag{3.9}$$

which has the roots

$$\Theta_{1,2}^2 = -\zeta \pm \sqrt{\zeta^2 + 2\zeta}, \tag{3.10}$$

with $\Theta_{th} = \Theta_1$ (since $\Theta_1 > 0$).

The solution of Eq. (3.7) with initial condition $\Theta(0) = 1$ and an arbitrary value of ζ can be written

$$\Theta^2(x) = \frac{\Theta_1^2(1-\Theta_2^2) - \Theta_2^2(1-\Theta_1^2)\exp[-(x/2\zeta)(\Theta_1^2-\Theta_2^2)]}{1-\Theta_2^2-(1-\Theta_1^2)\exp[-(x/2\zeta)(\Theta_1^2-\Theta_2^2)]}. \tag{3.11}$$

The behavior of $\Theta(x)$ is governed by a single parameter, ζ, whose value is fixed by the initial electron temperature $T_e(0)$. If we introduce a temperature T^* such that $c_e(T^*) = c_p(T^*)$ (for ordinary metals $T^* \sim 1$ K), then for $T_e(0) \gg T^*$, $\zeta \ll 1$; in the opposite limit $T_e(0) \ll T^*$ but $T_e(0) \gg T_s$ (it is difficult to realize this situation in experiments) we have $\zeta \gg 1$. From (3.7) and (3.11) it follows that for small ζ, $\Theta(x)$ first rapidly decreases in a characteristic

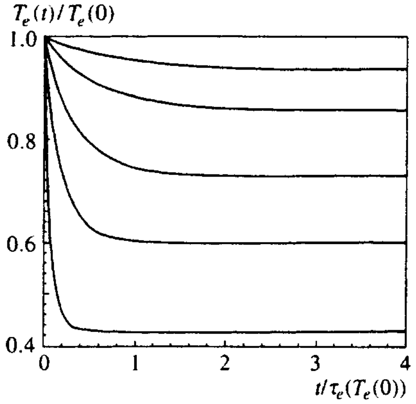


FIG. 1. Relaxation of the normalized electron temperature in the thermalization of the electron–phonon system of an effectively thick film at $r=-1$. The curves correspond to $\zeta=0.02$, 0.1, 0.3, 1, and 3 (bottom to top). The horizontal asymptotes of the curves represent the respective thermalization temperatures Θ_{th}.

time $\tau_e(T_e(0))=\zeta\tau_{pe}$, and then over a time $\tau_e(T_{th})\sim\zeta^{1/2}\tau_{pe}$ acquires the stationary value $\Theta_{th}\approx(2\zeta)^{1/4}$. Note that both time scales are much shorter than the characteristic phonon–electron collision time τ_{pe}. In the limit $\zeta\gg1$, the thermalization temperature $\Theta_{th}\approx1-\zeta^{-1}$, i.e., it differs little from the initial temperature, and the characteristic time it takes the system to become stationary coincides with τ_{pe}. The function $\Theta(x)$ for several values of ζ is depicted in Fig. 1.

Now assume that r is arbitrary. As the analysis of the case with $r=-1$ shows, the duration of the thermalization process at $T_e(0)\gg T^*$ is governed by the electron–phonon collision time $\tau_e(T_{th})$, which is much shorter than the phonon–electron collision time. This makes it possible to replace the exponential $\exp[-\nu(t-t')]$ in Eq. (3.5) by unity, and to integrate over phonon wave vectors. As a result, we obtain an equation for the normalized temperature $\Theta(x)$ (which is assumed high compared to Θ_B):

$$\zeta\Theta\ \frac{d\Theta}{dx}=-\frac{\Theta^{5+r}}{5+r}+\frac{F_r(3+r)}{(5+r)^2}\int_0^x dx'\,\Theta^{6+2r}(x'),\tag{3.12}$$

where

$$F_r=\frac{4\pi^4 D_{6+2r}}{15(3+r)F_{5+r}^2}.$$

In the first (initial) stage of relaxation, we can neglect the integral and obtain the solution (3.2) of the truncated equation. Since $\Theta_{th}\ll1$, we can assume that the initial stage of relaxation provides the main contribution to the integral term in (3.12), which means that in the integral we can replace the desired function $\Theta(x)$ by the function $\Theta_0(x)$ of (3.2). Note that

$$\int_0^x dx'\,\Theta_0^{6+2r}(x')=\frac{5+r}{3+r}\,[1-\Theta_0^{3+r}(x)].\tag{3.13}$$

If we now plug (3.13) into (3.12) and replace $\Theta_0(x)$ with $\Theta(x)$, we obtain the ordinary differential equation

$$\zeta\Theta\ \frac{d\Theta}{dx}=-\frac{1}{5+r}\,[\Theta^{5+r}-F_r\zeta(1-\Theta^{3+r})],\tag{3.14}$$

which coincides at $r=-1$ with the exact equation (3.6), since $F_{-1}=2$. Strictly speaking, Eq. (3.14) is valid in the region where the first term in the square brackets is considerably larger than the second. At the same time we hope that the first term also correctly describes the behavior of the electron temperature in the region where x is large, yielding, in particular, the stationary value

$$\Theta_{th}\approx(F_r\zeta)^{1/(5+r)},$$

which for $-1<r<0$ differs little from the value $\Theta_{th}\approx(2\zeta)^{1/4}$ that follows from the exact equation (3.9). (Note that for $T_e(0)\gg T^*$, the parameter ζ is much smaller than unity.)

To find the characteristic system thermalization time for arbitrary values of r, we examine the solution of Eq. (3.14) for large values of x. Expanding the right-hand side of Eq. (3.14) in powers of $\Theta_1'=\Theta-\Theta_{th}$, we obtain a linear equation for Θ_1',

$$\zeta\ \frac{d\Theta_1'}{dx}=-\Theta_{th}^{3+r}\Theta_1',\tag{3.15}$$

which implies that the electron temperature approaches its stationary value T_{th} exponentially, with a characteristic thermalization time $t_{th}\sim\tau_e(T_{th})$.

Thus, in time t_{th} following the absorption of a laser pulse, the phonon distribution function over the entire film (except for a small region of thickness $st_{th}\ll l_{pe}$ near the film/substrate interface) becomes a Bose distribution function, and is characterized by a phonon temperature coinciding with the electron temperature. Further relaxation of the excess energy does not disrupt the electron–phonon thermalization process.

To analyze the last (third) stage in the cooling of an effectively thick film, we examine a somewhat more general case, in which the power $W(t)$ absorbed by the film varies considerably over time intervals much longer than the characteristic electron–phonon collision time τ_{pe}, and we then set W to zero. Here the time-integral term in Eq. (2.37) reduces to a local term. Indeed, for $\tau_{es}\gg\tau_{pe}$ the kernel of the integral decreases exponentially in times $t-t'\sim\tau_{pe}$, so that for slower variations of $W(t)$ we can expand the function $n_q(T_e(t'))$ in powers of $t-t'$ and keep only the linear terms of the expansion. Integrating with respect to t', we obtain, at $T=T_e=T_p$,

$$[c_e(T)+c_p(T)]\ \frac{dT}{dt}=W(t)-\int\frac{d^3q}{(2\pi)^3}\ \Omega\ \frac{\alpha(\vartheta)s_z}{d}\times[n_q(T(t))-n_q(T_s)].\tag{3.16}$$

At moderate temperatures, with $c_p(T)\gg c_e(T)$, i.e., $T\gg T^*$, it is convenient to express the left- and right-hand sides of Eq. (3.16) in terms of the phonon energy density

$$\mathscr{E}_p(T)=\int\frac{d^3q}{(2\pi)^3}\ \Omega n_q(T).\tag{3.17}$$

Separating integration over angles from integration over the absolute value of the wave vector **q** on the right-hand side of Eq. (3.16), we obtain a linear differential equation for $\mathscr{E}_p(T)$:

$$\frac{d\mathscr{E}_p(T(t))}{dt}=W(t)-\frac{1}{\tau_{es}}\,[\mathscr{E}_p(T(t))-\mathscr{E}_p(T_s)]. \qquad (3.18)$$

The solution of this equation can be written in the general form

$$\mathscr{E}_p(T(t))=\mathscr{E}_p(T_s)+[\mathscr{E}_p(T(t_0))-\mathscr{E}_p(T_s)]\times\exp\left(-\frac{t-t_0}{\tau_{es}}\right)+\int_{t_0}^{t}W(t')\exp\left(-\frac{t-t'}{\tau_{es}}\right)dt', \qquad (3.19)$$

with no restrictions imposed on the magnitude of the absorbed power $W(t)$; the time $t_0\sim t_{th}$.

To describe relaxation of the excess energy of a heated film in the last stage, we must set $W(t)$ to zero in the solution (3.19). Here it becomes obvious that in the nonlinear regime the exponentially relaxing quantity is the energy of the system, rather than the temperature. The film temperature relaxes exponentially when $T(t)-T_s\ll T_s$.

What we have just discussed refers to the nonlinear relaxation of electron temperature when the film is under intense heating. However, experiments are possible in which the film is heated by a low-intensity laser pulse. We now turn to energy relaxation in a weakly nonequilibrium electron–phonon system, where $T_e(0)-T_s\ll T_s$.

After an effectively thin film has been weakly heated, the nonequilibrium addition to the electron temperature, $T_1(t)=T_e(t)-T_s$, obeys the equation

$$\tau_e(T_s)\,\frac{dT_1}{dt}=-T_1, \qquad (3.20)$$

which has a solution in the form of an exponential function:

$$T_1(t)=T_1(0)\exp\left[-\frac{t}{\tau_e(T_s)}\right]. \qquad (3.21)$$

In the case of initial low-intensity heating of an effectively thick film, the thermalization of the electrons and phonons is described by the following integro-differential equation (cf. Eq. (3.5)):

$$c_e(T_s)\,\frac{dT_1}{dt}=-\int\frac{d^3q}{(2\pi)^3}\,\hbar\omega_q\nu\,\frac{dn_q}{dT_s}\left[T_1(t)-\int_0^t dt'\,T_1(t')\,\nu e^{-\nu(t-t')}\right]. \qquad (3.22)$$

Let us first obtain the exact solution of Eq. (3.22) for $r=-1$, in which the phonon–electron collision rate ν is independent of the phonon wave vector **q**. Differentiation with respect to time transforms Eq. (3.22) to the differential equation

$$\zeta_s\,\frac{d^2T_1}{dx^2}=-(1+\zeta_s)\,\frac{dT_1}{dx}, \qquad (3.23)$$

whose solution is

$$T_1(x)=[T_1(0)-T_1(\infty)]\exp\left[-\frac{x(1+\zeta_s)}{\zeta_s}\right]+T_1(\infty). \qquad (3.24)$$

Here the parameter ζ_B is equal to $c_e(T_B)/c_p(T_B)$, and the thermalization temperature is

$$T_1(\infty)=T_1(0)\,\frac{\zeta_s}{1+\zeta_s}.$$

Thus, the duration of the thermalization process is governed by the electron–phonon collision time $\tau_e(T_s)$ when $\zeta_s\ll 1$, and by the phonon–electron collision time $\tau_{pe}(T_s)$ when $\zeta_s\gg 1$.

We now consider arbitrary values of r in the limit $\zeta_s\ll 1$, where the system's thermalization time is short compared to $\tau_{pe}(T_s)$. This makes it possible to replace the exponential $\exp[-\nu(t-t')]$ in Eq. (3.22) by unity and obtain, after differentiating with respect to time,

$$\zeta_s\,\frac{d^2T_1}{dx^2}+\frac{dT_1}{dx}-K_rT_1=0, \qquad (3.25)$$

where

$$K_r=\frac{4\pi^4(6+2r)D_{6+2r}}{15(5+r)^2D_{5+r}^2}.$$

The solution of Eq. (3.25) that is bounded at $x=0$ has the form

$$T_1(x)=T_1(0)\exp\left[-\left(\frac{1}{\zeta_s}+K_r\right)x\right]. \qquad (3.26)$$

Note that at $r=-1$, K_{-1} is equal to unity, i.e., the expression in square brackets in (3.26) coincides with the power of the exponential function in the exact solution (3.24). This solution correctly describes the initial stage in the energy relaxation of the system, but leads to incorrect asymptotic behavior as $x\rightarrow\infty$. Correcting this deficiency, we use (3.24) and (3.26) to write the interpolation formula

$$T_1(x)=[T_1(0)-T_1(\infty)]\exp\left[-\left(\frac{1}{\zeta_s}+K_r\right)x\right]+T_1(\infty), \qquad (3.27)$$

which is valid for both small and large values of x.

After the electron–phonon system becomes thermalized, the film temperature relaxes exponentially:

$$T_1(t)=T_1(t_0)\exp\left(-\frac{t-t_0}{\tau_{es}}\right), \qquad (3.28)$$

where $T_1(t_0)$ coincides with $T_1(\infty)=\zeta_sT_1(0)$, and t_0 is of the order of the system's thermalization time. Combining the results (3.27) and (3.28), we can write a single interpolation formula for all stages of relaxation of the electron temperature in an effectively thick film subjected to initial low-intensity heating:

$$T_1(x)=T_1(0)\left\{(1-\zeta_s)\exp\left[-\frac{(1+\zeta_sK_r)t}{\tau_e(T_s)}\right]+\zeta_s\right\}$$

$$\times\exp\left(-\frac{t}{\tau_{es}}\right). \tag{3.29}$$

4. NONLINEAR RELAXATION OF THE ELECTRICAL CONDUCTIVITY OF A COOLING FILM

The electron temperature T_e is not a quantity that can be directly measured in experiments. At the same time, in this section we show that the time dependence of the electron temperature, $T_e(t)$, can be studied indirectly by measuring the relaxation dynamics of the electrical conductivity of a film. To this end, below we consider the case in which a weak (''measuring'') direct current of average density $\langle j(z)\rangle = j_0$ flows through a film that at time $t=0$ is heated by an energy pulse $\mathscr{E}_0$. We assume that j_0 is so low that we can ignore heating of the film by the current.

Let the field strength have the form $E(t)=E_0+E_1(t)$, where E_0 is the electric field at $t<0$. The first term, E_0, is determined by both the residual resistance of the film and the phonon contribution to the resistance of the film, whose temperature is that of the heat bath.

Before proceeding with the rather cumbersome calculations, we discuss the general plan for calculating the nonstationary component of the electric field, $E_1(t)$, by perturbative techniques. Below we show (see Eq. (4.6)) that we have the following expression for the current density averaged over the film thickness:

$$j_0=[\sigma_0+\sigma_{ph}(t)]E(t)-\sigma_0\tau_i\,\frac{dE}{dt}, \tag{4.1}$$

where the conductivity contains contributions from the scattering of electrons by impurities, σ_0, and by phonons, $\sigma_{ph}(t)$. Since the current that passes through the film is fixed and is independent of the heating intensity, we can write

$$j_0=[\sigma_0+\sigma_{ph}(T_s)]E_0=[\sigma_0+\sigma_{ph}(t)][E_0+E_1(t)]. \tag{4.2}$$

Here we have allowed for the fact that the term with the dE/dt is small compared to $\sigma_0 E_1$ relative to the smallness of the ratio $\tau_i/\tau_e \ll 1$ (see below). Since usually in films $\sigma_0 \gg |\sigma_{ph}|$ and (in our setting) the stationary component of the electric field, E_0, is much larger than the time-dependent component E_1, keeping only the main contribution we can arrive at a simple expression for the time-dependent component of the field:

$$E_1(t)=E_0\,\frac{\sigma_{ph}(T_s)-\sigma_{ph}(t)}{\sigma_0}. \tag{4.3}$$

Thus, to calculate $E_1(t)$ we must find in explicit form the phonon contribution to the electrical conductivity of the film, $\sigma_{ph}(t)$.

To find the relationship that links E and T_e in the nonstationary case (the DC heating of a film is discussed in Ref. 5), we plug the anisotropic part of the electron distribution function defined by Eqs. (2.16)–(2.18) into the averaged current density (2.22). If we allow for the explicit expression for the electron–phonon collision integral, we obtain

$$\begin{aligned}\mathbf{j}_0=&\,\sigma_0\mathbf{E}(t)-\sigma_0\tau_i\,\frac{d\mathbf{E}}{dt}\\&-\left(\frac{e\tau_i\hbar}{m}\right)^2\frac{1}{d}\int_0^d dz\int\frac{2\,d^3k}{(2\pi)^3}\int\frac{d^3q}{(2\pi)^3}\,w(q)\\&\times\mathbf{q}(\mathbf{E}\cdot\mathbf{k})\,\frac{\partial f_0}{\partial\varepsilon_{\mathbf{k}}}\{[f_0(\Omega-\varepsilon_{\mathbf{k}})+N_{-\mathbf{q}}]\\&\times\delta(\varepsilon_{\mathbf{k}+\mathbf{q}}-\varepsilon_{\mathbf{k}}+\Omega)\\&+[f_0(\varepsilon_{\mathbf{k}}+\Omega)+N_{\mathbf{q}}]\delta(\varepsilon_{\mathbf{k}+\mathbf{q}}-\varepsilon_{\mathbf{k}}-\Omega)\},\end{aligned} \tag{4.4}$$

where the film's residual conductivity $\sigma_0=n_0e^2\tau_i/m$, with n_0 the concentration of free electrons. To calculate the integrals in Eq. (4.4), we use a spherical coordinate system in which the polar axis is perpendicular to the film and the azimuthal angle is measured relative to the current density vector $\mathbf{j}_0$.

Denoting the polar and azimuthal angles of the vectors $\mathbf{q}$ and $\mathbf{k}$ by ϑ_1,φ_1 and ϑ_2,φ_2, respectively, and allowing for the fact that $N_{\mathbf{q}}$ is independent of φ_1, we integrate over azimuthal angles. To this end, we note that if a function Φ depends on the difference of angles $\varphi_1-\varphi_2$,

$$\begin{aligned}&\int_0^{2\pi}d\varphi_1\int_0^{2\pi}d\varphi_2\,\cos\varphi_1\,\cos\varphi_2\Phi(\varphi_1-\varphi_2)\\&\quad=\pi\int_0^{2\pi}d\psi\,\cos\psi\Phi(\psi).\end{aligned} \tag{4.5}$$

Next, employing the fact that (4.4) contains delta functions, we can integrate over ψ. The result of integration becomes considerably simpler if we note that in the temperature range $T>\Theta_D^2/\varepsilon_F$, where electron–phonon collision predominate over electron–electron collisions, $q/k\sim T/\Theta_D \gg s/v_F$. On the basis of this inequality we can neglect the difference in the arguments of the delta functions and write

$$\begin{aligned}j_0=&\,\sigma_0E(t)-\sigma_0\tau_i\,\frac{dE}{dt}-\frac{e^2\tau_i^2E(t)}{16\pi^5md}\\&\times\int_0^d dz\int k^2dk\,\frac{\partial f_0}{\partial\varepsilon_{\mathbf{k}}}\int_{-1}^1 dv\int dq\;q^2w(q)\\&\times\int_{-1}^1 du[f_0(\Omega-\varepsilon_k)+f_0(\varepsilon_k+\Omega)\\&+N_{\mathbf{q}}+N_{-\mathbf{q}}]\,\frac{(uv+\delta)\chi(Y)}{\sqrt{Y}}.\end{aligned} \tag{4.6}$$

Here $Y=[(1-u^2)(1-v^2)-(uv+\delta^2)]$, $\delta=q/2k$, $u=\cos\vartheta_1$, $v=\cos\vartheta_2$, and $\chi(Y)$ is the Heaviside step function ($\chi=1$ for $Y>0$ and $\chi=0$ for $Y<0$). Integration with respect to v in (4.6) is done according to the formula

$$\int_{v_1}^{v_2}dv\,\frac{uv+\delta}{\sqrt{Y}}=\pi\delta(1-u^2), \tag{4.7}$$

where v_1 and v_2 are the roots of the equation $Y(v)=0$. Evaluation of the integral with respect to k yields the following expression for the phonon contribution to the conductivity:

$$\sigma_{ph}=-\frac{e^2\tau_i^2N(0)}{32\pi^2mk_F}\frac{1}{d}\int_0^d dz\int_0^{q_D}dq\ q^3w(q)\times\int_{-1}^{1}du(1-u^2)[R(\Omega)+N_{\mathbf{q}}+N_{-\mathbf{q}}], \tag{4.8}$$

where

$$R(\Omega)=-\int_{-\infty}^{\infty}d\xi\,\frac{df_0}{d\xi}\,[f_0(\Omega-\xi)+f_0(\xi+\Omega)]=-2n_q(T_e)+2n_q(T_e)[1+n_q(T_e)]\,\frac{\Omega}{k_BT_e}. \tag{4.9}$$

The expression (4.8) splits into two terms in a natural manner. Let us introduce the following quantities:

$$\sigma_{ph}^{(1)}=-\frac{e^2\tau_i^2N(0)}{24\pi^2mk_F}\int_0^{q_D}dq\ q^3w(q)\times n_q(T_e)[1+n_q(T_e)]\,\frac{\Omega}{k_BT_e}, \tag{4.10}$$

$$\sigma_{ph}^{(2)}=\frac{e^2\tau_i^2N(0)}{32\pi^2mk_F}\frac{1}{d}\int_0^d dz\int_0^{q_D}dq\ q^3w(q)\int_{-1}^{1}du(1-u^2)\times[2n_q(T_e)-N_{\mathbf{q}}-N_{-\mathbf{q}}]. \tag{4.11}$$

Equation (10) gives the contribution to the film's conductivity associated with the scattering of electrons whose temperature is $T_e(t)$ by phonons with the same temperature. In the time-independent uniform case, $\sigma_{ph}^{(1)}$ represents the familiar phonon contribution to the electrical conductivity, which for $w(q)\propto q$ is proportional to T_e^5 (Bloch–Grüneisen relation). In the general case (2.9′), where $w(q)\propto q^{1+r}$, we have the following expression for $\sigma_{ph}^{(1)}$:

$$\sigma_{ph}^{(1)}(T_e(t))=-\frac{(5+r)D_{5+r}e^2\tau_i^2N(0)C_r}{12\pi^2mk_F(\hbar s)^{5+r}}\,[k_BT_e(t)]^{5+r}. \tag{4.12}$$

The corresponding addition to the film's resistivity with allowance for the inequality $|\sigma_{ph}^{(1)}|\ll\sigma_0$ and the explicit form of C_r is given by the following formula:

$$\rho_{ph}^{(1)}=-\frac{\sigma_{ph}^{(1)}}{\sigma_0^2}=\frac{(5+r)D_{5+r}m^2\mu_1^2}{12\pi^3\hbar^7\rho s^4(n_0es)^2}\,[k_BT_e(t)]^{5+r}. \tag{4.13}$$

The phonon contribution $\sigma_{ph}^{(2)}$ is finite when the phonon distribution function averaged over the film thickness differs from $n_q(T_e)$. The reason for this difference in the time-independent case is the nonuniformity of the phonon distribution function, related to phonon escape to the substrate. Under time-dependent conditions, the dynamics of the phonon distribution function may provide another reason for the above-noted difference.

Plugging the expression for the phonon distribution function into (4.11) and integrating with respect to z, we get

$$\sigma_{ph}^{(2)}=\frac{e^2\tau_i^2N(0)}{8\pi^2mk_F}\int_0^{q_D}dq\ q^3w(q)\int_0^1du(1-u^2)\times\left\{n_q(T_e(t))n_q(T_s)-\int_0^t dt'[n_q(T_e(t'))-n_q(T_s)]\nu e^{-\nu(t-t')}\beta^{[\tau]}(1-\alpha\{\tau\})\right\}, \tag{4.14}$$

with $T_e(t)$ generally specified by Eq. (2.37). The rather complicated dependence of (4.14) on $u=\cos\vartheta_1$ is due primarily to the general nature of this expression, which holds for an arbitrary relationship between d and l_{pe}. The matter simplifies considerably when the limiting cases are examined. For instance, for an effectively thin film, where there is almost no phonon reabsorption by electrons, we have $N_{\mathbf{q}}\approx n_q(T_B)$, and Eq. (4.14) yields

$$\sigma_{ph}^{(2)}[T_e(t)]=\frac{D_{5+r}e^2\tau_i^2N(0)C_r}{12\pi^2mk_F(\hbar s)^{5+r}}\,k_B^{5+r}[T_e^{5+r}(t)-T_s^{5+r}]. \tag{4.15}$$

Combining this result with (4.12), we find that for effectively thin films

$$\sigma_{ph}=-\frac{D_{5+r}e^2\tau_i^2N(0)C_r}{12\pi^2mk_F(\hbar s)^{5+r}}\,k_B^{5+r}\{(5+r)T_s^{5+r}+(4+r)\times[T_e^{5+r}(t)-T_s^{5+r}]\}. \tag{4.16}$$

We have written (4.16) in such a way so as to isolate explicitly the contribution to the film's conductivity determined by electrons that are heated in relation to the ''cold'' phonons, whose temperature is that of the heat bath, T_s (the last term in the braces); this contribution is characteristic of the electron overheating regime. The fact that this contribution to conductivity tracks the electron temperature adiabatically, i.e., without a time lag, is directly related to the above-noted stationarity of the phonon distribution function in the electron overheating regime. One must also bear in mind that the result (4.16) holds for an arbitrary function $T_e(t)$, which means that it can be used to describe conductivity relaxation after pulsed heating as well. Plugging the characteristic electron–phonon collision time into (4.16), we obtain

$$\sigma_{ph}(\Theta(t))=-\frac{\pi^2\sigma_0\tau_i}{3(5+r)\tau_e(T_e(0))}\left[\frac{T_e(0)}{\Theta_D}\right]^2\times[(4+r)\Theta^{5+r}(t)+\Theta_B^{5+r}]. \tag{4.17}$$

Here the factor $[T_e(0)/\Theta_D]^2$ reflects the familiar idea that at low temperatures the contribution of electron–phonon collisions to the film's resistance is ineffective. The time dependence of the dimensionless temperature in (4.17), $\Theta(t)$, is determined by Eq. (3.1).

Plugging (4.17) into (4.3) yields (for an effectively thin film)

$$E_1(t)=E_0\,\frac{(4+r)\pi^2\tau_i}{3(5+r)\tau_e(T_e(0))}\left[\frac{T_e(0)}{\Theta_D}\right]^2\times[\Theta^{5+r}(t)-\Theta_B^{5+r}].$$

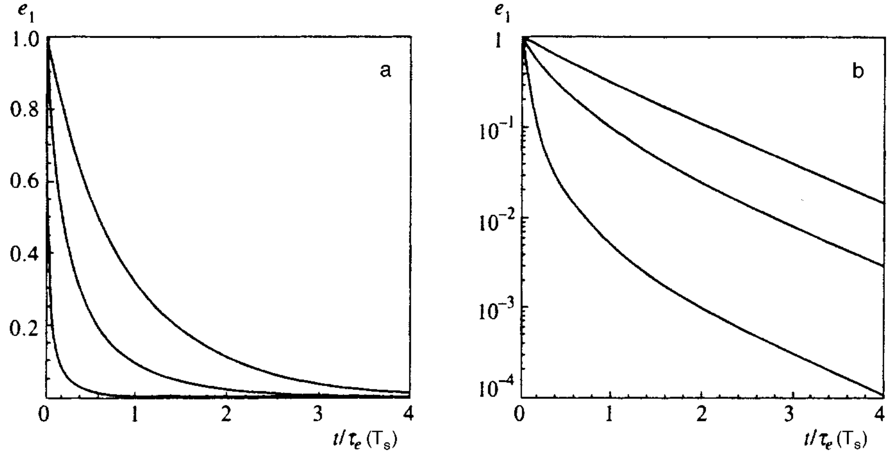


FIG. 2. Relaxation of the normalized correction to the electric field that emerges in an effectively thin film after pulse heating on linear (a) and logarithmic (b) scales. The left curve corresponds to strong initial heating, $\Theta_s = T_s/T_e(0) = 0.2$; the middle curve to moderate heating, $\Theta_s = 0.5$; and the right curve represents the relaxation of e_1 in the case of low-intensity heating, $\Theta_s = 0.9$.

Generally, $\Theta(t)$ is specified by Eq. (3.1). In the case of intense heating, where in the initial stage of electron cooling $\Theta(t) \gg \Theta_s$ and $\Theta(t)$ is given by formula (3.2), conductivity relaxation obeys a power law. Here the nonstationary component of the electric field in the film has the form

$$E_1(t) = \frac{(4+r)\pi^2 E_0 \tau_i}{3(5+r)\tau_e(T_e(0))} \left[\frac{T_e(0)}{\Theta_D}\right]^2 \times \left[1 + \frac{3+r}{5+r}\,\frac{t}{\tau_e(T_e(0))}\right]^{-(5+r)/(3+r)}. \quad (4.18)$$

If the initial heating of the electrons is weak and $\Theta(t) - \Theta_s \ll \Theta_s$, the nonstationary response of a thin film can be written, on the basis of (3.21), as

$$E_1(t) = (4+r)\sigma_0 \rho_{ph}(T_s)\, \frac{T_1(0)}{T_s}\, \exp\left[-\frac{t}{\tau_e(T_s)}\right] E_0. \quad (4.19)$$

Here $\rho_{ph}(T_s)$ is the phonon contribution to the resistivity of the film at the temperature of the heat bath:

$$\rho_{ph}(T_s) = \frac{\pi^2 m}{3 n_0 e^2 \tau_e(T_s)} \left(\frac{T_s}{\Theta_D}\right)^2.$$

Thus, measurements of the time dependence of E_1 in the case of low-intensity heating provide important information about the temperature dependence of the electron–phonon collision time and, correspondingly, about the frequency dependence of the Eliashberg function.

For the special case $r = -1$, Eq. (3.1) can be integrated, which makes it possible to write the relative value of the nonstationary component of the electric field explicitly:

$$e_1(t) = \frac{E_1(t)}{E_1(0)} = \frac{\Theta_s^2}{1-\Theta_s^4} \times \left\{ \left[\frac{(1+\Theta_s^2) + (1-\Theta_s^2)\exp[-t/\tau_e(T_s)]}{(1+\Theta_s^2) - (1-\Theta_s^2)\exp[-t/\tau_e(T_s)]} \right]^2 - 1 \right\} = \Theta_s^4 \left(\Theta_s^2 \cosh \frac{t}{2\tau_e(T_s)} + \sinh \frac{t}{2\tau_e(T_s)} \right)^{-2}. \quad (4.20)$$

This expression holds for an arbitrary intensity of initial heating of a thin film, i.e., for any value $\Theta_s < 1$. The function $e_1(t)$ for $\Theta = 0.2$, 0.5, and 0.9 is depicted in Fig. 2.

To analyze the nonlinear relaxation of the conductivity of an effectively thick film, we write the conductivity as a sum of two terms, $\sigma_{ph} = \sigma_{ph}^{(1)} + \sigma_{ph}^{(2)}$, where $\sigma_{ph}^{(1)}$ and $\sigma_{ph}^{(2)}$ are given by (4.12) and (4.14). The reader will recall that $\sigma_{ph}^{(1)}$ is the contribution to the film conductivity related to the scattering of electrons whose temperature is $T_e(t)$ by phonons with same temperature. The contribution $\sigma_{ph}^{(2)}$ reflects the deviation of the phonon distribution function from the thermalized one. The general expression for $\sigma_{ph}^{(2)}$ can be made simpler if we consider the different stages in the energy relaxation. For instance, in the first two stages, in which electrons primarily emit nonequilibrium phonons, and thermal equilibrium between electrons and phonons sets in due to reabsorption of these phonons, we can ignore the escape of phonons to the substrate and put $\alpha = 0$ and $\beta = 1$ in (4.14). Then

$$\sigma_{ph}^{(2)} = \frac{e^2 \tau_i^2 N(0)}{12\pi^2 m k_F} \int_0^{q_D} dq\, q^3 w(q) \left\{ n_q(T_e(t)) - n_q(T_s) - \int_0^t dt' [n_q(T_e(t')) - n_q(T_s)] \nu e^{-\nu(t-t')} \right\}. \quad (4.21)$$

Plugging

$$w(q) = \frac{2\pi^2 \hbar^3}{m^2 s}\, \nu(q)$$

into (4.21) and comparing the result with (3.5), we arrive at a simple formula:

$$\sigma_{ph}^{(2)}(t) = \frac{e^2 \tau_i^2}{6 m^2 s^2}\, c_e(T_e(t))\, \frac{dT_e}{dt}. \quad (4.22)$$

On the basis of Eq. (3.14), which holds for $T_e(t) \gg T_s$, we can express the right-hand side of this expression in powers of the electron temperature:

$$\sigma_{ph}^{(2)}(T_e(t)) = \frac{D_{5+r} e^2 \tau_i^2 N(0) C_r}{12\pi^2 m k_F (\hbar s)^{5+r}}\, k_B^{5+r} \left\{ T_e^{5+r}(t) \right.$$

$$-\frac{5F_rN(0)\hbar^3s^3}{2k_B^2}\,[T_e^{3+r}(0)-T_e^{3+r}(t)]\bigg\}. \tag{4.23}$$

Combining the two contributions in σ_{ph} and going over to the normalized temperature, we obtain

$$\sigma_{ph}(\Theta(t))=-\frac{\pi^2\sigma_0\tau_i}{3(5+r)\tau_e(T_e(0))}\left[\frac{T_e(0)}{\Theta_D}\right]^2 \times[(4+r)\Theta^{5+r}+F_r\zeta(1-\Theta^{3+r})], \tag{4.24}$$

where the $\Theta(t)$ is given by Eq. (3.14).

In the last stage of energy relaxation, the electron–phonon system of an effectively thick film becomes thermalized, i.e., the electron and phonon temperatures become equal. In this case we can drop the contribution $\sigma_{ph}^{(2)}$ and assume that $\sigma_{ph}=\sigma_{ph}^{(1)}(T(t))$. The dynamical equation for the temperature T follows from (3.19) if we put $W(t)=0$ and express the energy $\mathscr{E}_p$ in terms of temperature:

$$\mathscr{E}_p=\frac{\pi^2k_B^4T^4}{30\hbar^3s^3}. \tag{4.25}$$

As a result we arrive at an equation for $T(t)$,

$$T(t)=\left[T_s^4+(T_{th}^4-T_s^4)\exp\left(-\frac{t-t_0}{\tau_{es}}\right)\right]^{1/4}, \tag{4.26}$$

in which the time t_0 is of the order of the thermalization time for the electron–phonon system of an effectively thick film, t_{th}, and the thermalization temperature is given by

$$T_{th}=T_e(0)[(\zeta^2+2\zeta)^{1/2}-\zeta]^{1/2}. \tag{4.27}$$

Thus, in the last stage of relaxation of the electron–phonon system,

$$\sigma_{ph}(t)=-\frac{\pi^2\sigma_0\tau_i}{3\tau_e(T(t))}\left[\frac{T(t)}{\Theta_D}\right]^2, \tag{4.28}$$

where $T(t)$ is given by Eq. (4.26).

It can be shown that for an effectively thick film the time dependence of the relative value of the electric response is

$$e_1(t)=\Theta^{5+r}+F_r\zeta\,\frac{1-\Theta^{3+r}}{4+r} \tag{4.29}$$

in the first two relaxation stages (the normalized temperature $\Theta(t)$ is the solution of Eq. (3.14)), and

$$E_1(t)=E_0\,\frac{\pi^2\tau_i}{3\tau_e(T_s)}\left(\frac{T_s}{\Theta_D}\right)^2\left\{\left[\frac{T(t)}{T_s}\right]^{5+r}-1\right\} \tag{4.30}$$

in the third stage. Here $T(t)$ is given by Eq. (4.26).

When the heating of an effectively thick film is weak, plugging (3.24) into Eqs. (4.12) and (4.22) yields an expression for the first two relaxation stages:

$$\sigma_{ph}(t)=\sigma_{ph}(T_s)-\frac{\pi^2\sigma_0\tau_i}{3\tau_e(T_s)}\left[\frac{T_s}{\Theta_D}\right]^2\,\frac{T_1(0)}{T_s} \times\left\{(4+r)\exp\left[-\frac{(1+\zeta_sK_r)t}{\tau_e(T_s)}\right]+(5+r)\zeta_s\right\}. \tag{4.31}$$

If we now use Eq. (4.3) and allow for the dynamics of phonon escape to the substrate, for the time-dependent part of the electric field, $E_1(t)$, we can easily obtain an interpolation formula suitable for describing all three stages of relaxation of the electron energy:

$$E_1(t)=E_0\sigma_0\rho_{ph}(T_s)\,\frac{T_1(0)}{T_s} \times\left\{(4+r)\exp\left[-\frac{(1+\zeta_sK_r)t}{\tau_e(T_s)}\right] +(5+r)\zeta_s\right\}\exp\left(-\frac{t}{\tau_{es}}\right). \tag{4.32}$$

Thus, for an effectively thick film, $E_1(t)$ has a characteristic biexponential form with very different time constants.

5. DISCUSSION OF THE EXPERIMENTAL SITUATION

Experiments on the energy relaxation of electrons in metals usually study either variations in the optical characteristics of films (for example, the reflection coefficient)[1–3] or the dynamics of the film's electrical resistance. The time-dependent correction to the resistance of thin films has been measured primarily for pulsed heating (see, e.g., Refs. 27–30). In accordance with the results of Sec. 4, in intense pulsed heating of effectively thick films and effectively thin films there should be an initial power-like relaxation of resistance with a characteristic time of the order of the electron–phonon interaction time at the initial electron temperature, i.e. $E_1(t)\propto 1-t/\tau_e(T_e(0))$ (see Fig. 2a). This implies that, knowing the initial asymptotic behavior of $E_1(t)$ from experiments, we can find the electron–phonon collision time $\tau_e(T_e(0))$. Here the phonon–electron collision time $\tau_{pe}(T_e(t))$ can be found on the basis of Eq. (2.48) by multiplying $\tau_e(T_e(0))$ by the ratio of the phonon and electron specific heats, $c_p(T_e(0))/c_e(T_e(0))$. Note that for a thin film the electron-phonon collision time at the heat bath temperature, $\tau_e(T_s)$, can be found from the asymptotic behavior of $e_1(t)$ for $t\gg\tau_e(T_e(0))$ (see Eq. (4.20) and Fig. 2b). Thus, a thin-film relaxation experiment provides the means for extracting the ''initial'' and ''final'' relaxation times over a broad temperature range by varying T_s and the intensity of the initial heating.

In effectively thick films, the fast component of the photoresponse, which is related to the thermal equilibration that takes place between electrons and phonons in the film, is replaced by the slow bolometric component, which is related to the escape of nonequilibrium phonons to the substrate. Because of the finite temporal resolution of experimental devices, a researcher can usually accurately observe only the relaxation of the bolometric component of the photoresponse, yielding the average time of phonon escape from film to substrate, τ_{es} (see Refs. 29 and 30). One of the possible ways of observing the fast photoresponse component (4.34) is to reduce T_s, since then (in low-intensity heating) the electron–phonon relaxation time grows sharply.

In this paper we assumed that the ballistic regime of heat removal is realized, i.e., phonons emitted into the substrate do not return to the film. (The conditions under which such a regime is possible were discussed in the concluding section of Ref. 5.) At the same time, in thick films, where there is no heat outflow and the transparency of the film/substrate interface is close to unity, the flux of nonequilibrium phonons returning to the film from the substrate proves to be considerable. Experiments[31] have shown that the flux of ''returning'' phonons leads, over long times $t \gg \tau_r$, to a transition from the ballistic regime of heat removal to the diffusion regime, i.e., from the exponential relaxation of the film temperature to a power-law relaxation (proportional to $t^{-1/2}$). The time $\tau_r \sim \tau_{pi,s}/\langle \alpha \rangle^2$ is the characteristic time of nonequilibrium phonon return to film from substrate ($\tau_{pi,s}$ is the elastic phonon scattering time in the substrate).[31]

6. CONCLUSION

The main object of study in this work was the low-temperature ($T_e \ll \Theta_D$) dynamics of the phonon relaxation of ''hot'' electrons in a metallic film on a substrate of an insulator characterized by high thermal conductivity. The physics of the problem clearly shows that energy relaxation plays a key role in forming the response of the electron–phonon system under time-dependent electron heating. To analyze the relaxation dynamics of the system by theoretical methods, we examined (in Sec. 2) a fairly simple but otherwise realistic microscopic model based primarily on the concept of a time-dependent electron temperature. The main feature of this model is the possibility of consistently describing phonon dynamics in the film, using the language of nonuniform and nonstationary phonon distribution functions without employing a commonly adopted (and often uncontrolable) simplifying assumption concerning the phonon temperature. This feature of our approach makes it possible to analyze phonon kinetics in all the basic stages in the electron relaxation process.

Another feature of our model is the use of the fairly general relationship (2.9′), which describes the frequency dependence of the matrix element of the electron–phonon interaction, which is actually equivalent to the assumption that the frequency dependence of the spectral function of the electron–phonon interaction (the Eliashberg function) is arbitrary. We believe that the latter feature makes it possible to compare the theoretical results of this work with experimental data in the fairly simple cases in which the Eliashberg function of a real metallic film differs from the standard Eliashberg function.

We also note that using an arbitrary frequency dependence of the Eliashberg function makes it possible not only to analyze the experimental data but also to obtain in the typical case of $r=-1$ for ''contaminated'' films an exact analytic solution of the problem of the energy relaxation of electrons in the thermalization stage. The exact solution obtained in Sec. 3 can serve as a basis for verifying approximate (within the same model) solutions with $r \neq -1$.

Here is a brief list of the stages in our approach. The basis for our calculations was a system of two nonlinearly coupled (through collision integrals) nonstationary kinetic equations for the electron and phonon distribution functions (Eqs. (2.1) and (2.2), respectively) and a thermal balance equation (Eq. (2.21)) that follows from the system. The main simplifying assumption that allowed obtaining a closed analytic solution of this complicated nonstationary and nonlinear problem was the assumption (often justified; see Sec. 2) that under conditions of predominant elastic electron scattering, the desired electron distribution function is of the Fermi type with a temperature $T_e(t)$ that is time-dependent and uniform over the film thickness.

The latter is justified in view of the high electron (in comparison to phonon) thermal conductivity of good (i.e., uncontaminated) metals. All this made it possible to obtain a closed (in terms of $T_e(t)$) analytic solution of the linear equation (2.29) for the time-dependent and nonuniform phonon distribution function (see (2.34) and (2.35)) without *a priori* assumptions concerning the form of this function. Plugging this solution into the thermal balance equation (2.21), we arrived at a closed integrodifferential equation (Eq. (2.37)), the main result of this paper needed for further studies. The solution of this equation makes it possible to analytically describe the main stages in the evolution of $T_e(t)$ at the microscopic level (Sec. 3). Since $T_e(t)$ cannot be observed directly, we suggested using the related temporal behavior of the total conductivity of the film, $\sigma(t)$, as an ''indicator'' for $T_e(t)$. In Sec. 4 we derived the necessary formulas and analyzed their asymptotic behavior in a number of interesting (from our viewpoint) experimental situations.

[1] H. E. Elsayed-Ali, T. B. Norris, M. A. Pessot, and G. A. Mourou, Phys. Rev. Lett. **58**, 1212 (1987).

[2] R. W. Schoenlien, W. Z. Lin, J. G. Fujimoto, and G. L. Eesley, Phys. Rev. Lett. **58**, 1680 (1987).

[3] S. D. Brorson, A. Kazeroonian, J. S. Modera, D. W. Face, T. K. Cheng, E. P. Ippen, M. S. Dresselhaus, and G. Dresselhaus, Phys. Rev. Lett. **64**, 2172 (1990).

[4] M. I. Kaganov, I. M. Lifshits, and L. V. Tanatarov, Zh. Éksp. Teor. Fiz. **31**, 232 (1956) [Sov. Phys. JETP **4**, 173 (1957)].

[5] V. A. Shklovskiĭ, Pis'ma Zh. Éksp. Teor. Fiz. **26**, 679 (1977); Zh. Éksp. Teor. Fiz. **78**, 1281 (1980) [Sov. Phys. JETP **51**, 646 (1980)].

[6] V. A. Shklovskij, J. Low Temp. Phys. **41**, 375 (1980).

[7] K. V. Maslov and V. A. Shklovskiĭ, Zh. Éksp. Teor. Fiz. **71**, 1514 (1976) [Sov. Phys. JETP **44**, 792 (1976)]; Zh. Éksp. Teor. Fiz. **78**, 3 (1980) [Sov. Phys. JETP **51**, 1 (1980)].

[8] M. I. Kaganov and V. G. Peschanskiĭ, Zh. Éksp. Teor. Fiz. **33**, 1261 (1957) [Sov. Phys. JETP **6**, 970 (1958)].

[9] G. Bergmann, W. Wei, Y. Zou, and R. M. Mueller, Phys. Rev. B **41**, 7386 (1990).

[10] E. M. Gershenzon, M. E. Gershenzon, G. N. Gol'tsman, A. M. Lyul'kin, A. D. Semenov, and A. V. Sergeev, Zh. Éksp. Teor. Fiz. **97**, 901 (1990) [Sov. Phys. JETP **70**, 505 (1990)].

[11] A. W. Little, Can. J. Phys. **37**, 334 (1959).

[12] P. B. Allen, Phys. Rev. Lett. **59**, 1460 (1987).

[13] N. Perrin and H. Budd, Phys. Rev. Lett. **28**, 1701 (1972).

[14] D. Belitz, Phys. Rev. B **36**, 2513 (1987).

[15] D. Pines, *Elementary Excitations in Solids*, W. A. Benjamin, New York (1963).

[16] E. M. Gershenzon, G. N. Gol'tsman, A. I. Elant'ev, B. S. Karasik, and S. E. Potoskuev, Fiz. Nizk. Temp. **14**, 753 (1988) [Sov. J. Low Temp. Phys. **14**, 414 (1988)].

[17] G. Bergmann, Phys. Rep. **27**, 159 (1976).

[18] J. J. Lin and C. E. Wu, Europhys. Lett. **29**, 141 (1995).

[19] W. S. Fann, R. Storz, H. W. K. Tom, and J. Bokor, Phys. Rev. B **46**, 13592 (1992).

[20] C.-K. Sun, F. Vallee, L. Acioli, E. P. Ippen, and J. G. Fujimoto, Phys. Rev. B **48**, 12365 (1993).

[21] S. B. Kaplan, J. Low Temp. Phys. **37**, 343 (1979).

[22] A. I. Bezuglyj and V. A. Shklovskij, Physica C **202**, 234 (1992).

[23] G. Bergmann, Solid State Commun. **46**, 347 (1983).

[24] G. Bergmann, Phys. Rep. **107**, 1 (1984).

[25] S. J. Dorozhkin, F. Lell, and W. Shoepe, Solid State Commun. **60**, 245 (1986).
[26] E. M. Gershenzon, M. E. Gershenzon, G. N. Gol'tsman *et al.*, Zh. Éksp. Teor. Fiz. **86**, 758 (1984) [Sov. Phys. JETP **59**, 442 (1984)].
[27] A. L. Shi, G. L. Huang, C. Lehane, D. Kim, H. S. Kwok, J. Swiatkiewicz, G. C. Xu, and P. N. Prasad, Phys. Rev. B **48**, 6550 (1993).
[28] N. Bluzer, Phys. Rev. B **44**, 10222 (1991).
[29] N. Bluzer, J. Appl. Phys. **71**, 1336 (1992).
[30] N. Bluzer, Phys. Rev. B **46**, 1033 (1992).
[31] A. V. Sergeev, A. D. Semyonov, P. Kouminov, V. Trifonov, I. G. Goghidze, B. S. Karasik, G. N. Gol'tsman, and E. M. Gershenzon, Phys. Rev. B **49**, 9091 (1994).

Translated by Eugene Yankovsky

Stimulation effect in anisotropic superconductors

A. V. Galaktionov

P. N. Lebedev Physics Institute, Russian Academy of Sciences, 117924 Moscow, Russia

(Submitted 16 September 1996)
Zh. Éksp. Teor. Fiz. **111**, 2134–2146 (June 1997)

The stimulation of superconductivity in anisotropic superconductors by electromagnetic and acoustic pumping as well as by the injection of a tunnel current at temperatures close to the superconducting transition temperature is studied. The features distinguishing the stimulation effect from the isotropic case are indicated. © *1997 American Institute of Physics.* [S1063-7761(97)01506-0]

1. INTRODUCTION

The stimulation of superconductivity in ordinary isotropic superconductors was in its time the subject of a large number of theoretical and experimental investigations (see the reviews Refs. 1 and 2). The stimulation effect, first studied by Eliashberg,[3,4] is a physically transparent example of the nonequilibrium properties of a superconductor that can be directly checked experimentally. Such a check is understood to mean an increase produced in the superconducting gap by a change in the quasiparticle distribution function under the action of electromagnetic pumping, a powerful sound wave, injection of a tunneling current, and so on. For example, if the frequency ω of the external electromagnetic radiation is less than twice the superconducting gap 2Δ, then new excitations cannot be created, while existing excitations are redistributed into the region of higher energies. In the self-consistency equation

$$\frac{1}{g}=\int_{\Delta}^{\omega_D}d\epsilon\,\frac{1-2n_{\epsilon}}{\sqrt{\epsilon^2-\Delta^2}} \tag{1}$$

the quasiparticle distribution function n_{ϵ} decreases in the characteristic region of integration, i.e., the superconducting gap increases.

The magnitude of this effect is governed by the ratio ω/Δ. In the case of anisotropic pairing, where the gap at the Fermi surface varies, it can therefore be expected that the stimulation effect will have a different frequency dependence. The fact that the coherence factors averaged over the Fermi surface (see, for example, Ref. 5), which appear in the expression for the electromagnetic absorption, are substantially different for a superconductor with an anisotropic gap whose sign varies over the Fermi surface must also be taken into account.

Similarly, the elastic relaxation time in anisotropic superconductors[6] is different from that in the isotropic case. It will be assumed below that there exists a symmetry transformation of the normal state of a metal under which the superconducting order parameter changes sign. Such superconducting ordering probably occurs in certain compounds with heavy fermions (see the review in Ref. 7). Likewise, the possibility of d-type pairing in high-T_c superconductivity has been widely discussed recently in the literature. Experiments on the measurement of the magnetic flux in a three-crystal configuration,[8–10] the paramagnetic behavior of high-T_c compounds,[11] and other measurements attest to this possibility.

For simplicity, we assume that the Fermi surface is cylindrical or spherical and that the pairing is a singlet pairing. The type of interelectronic interaction leading to anisotropic superconductivity is not important for the questions examined in this paper (we call the carrier of this interaction a ''boson''). In Sec. 2, superconducting stimulation under the action of electromagnetic and acoustic pumping is calculated. It is shown that under electromagnetic pumping, the stimulation effect has a sharper maximum as a function of the radiation frequency (compared with the isotropic case). A sound wave with frequency $\omega>2\Delta$ suppresses superconductivity in an isotropic superconductor, while in the anisotropic case the stimulation effect is positive over the entire experimental frequency range. Stimulation of superconductivity by tunneling injection is studied in Ref. 3. Since the probability of tunneling through an insulating barrier between superconductors depends on the direction of the momentum at the Fermi surface, the stimulation effect depends on the orientation of the plane of the tunnel junction. In the isotropic case, the stimulation effect vanishes for voltages $|eV|>|\Delta|+|\Delta'|$ (where Δ' is the gap in the superconductor with which the stimulated sample is in contact), whereas in the anisotropic case there is no threshold voltage, and for high voltages the effect can be either positive or negative, depending on the orientation.

2. ELECTROMAGNETIC AND ACOUSTIC STIMULATION

Let the superconducting sample be a thin film in good acoustic contact with a substrate with high thermal conductivity (i.e., the substrate can be treated as a heat bath). The film thickness d is assumed to much less than both the skin depth and the London penetration depth at fixed temperature. Thus, the electromagnetic field can be assumed to be uniform in the film. Nonetheless, the thickness d must be much greater than the temperature-dependent coherence length $\xi(T)$ (since near a boundary the order parameter of an isotropic superconductor varies over a characteristic scale $\xi(T)^{12,13}$). By virtue of the latter condition, boundary effects can be neglected and it can be assumed that the gap width equals the bulk value.

It will be assumed below that the temperature is close to the superconducting transition temperature T_c, so that the gap is small compared with T. The frequency ω of the incident radiation is also assumed to be much less than T but, at the same time, much greater than the reciprocal γ of the inelastic relaxation time. On account of the latter condition, radiation absorption is mainly a one-photon process.[2,3] Since elastic scattering by impurities destroys anisotropic superconductivity, we require that the reciprocal $1/\tau$ of the relaxation time satisify $1/\tau \ll T$, but we nonetheless assume that $1/\tau \gg \gamma$. Both the electron–electron interaction $\sim T^2/\epsilon_F$ and the electron–boson interaction, likewise characterized by a power of the temperature higher than the first power (T^3/ω_D^2 for phonons), make the main contribution to γ. Thus, at sufficiently low temperatures (corresponding to small coupling constants) and pure metals, the hierarchy

$$\gamma \ll 1/\tau \ll T, \tag{2}$$

is possible and will be used below. In standard low-temperature superconductors $\gamma \sim 10^7 - 10^9\ \mathrm{s}^{-1}$. Anticipating that the electron–phonon interaction is important for describing the properties of Y–Ba–Cu–O compounds, γ can be estimated[14] to be $10^{13}\ \mathrm{s}^{-1}$ in the case of optimal doping. This is determined by the proportionality $\gamma \propto T$ for $T > \omega_D/2\pi$ and $\gamma \propto T^3/\omega_D^3$ in the opposite limit. Since in this compound $\omega_D \sim 600$ K, the estimate $\gamma \sim T$ is valid in the case of optimal doping.[14] If the electron–phonon interaction is important, the hierarchy (2) can be realized only in the case of nonoptimal doping. The plasma frequency in a high-T_c superconductor is ~ 1 eV,[15] so that the contribution $\sim T^2/\epsilon_F$ to γ is much smaller.

We define $I(\epsilon)$ to be the rate of change of the number of quasiparticles, normalized to the density of states $N(0)$ in the normal metal, with energy ϵ and a prescribed direction of spin per unit volume. The problem of finding the rate I under the action of an electromagnetic pump is equivalent to the problem of finding the real part of the conductivity (i.e., the absorption) for the case of anisotropic superconductors in Ref. 16. The result obtained in Ref. 16 can be understood using a qualitative approach,[5] which leads to the Mattis–Bardeen formula in the case of ordinary superconductors.

The electromagnetic interaction Hamiltonian has the structure

$$H' \propto \sum_{\mathbf{p}} \mathbf{A}(\mathbf{p}+\mathbf{p}')a^{+}_{\mathbf{p}'}a_{\mathbf{p}}e^{-i\omega t} + \mathrm{c.c.} \tag{3}$$

Here $\mathbf{A}$ is the vector potential. The electric field $\mathbf{E}$ in the case of a cylindrical Fermi surface is assumed to be radial. Next, the expression (3) can be rewritten in terms of Bogolyubov quasiparticles, using $u_{\mathbf{p}}$ and $v_{\mathbf{p}}$ (if $\Delta(T) \gg 1/\tau$, these quasiparticles are well defined), and the standard coherence factors for the electromagnetic absorption can be used:[5]

$$\begin{aligned} l(\mathbf{p},\mathbf{p}') &= (\mathbf{p}+\mathbf{p}')_k(u_{\mathbf{p}}u^{*}_{\mathbf{p}'} + v^{*}_{\mathbf{p}}v_{\mathbf{p}'}), \\ f(\mathbf{p},\mathbf{p}') &= (\mathbf{p}+\mathbf{p}')_k(v_{\mathbf{p}}u^{*}_{\mathbf{p}'} - u^{*}_{\mathbf{p}}v_{\mathbf{p}'}). \end{aligned} \tag{4}$$

The subscript k denotes the direction of the electric field. The first coherence factor corresponds to processes without a change in the number of quasiparticles, and the second one refers to processes with a change in the number of quasiparticles by 2. Free electrons cannot absorb photons because of the impossibility of satisfying the laws of conservation of energy and momentum at the same time. The absorption of light is associated with scattering by impurities, and the coherence factors $|l|^2$ and $|f|^2$ must therefore be averaged over the Fermi surface (bearing in mind the momentum dependence of the density of states). Performing this procedure for quasiparticle states with energies ϵ and ϵ' (in so doing, states with momenta greater and less than the Fermi momenta must be taken into account), we arrive in the case of an isotropic superconductor at the possibility of the following substitutions for the coherence factors in the expressions for the absorption and emission probabilities:

$$\begin{aligned} |l|^2 &= (p_k^2 + p_k'^2)\left(1 + \frac{\Delta^2}{\epsilon\epsilon'}\right), \\ |f|^2 &= (p_k^2 + p_k'^2)\left(1 - \frac{\Delta^2}{\epsilon\epsilon'}\right). \end{aligned} \tag{5}$$

However, in the case of anisotropic pairing, when a symmetry transformation under which the gap changes sign exists,

$$|f|^2 = |l|^2 = (p_k^2 + p_k'^2). \tag{6}$$

We note that the term containing $p_k p_k'$ vanishes since $\langle p_k \Delta(\mathbf{p})\rangle = 0$, where the average is taken over the Fermi surface. This condition (which holds for singlet pairing) appears in Ref. 16 as a criterion for the possibility of neglecting renormalization of the electromagnetic interaction vertex as a result of scattering by impurities. As usual, in the calculation of transitions from the initial state i to a final state f, in the linear approximation in the pump power, the population n of the levels must be taken into account. The decrease in the distribution function as a result of absorption is proportional to $-n_i(1-n_f)$, and the increase due to absorption is proportional to $n_f(1-n_i)$, so that the effect as a whole will be proportional to $n_f - n_i$ (spontaneous emission can be neglected).

Thus, if the coherence factors (5) are used, we obtain in the isotropic case

$$\begin{aligned} I(\epsilon) = \frac{\sigma(\omega)\overline{E^2}}{2N(0)\omega^2}\, \nu_\epsilon \bigg[&\theta(\epsilon-\omega)\nu_{\epsilon-\omega}(n^0_{\epsilon-\omega} - n^0_\epsilon) \\ &\times\left(1 + \frac{\Delta^2}{\epsilon(\epsilon-\omega)}\right) - \nu_{\epsilon+\omega}(n^0_\epsilon - n^0_{\epsilon+\omega}) \\ &\times\left(1 + \frac{\Delta^2}{\epsilon(\epsilon+\omega)}\right) + \theta(\omega-\epsilon)\nu_{\omega-\epsilon}(1 - n^0_{\omega-\epsilon} - n^0_\epsilon) \\ &\times\left(1 - \frac{\Delta^2}{\epsilon(\omega-\epsilon)}\right)\bigg], \end{aligned} \tag{7}$$

which agrees with Ref. 1. Here $\sigma(\omega)$ is the conductivity of the normal metal, $\nu_\epsilon = \epsilon\theta(\epsilon-\Delta)(\epsilon^2-\Delta^2)^{-1/2}$ is the superconducting density of states normalized to $N(0)$, and $n^0_\epsilon = [1+\exp(\epsilon/T)]^{-1}$. The coefficient of proportionality was based on absorption in the normal state.

To derive I in the anisotropic case, it is convenient to introduce the definition

$$\beta(\epsilon,\epsilon') = [\langle \nu(\mathbf{p},\epsilon)p_k^2\rangle\langle \nu(\mathbf{p}',\epsilon')\rangle + \langle \nu(\mathbf{p},\epsilon)\rangle$$

$$\times\langle \nu(\mathbf{p}',\epsilon')p_k'^2\rangle]/2\langle p_k^2\rangle. \tag{8}$$

Here $\langle g\rangle = \int g d^2S/\int d^2S$, where the integration extends over the Fermi surface,

$$\nu(\mathbf{p},\epsilon)=\frac{\epsilon\theta(\epsilon-\Delta(\mathbf{p}))}{\sqrt{\epsilon^2-\Delta^2(\mathbf{p})}},$$

and the notation $\langle \nu(\mathbf{p},\epsilon)\rangle = \nu_\epsilon$ has been introduced. Using the definition (8) and the expression (6), we have in the anisotropic case

$$\begin{aligned} I=\frac{\sigma(\omega)\overline{E^2}}{2N(0)\omega^2}&[\theta(\epsilon-\omega)\beta(\epsilon-\omega,\epsilon)(n^0_{\epsilon-\omega}-n^0_\epsilon)\\ &-\beta(\epsilon+\omega,\epsilon)(n^0_\epsilon-n^0_{\epsilon+\omega})\\ &+\theta(\omega-\epsilon)\beta(\omega-\epsilon,\epsilon)(1-n^0_{\omega-\epsilon}-n^0_\epsilon)]. \end{aligned} \tag{9}$$

Defining $\beta(\pm\epsilon,\pm\epsilon')=\beta(\epsilon,\epsilon')$, the expression for the real part of the conductivity in an anisotropic superconductor can be written, in agreement with Ref. 16 as,

$$\frac{\sigma^S(\omega)}{\sigma(\omega)}=\frac{1}{\omega}\int_{-\infty}^{\infty}d\epsilon\beta(\epsilon-\omega,\epsilon)(n^0_{\epsilon-\omega}-n^0_\epsilon). \tag{10}$$

It is easy to show that I can be characterized by the energy scales ω and Δ. The τ approximation for the collision integral is valid in the case when the variation of the equilibrium distribution function is localized in an energy interval much less than T near the Fermi surface.[1,4] The change in the distribution function enters into the collision integral either as a factor or in the integrand. But since the characteristic integration scale is T, the integrated perturbation of the distribution function has the additional smallness $\sim\omega/T$, Δ/T as compared with the contributions to the collision integral where the change in the distribution function enters as a multiplicative factor. Since the reciprocal γ of the energy relaxation time is much less than the reciprocal of the impurity scattering time, γ does not depend on the direction of the momentum at the Fermi surface. Likewise, coherence factors do not enter into the expression for γ, since particles with energy $\sim T$ for which $\Delta^2/\epsilon\epsilon'\approx 0$ participate in energy transfer. Thus, the change δn_ϵ in the distribution function is

$$\gamma\delta n_\epsilon\nu_\epsilon = I. \tag{11}$$

Substituting δn_ϵ into the self-consistency equation

$$\Delta(\mathbf{p})=-\frac{1}{2}\sum_{\mathbf{p}'}V(\mathbf{p},\mathbf{p}')\Delta(\mathbf{p}')\frac{1-2n_{\epsilon'}-2\delta n_{\epsilon'}}{\epsilon_{\mathbf{p}'}}, \tag{12}$$

we obtain

$$\begin{aligned} &\frac{T_c-T}{T_c}\langle|\Delta(\mathbf{p})|^2\rangle-\frac{7\zeta(3)}{8\pi^2T^2}\langle|\Delta(\mathbf{p})|^4\rangle\\ &\quad-2\int_0^\infty(\epsilon\nu_\epsilon-\Phi_\epsilon)\delta n_\epsilon d\epsilon=0. \end{aligned} \tag{13}$$

In Eq. (12), $\epsilon_{\mathbf{p}}=(\xi^2+|\Delta(\mathbf{p})|^2)^{1/2}$ (here $\xi=v_F|p-p_F|$) and $V(\mathbf{p},\mathbf{p}')$ is the pairing potential, and in (13) we have introduced the notation

$$\Phi_\epsilon=\int_0^\epsilon \nu_{\epsilon'}d\epsilon'.$$

The derivation of Eq. (13) employed the equality $\langle|\Delta|^2/\xi\rangle=\epsilon\nu_\epsilon-\Phi_\epsilon$, where the averaging occurs over a level with fixed energy ϵ. The expression (13) has the same form as the ordinary Ginzburg–Landau equation with an additional nonequilibrium term. The dynamical contribution $A^2|\Delta|^2$ can be neglected if $\omega\tau\ll T_c/\gamma$ (it is assumed that $\omega\tau\gg 1$). The latter condition does not impose any additional constraints, since it is a consequence of the preceding conditions $\omega\ll T_c$ and $\tau\ll 1/\gamma$.

The function $\beta(\epsilon,\epsilon')$ simplifies in the case of a cylindrical Fermi surface, if a rotation by an angle $\pi/2$ around the axis of the cylinder does not change the magnitude of the gap. This happens, for example, when the order parameter transforms according to a nontrivial irreducible representation of the group D_{4h}. Then

$$\beta(\epsilon,\epsilon')=\nu_\epsilon\nu_{\epsilon'}. \tag{14}$$

Since the characteristic values of the energy in Eq. (9) are much smaller than T, the equilibrium distribution function can be expanded and we arrive at the following equation:

$$\frac{T_c-T}{T_c}-b\Delta_0^2(T)+\kappa F\left(\frac{\Delta_0(T)}{\omega}\right)=0. \tag{15}$$

where $\Delta_0(T)$ is the amplitude of the gap, i.e., $\Delta(\mathbf{p},T)=\Delta_0(T)\phi(\mathbf{p})$ $(|\phi(\mathbf{p})|\leqslant 1)$. The function F is given by

$$F\left(\frac{\Delta_0}{\omega}\right)=\frac{1}{\langle|\Delta(\mathbf{p})|^2\rangle}\int_0^\infty(\epsilon\nu_\epsilon-\Phi_\epsilon)(\nu_{\epsilon+\omega}-\nu_{\epsilon-\omega})d\epsilon, \tag{16}$$

where $\nu_{-\epsilon}=\nu_\epsilon$ by definition, and

$$\kappa=\frac{\sigma(\omega)\overline{E^2}}{4\gamma\omega TN(0)}. \tag{17}$$

Thus, the frequency dependence of the nonequilibrium term is determined by an integral transformation of the density of states. The condition $\omega\gamma/\gamma'\ll T$ must be satisfied in order that the suppression of superconductivity by heating of the superconducting film not exceed stimulation (15). Here γ' is the reciprocal of the energy relaxation time under the conditions of heat exchange with the heat bath, and, generally speaking, can be less than γ.

For large x

$$F(x)=\frac{1}{x\langle|\phi^2(\mathbf{p})|\rangle}\int_0^\infty(\nu_y-1)^2dy. \tag{18}$$

In this expression the density of states is represented as a function of the dimensionless variable $y=\epsilon/\Delta_0$. Furthermore, the equality

$$\int_0^\infty(\nu_y-1)dy=0$$

was used.

The integral in Eq. (18) converges in the case of isotropic pairing, but diverges if the gap is isotropic, so that for

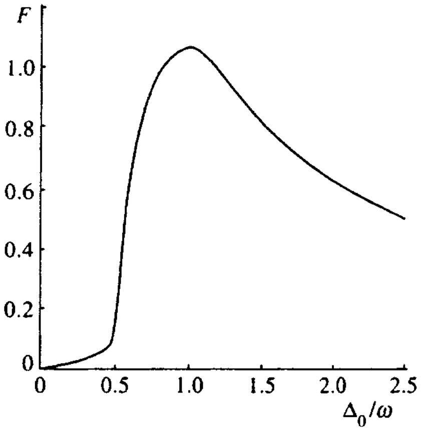


FIG. 1. Nonequilibrium term F (16) in a superconductor with a cylindrical Fermi surface and $\Delta(\varphi)=\Delta_0 \cos 2\varphi$.

ordinary superconductors the asymptotic behavior is different: $F(x) \propto \ln x/x$. For small values of x we have to logarithmic accuracy

$$F(x)=\frac{1}{2}\langle|\phi^2(\mathbf{p})|\rangle x^2 \ln\frac{1}{x}, \tag{19}$$

while in ordinary superconductors $F(x) \propto x$.

This difference is due to the different values of the averaged coherence factors $|l|^2$ and $|f|^2$ in the anisotropic case. It can be shown that the asymptotic behavior $\propto 1/x$ for $x \gg 1$ and $\propto x^2 \ln(1/x)$ for $x \ll 1$ remains in the general anisotropic case, if there exists a symmetry transformation under which the order parameter changes sign. Thus, for small x, the nonequilibrium term in the Ginzburg–Landau equation is much smaller than in the isotropic case.

The function $F(x)$ for a cylindrical Fermi surface and for $\Delta(\varphi)=\Delta_0 \cos 2\varphi$ is displayed in Fig. 1. In this figure, $F(x)$ decreases much more rapidly for $x<x_{\max}$ than in the isotropic case. We also note that in the isotropic case the stimulation effect is maximum for $\omega=2\Delta_0$, whereas in the special case at hand this occurs at $\omega=\Delta_0$. At the same time, the maximum of the differential conductivity dj/dV in superconductor–superconductor tunneling junctions in both anisotropic and isotropic cases occurs at $eV=2\Delta_0$ for both diffuse[17] and specular[13] reflection at the boundary.

The stimulation effect can probably be studied experimentally, as in ordinary superconductors, by measuring the critical field of a film at temperatures below T_c. Since the film thickness in the anisotropic case must be greater than $\xi(T)$, metastability makes measurements of the stimulation effect above T_c inopportune.

We now examine stimulation of superconductivity by a powerful sound wave. The deformation potential describing the interaction with longitudinal sound (see, for example, Ref. 18) is of a scalar character and depends on the magnitude of the transferred momentum. If the wavelength of sound is much greater than the electron mean free path (we note that the mean free path in high-T_c superconductors is short, and this ratio can be assumed to hold), then sound absorption is of a diffuse character[18] and is accompanied by scattering by impurities.

The coherence factors characterizing sound damping are different from those for electromagnetic absorption. They are determined by the expressions[5]

$$q(\mathbf{p},\mathbf{p}')=u_{\mathbf{p}}u^*_{\mathbf{p}'}-v^*_{\mathbf{p}}v_{\mathbf{p}'},$$

$$m(\mathbf{p},\mathbf{p}')=v_{\mathbf{p}}u^*_{\mathbf{p}'}+u^*_{\mathbf{p}}v_{\mathbf{p}'}. \tag{20}$$

The first factor, once again, corresponds to processes without a change in the number of quasiparticles, and the second factor corresponds to processes with a change in this number by 2. After averaging over the Fermi surface

$$|q|^2=|m|^2=1, \tag{21}$$

in the anisotropic case and

$$|q|^2=1-\frac{\Delta^2}{\epsilon\epsilon'}, \tag{22}$$

in the isotropic case

$$|m|^2=1+\frac{\Delta^2}{\epsilon\epsilon'}.$$

In the acoustic pump I in the isotropic case, besides the coefficient of proportionality, the sign of the terms containing Δ^2 in Eq. (7) must therefore be changed. For anisotropic superconductors the stimulation effect can be described, once again, by (15) and (16) with κ replaced by κ'

$$\kappa'=\frac{W}{4\gamma\omega TN(0)}, \tag{23}$$

where W is the absorbed sound power per unit volume.

In the isotropic case[1] with $\omega>2\Delta_0$ the stimulation effect decreases abruptly and becomes negative, whereas in the anisotropic case it is positive over the entire frequency range studied. This is due to the different values of $|q|^2$ and $|m|^2$ in the anisotropic case.

3. TUNNELING STIMULATION

In the Ginzburg–Landau equation, the nonequilibrium term describing the previously studied electromagnetic and acoustic stimulation is characterized by the parameter Δ_0/ω. In the case of stimulation by a tunneling current, the magnitude of the effect is determined by two parameters: Δ_0/eV and Δ_0/Δ_0', where eV is the voltage applied to the junction and Δ_0' is the superconducting gap in the superconductor with which the experimental sample forms a junction. This makes for additional experimental possibilities. A description of the measurement scheme is given in Ref. 1.

In the anisotropic case, a substantial simplification due to the possibility of relaxation of the electron–hole imbalance by the method of impurity scattering arises.[19] If the superconducting gap changes sign at the Fermi surface, then the reciprocal of the imbalance relaxation time for $\epsilon\sim\Delta_0$ can be estimated to be $\gamma_{\text{imb}}\sim 1/\tau$. When the hierarchy (2) holds, the pumping of the electron and hole branches at fixed ϵ must therefore be averaged in order to find $I(\epsilon)$. The function $I(\epsilon)$ will thereby have the same structure as in the symmetric scheme (see Fig. 1) of stimulation of isotropic superconductors:

$$I(\epsilon)=\frac{\mathscr{A}}{2\mathscr{V}N(0)}\int_{v_x>0}\frac{d^2S}{(2\pi)^3v_F}\,v_xD(\hat{\mathbf{p}})$$
$$\times[\nu(\mathbf{p},\epsilon)\nu'(\mathbf{p},\epsilon-eV)(n'_{\epsilon-eV}-n_\epsilon)$$
$$+\nu(\mathbf{p},\epsilon)\nu'(\mathbf{p},\epsilon+eV)(n'_{\epsilon+eV}-n_\epsilon)]. \qquad (24)$$

The plane of the tunneling junction is assumed to be perpendicular to the x axis, and the integration in Eq. (24) extends over the half of the Fermi surface with $v_x>0$. Primed quantities refer to the second superconductor forming the junction, $\mathscr{A}$ is the area of the junction, $D(\hat{\mathbf{p}})$ is the transmission coefficient of the tunneling barrier for an electron with momentum $\hat{\mathbf{p}}$, and $\mathscr{V}$ is the volume of the stimulated sample. The factor $v_xD(\hat{\mathbf{p}})$ follows from a comparison with the microscopic expression for the tunneling current,[20] but it has a transparent physical meaning, since the probability of tunneling is proportional to $D(\hat{\mathbf{p}})$ and the electron–barrier collision frequency ($\propto v_x$). Since I is an even function of eV, for convenience we assume below that $eV>0$.

Once again, the distribution function in Eq. (24) can be expanded if Δ, Δ', $eV\ll T$ and the expression for the normal resistance of the contact is taken into account:

$$\frac{1}{R_N}=2e^2\int_{v_x>0}\frac{d^2S}{(2\pi)^3v_F}\,v_xD(\hat{\mathbf{p}})\mathscr{A}. \qquad (25)$$

As a result, we arrive at the following expression for the nonequilibrium term in the Ginzburg–Landau equation:

$$F=\frac{\kappa^t}{\langle|\Delta(\mathbf{p})|^2\rangle}\int_0^\infty\frac{\epsilon\nu_\epsilon-\Phi_\epsilon}{\nu_\epsilon}\,d\epsilon$$
$$\times\int_{v_x>0}d^2Sv_xD(\hat{\mathbf{p}})\nu(\mathbf{p},\epsilon)[\nu'(\mathbf{p},\epsilon+eV)$$
$$-\nu'(\mathbf{p},\epsilon-eV)]\left[\int_{v_x>0}d^2Sv_xD(\hat{\mathbf{p}})\right]^{-1}. \qquad (26)$$

Here the dimensionless quantity κ^t is defined as

$$\kappa^t=\frac{V}{8eR_NT\mathscr{V}\gamma N(0)}. \qquad (27)$$

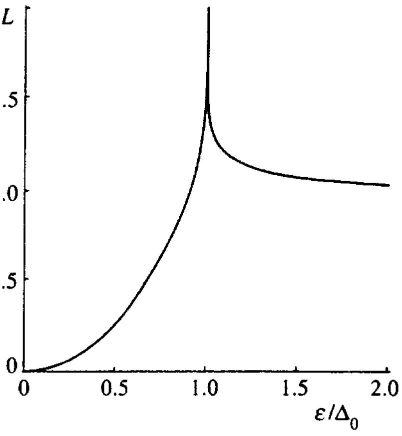


FIG. 2. $L(\epsilon)$ (30) in a superconductor with a cylindrical Fermi surface and $\Delta(\varphi)=\Delta_0\cos 2\varphi$.

For wide insulating interlayers, the transmission factor makes it necessary to take account of only the electrons whose velocity lies in a narrow cone near the normal to the plane of the junction. In this case the nonequilibrium term simplifies,

$$F=\kappa^t\int_{\Delta_+}^\infty\frac{L(\epsilon)}{\sqrt{\epsilon^2-\Delta_+^2}}[\nu_{is}(\Delta'_+,\epsilon+eV)$$
$$-\nu_{is}(\Delta'_+,\epsilon-eV)]d\epsilon. \qquad (28)$$

In this expression Δ_+, Δ'_+ are the superconducting gaps in the superconductors for momenta normal to the junction, and ν_{is} is the density of states in the isotropic superconductor:

$$\nu_{is}(\Delta,\epsilon)=\frac{|\epsilon|}{\sqrt{\epsilon^2-|\Delta|^2}}\,\theta(|\epsilon|-|\Delta|). \qquad (29)$$

The dimensionless function $L(\epsilon)$ has the form

$$L(\epsilon)=\frac{\langle|\Delta(\mathbf{p})|^2/\xi(\mathbf{p},\epsilon)\rangle}{\langle|\Delta(\mathbf{p})|^2\rangle\langle1/\xi(\mathbf{p},\epsilon)\rangle}=\frac{\epsilon^2\nu_\epsilon-\epsilon\Phi_\epsilon}{\langle|\Delta(\mathbf{p})|^2\rangle\nu_\epsilon}. \qquad (30)$$

Here $1/\xi(\mathbf{p},\epsilon)=\theta(|\epsilon|-|\Delta(\mathbf{p})|)/\sqrt{\epsilon^2-|\Delta(\mathbf{p})|^2}$. The function $L(\epsilon)$ approaches 1 for large ϵ. For the previously examined cylindrical Fermi surface and gap $\Delta(\varphi)=\Delta_0\cos 2\varphi$, the function $L(\epsilon)$ can be expressed in terms of the dimensionless variable $x=\epsilon/\Delta_0$ as follows:

$$L(x)=\begin{cases}2[K(x)-E(x)]/K(x), & x\leqslant1,\\ 2x^2[K(1/x)-E(1/x)]/K(1/x), & x>1.\end{cases} \qquad (31)$$

Here $E,\ K(x)=\int_0^{\pi/2}(1-x^2\cos^2\alpha)^{\pm1/2}d\alpha$ are complete elliptic integrals. The function $L(\epsilon)$ is displayed in Fig. 2.

The stimulation of superconductivity as described by Eq. (28) exhibits many of the characteristics of the isotropic case. A logarithmic singularity occurs at the voltages $eV=||\Delta_+|-|\Delta'_+||$, and the stimulation effect decreases abruptly at $eV=|\Delta_+|+|\Delta'_+|$. More accurately, for $|\Delta'_+|>|\Delta_+|$ and $eV\approx|\Delta'_+|-|\Delta_+|$ we have to logarithmic accuracy

$$F=\frac{\kappa^tL(eV)}{2}\sqrt{\frac{|\Delta'_+|}{|\Delta_+|}}\ln\frac{|\Delta_+|}{|eV-|\Delta'_+|+|\Delta_+||}. \qquad (32)$$

For a different ratio of the gaps, when $|\Delta'_+|<|\Delta_+|$ and $eV\approx|\Delta_+|-|\Delta'_+|$,

$$F=-\frac{\kappa^tL(eV)}{2}\sqrt{\frac{|\Delta'_+|}{|\Delta_+|}}\ln\frac{|\Delta'_+|}{|eV-|\Delta_+|+|\Delta'_+||}, \qquad (33)$$

i.e., the stimulation effect is negative. If $eV\lesssim|\Delta_+|+|\Delta'_+|$, the stimulation effect is positive in both cases, but at $eV=|\Delta'_+|+|\Delta_+|$ it decreases abruptly by an amount

$$\Delta F=\frac{\pi}{2}\kappa^tL(eV)\sqrt{\frac{|\Delta'_+|}{|\Delta_+|}}. \qquad (34)$$

The most important difference from the isotropic case shows up in the asymptotic stimulation effect at high voltages. While in an isotropic superconductor the stimulation effect vanishes above the threshold voltage $eV=|\Delta'_+|+|\Delta_+|$, in the anisotropic case it decreases at high voltages

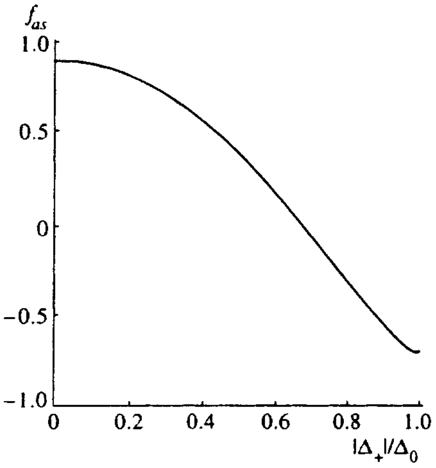


FIG. 3. Asymptotic behavior of $f_{as}(x)$ (36) in a superconductor with a cylindrical Fermi surface and $\Delta(\varphi)=\Delta_0\cos 2\varphi$.

$\propto 1/V^2$. The proportionality coefficient can be positive or negative, depending on Δ_+, i.e., on the orientation of the junction plane relative to the crystal axes.

It is convenient to write the expression for the decrease in F at high voltages, introducing the complex function $\mathscr{L}(\epsilon)$ defined by the first relation in Eqs. (30) with the substitution

$$\frac{1}{\xi(\mathbf{p},\epsilon)}\to\frac{\theta(|\epsilon|-|\Delta(\mathbf{p})|)}{\sqrt{\epsilon^2-|\Delta(\mathbf{p})|^2}}-i\,\frac{\theta(|\Delta(\mathbf{p})|-|\epsilon|)}{\sqrt{|\Delta(\mathbf{p})|^2-\epsilon^2}}. \tag{35}$$

With this definition we have for $eV\gg|\Delta_+|+|\Delta'_+|$

$$F=\frac{2\kappa^t|\Delta'_+|^2}{(eV)^3}\left[\int_0^{|\Delta_+|}\frac{\epsilon\,\mathrm{Im}\,\mathscr{L}}{\sqrt{|\Delta_+|^2-\epsilon^2}}\,d\epsilon+\int_{|\Delta_+|}^{\Delta_0}\frac{\epsilon(\mathrm{Re}\,\mathscr{L}-L)}{\sqrt{\epsilon^2-|\Delta_+|^2}}\,d\epsilon\right]=\frac{\kappa^t|\Delta'_+|^2\Delta_0}{(eV)^3}\,f_{as}\left(\frac{|\Delta_+|}{\Delta_0}\right). \tag{36}$$

Recall that κ^t in (27) contains the first power of V, and Δ_0 is the maximum magnitude of $\Delta(\mathbf{p})$. It is easy to show that $\mathrm{Re}\,\mathscr{L}-L>0$ and $\mathrm{Im}\,\mathscr{L}<0$. If the magnitude of the superconducting gap for momentum normal to the junction plane is close to its maximum value, then the stimulation effect will be negative at high voltages, and if it is close to the minimum value, then conversely the nonequilibrium term at high voltages will be positive. This assertion holds for the general case of anisotropic pairing, and not just for the special case studied in the present work.

For a cylindrical Fermi surface and $\Delta(\varphi)=\Delta_0\cos 2\varphi$, the use of the Legendre relation for the elliptic integrals gives

$$\mathrm{Im}\,\mathscr{L}(\epsilon)=-\frac{\pi}{K^2(\epsilon)+K^2(\sqrt{1-\epsilon^2})},$$

$$\mathrm{Re}\,\mathscr{L}(\epsilon)-L(\epsilon)=\frac{\pi K(\sqrt{1-\epsilon^2})}{K(\epsilon)[K^2(\epsilon)+K^2(\sqrt{1-\epsilon^2})]}. \tag{37}$$

The function $f_{as}(x)$ for this case is presented in Fig. 3.

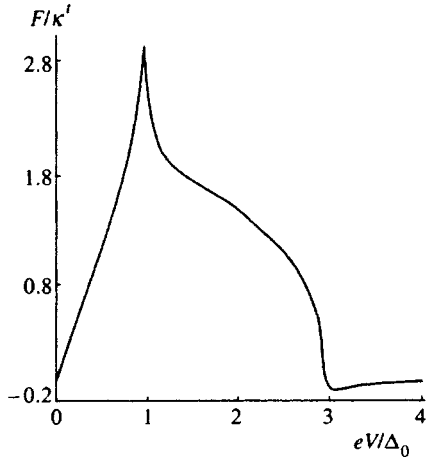


FIG. 4. Nonequilibrium term F/κ^t (26) for a contact of superconductors with cylindrical Fermi surfaces, $\Delta(\varphi)=\Delta_0\cos 2\varphi$, $\Delta'(\varphi)=2\Delta_0\cos 2\varphi$, and model transmittance $D(\hat{\mathbf{p}})=D_0p_x^2/p_F^2$.

For a thin insulating interlayer, the model transmittance $D(\hat{\mathbf{p}})=D_0p_x^2/p_F^2$ can be used. This, together with the factor v_x, have the effect that states with momenta lying in directions close to the normal to the junction plane appear with high weight in the expression for the nonequilibrium term.

The main features of the previously studied dependence of F on the voltage and the gaps in the superconductors forming the junction therefore remain the same. The logarithmic divergences (32) and (33) are smoothed in a natural manner, and lead to finite expressions. Figure 4 shows the nonequilibrium term for this model transmittance and cylindrical Fermi surfaces of superconductors whose gaps are given by the expressions $\Delta(\varphi)=\Delta_0\cos 2\varphi$ and $\Delta'(\varphi)=2\Delta_0\cos 2\varphi$ (the angle is measured from the x axis). The stimulation effect is strongest for $eV=|\Delta'_+|-|\Delta_+|=\Delta_0$ and drops sharply for $eV=|\Delta'_+|+|\Delta_+|=3\Delta_0$. For high voltages the nonequilibrium term is negative, since $|\Delta_+|=\Delta_0$.

In the particular configuration studied here, there is no suppression of the anisotropic order parameter near the plane of the tunneling junction if the reflection at the boundary is specular. Such suppression does occur in the general case (see the boundary conditions in Ref. 13). Since the change in the order parameter is determined by the scale of the temperature-dependent coherence length, and the expression for the current through the contact is characterized by nonlocality of the order of the coherence length at low temperature,[13] to calculate the pump I it is necessary to substitute into Eq. (24) the density of states near the junction plane. Specifically, the quantity Δ_+ in Eq. (28) must be calculated taking account of this suppression. Of course, $L(\epsilon)$ need not be changed, since this function describes the volume properties of the superconductor.

For a wide range of orientations near T_c, the gap at the boundary can be set equal to zero. The stimulation effect, as follows, for example, from Eq. (36), does not vanish in so doing if $\Delta'_+\neq 0$, which happens for certain orientations of the junction or if the stimulating superconductor is isotropic.

4. CONCLUSIONS

In the present paper the characteristic features of electromagnetic, acoustic, and tunneling stimulation in anisotropic superconductors were found. In the Ginzburg–Landau equation, the nonequilibrium term (15) described by the expressions (9) and (16) is characterized by a sharper maximum than in the isotropic case. Acoustic stimulation as a function of the frequency of the incident longitudinal sound wave is once again determined by the expression (16) and is positive in the entire frequency range studied, while for isotropic superconductors it is negative for $\omega > 2\Delta_0$. These differences are due to the different values of the averaged coherence factors in anisotropic superconductors. The nonequilibrium term accompanying tunneling stimulation depends on the orientation of the junction plane relative to the crystal axes (see Eqs. (32)–(34)) by virtue of the dependence of the electron tunneling probability on the direction of the momentum. In contrast to the isotropic case, it does not equal zero at high voltage (36). Its sign at high voltage also depends on the orientation.

I thank Yu. S. Barash and E. G. Maksimov for helpful discussions. This work was supported by the Russian Fund for Fundamental Research under grant No. 96-02-16249. I am also grateful to Forschungszentrum Jülich GmbH.

[1] G. M. Eliashberg and B. I. Ivlev, in *Nonequilibrium Superconductivity*, D. N. Langenberg and A. I. Larkin (eds.), North-Holland, Amsterdam (1986), p. 211.

[2] A. M. Gulyan and G. F. Zharkov, *Superconductors in External Fields (Nonequilibrium Phenomena)* [in Russian], Nauka, Moscow (1990).

[3] G. M. Éliashberg, JETP Lett. **11**, 114 (1970).

[4] G. M. Éliashberg, Zh. Éksp. Teor. Fiz. **61**, 1254 (1971) [Sov. Phys. JETP **34**, 668 (1972)].

[5] J. Schrieffer, *Theory of Superconductivity*, W. A. Benjamin, New York (1964).

[6] C. J. Pethick and D. Pines, Phys. Rev. Lett. **57**, 118 (1986).

[7] M. Sigrist and K. Ueda, Rev. Mod. Phys. **63**, 239 (1991).

[8] D. A. Wollman, D. J. van Harlingen, W. C. Lee *et al.*, Phys. Rev. Lett. **71**, 2134 (1993).

[9] D. A. Crawner and H. R. Ott, Phys. Rev. B **50**, 6530 (1994).

[10] C. C. Tsuei, J. R. Kirtley, C. C. Chi *et al.*, Phys. Rev. Lett. **73**, 593 (1994).

[11] W. Braunisch, N. Knauf, G. Bauer *et al.*, Phys. Rev. Lett. **68**, 1908 (1992); Phys. Rev. B **48**, 4030 (1993).

[12] E. A. Shapoval, Zh. Éksp. Teor. Fiz. **88**, 1073 (1985) [Sov. Phys. JETP **61**, 630 (1985)].

[13] Yu. S. Barash, A. V. Galaktionov, and A. D. Zaikin, Phys. Rev. B **52**, 665 (1995).

[14] S. V. Shulga, O. V. Dolgov, and E. G. Maksimov, Physica C **178**, 266 (1991).

[15] T. Timusk and D. B. Tanner, in *Physical Properties of High Temperature Superconductors*, D. M. Ginsberg (ed.), World Scientific, Singapore (1989), Vol. 1, p. 339; Vol. 3, 1992, p. 363.

[16] P. J. Hirshfeld, P. Wolfle, J. A. Sauls *et al.*, Phys. Rev. B **40**, 6695 (1989).

[17] H. Won and K. Maki, Phys. Rev. B **49**, 1397 (1994).

[18] C. Kittel, *Introduction to Solid State Physics*, Wiley, 4th ed., New York (1971).

[19] A. G. Aronov, Yu. M. Gal'perin, V. L. Gurevich, and V. I. Kozub, in *Nonequilibrium Superconductivity*, D. N. Langenberg and A. I. Larkin (eds.), North-Holland, Amsterdam (1986), p. 325.

[20] A. V. Zaĭtsev, Zh. Éksp. Teor. Fiz. **86**, 1742 (1984) [Sov. Phys. JETP **59**, 1015 (1984)].

Translated by M. E. Alferieff

Photoinduced phase transitions in a system with a transformable electron spectrum

A. L. Semenov[1)]

Ul'yanovsk Branch of the M. V. Lomonosov Moscow State University, 432700 Ul'yanovsk, Russia

(Submitted 17 November 1996)
Zh. Éksp. Teor. Fiz. **111**, 2147–2157 (June 1997)

An electron theory of photoinduced phase transitions and cavityless optical bistability in a light field with a finite optical-spectrum width is constructed. The Liouville equation for the density matrix is used to derive the existence criteria and calculate the main characteristics of the given phenomena. Broadening of the optical spectrum is shown to reduce the possibility of observing critical features (the existence criterion becomes more stringent and the hysteresis-loop area becomes smaller). Finally, results are compared with experimental data for CdS and amorphous GeS_2. © *1997 American Institute of Physics.* [S1063-7761(97)01606-5]

1. INTRODUCTION

A photoinduced transition in a system with a transformable electron spectrum manifests itself in sudden jumps in the band gap of the electron spectrum and the electron concentration in the conduction band of the semiconductor when the incident-light intensity reaches a critical value. The optical properties of the substance also change suddenly in the process. Near a phase transition there is cavityless optical bistability with increasing absorption.[1]

The sudden variation in the properties of the electron–phonon system of a solid can be initiated, as is known, by varying such parameters as pressure,[2] temperature,[2,3] the concentration of alloy impurities,[4–6] and the concentration of molecules adsorbed at the surface.[7] All these phase transitions occur between states that are in thermodynamic equilibrium, while a photoinduced phase transition is substantially nonequilibrium.

There are many mechanisms that could trigger a transformation of the electron spectrum in a light field: the exciton–exciton interaction and screening by electron–hole plasma in CdS (Refs. 8–10), the interaction of electrons and a static phonon mode at the edge of the Brillouin zone in VO_2 (Refs. 2, 12, and 12), the electron–defect interaction in amorphous GeS_2 (Ref. 13), and others.[1]

However, despite the fact that there are so many different mechanisms, all the photoinduced phase transitions in such substances have common features. In particular, it has been experimentally established that the frequency of the incident light triggering the phase transition is bounded from above.[8,12,13] For instance, in CdS and GeS_2 the frequency must be lower than the lower-edge optical-transition frequency at least by some fixed quantity (different for different substances).[8,13] From below the frequency of the light is bounded due to the saturation of optical band-to-band transitions. Thus, photoinduced phase transitions and cavityless optical bistability of the electron type are selective.[12]

The present paper develops a model of photoinduced phase transitions and cavityless optical bistability based on the fact that the electron spectrum of the system depends on the electron concentration in the conduction band. When the system is irradiated by quasimonochromatic light whose carrier frequency is lower than the lower-edge optical-transition frequency, positive feedback emerges in the system. Initially a nonresonance light field generates, due to the smearing of the lower edge of the optical transition, a small number of optical transitions of electrons to the conduction band, which leads to a slight decrease in the band gap. As a result, electron–phonon coupling becomes more resonant and hence stronger, which in turn leads to a further decrease in the band gap. Thus, positive feedback lowers the stability of the system. If positive feedback is strong, stability can be lost, and there is a sudden transition to a new state of equilibrium (a phase transition).

The phenomena of a photoinduced phase transition and cavityless optical bistability based on the rapid increase in the effectiveness of the interaction of a subband light field and a two-level system as the level-to-level transition frequency lowers have been examined by Andreev *et al.*[14] They assumed, however, that the electron spectrum has a sharp lower edge for optical transitions and that the overlap of the spectra of the electron subsystem and the light field is due to the finite width of the latter. In the present paper the smearing of the lower edge of optical transitions in the electron subsystem is the decisive factor, while the width of the optical spectrum of the incident light may be infinitesimal. The finiteness of this width leads, as we will shortly see, to more stringent conditions (in comparison to the case where the width of the incident spectrum is zero, $\Delta\omega=0$) for the existence of a photoinduced phase transition and cavityless optical bistability, conditions that constrain the parameters of the electron system and the carrier frequency of the light field.

2. BASIC EQUATIONS

The interaction of the electron subsystem and the light field is described by the Liouville equation[15]

$$i\hbar\,\frac{\partial\rho}{\partial t}=[H+V,\rho],\tag{1}$$

where ρ is the electron density matrix of the medium, H is the Hamiltonian of the electron system in the absence of light, and V is the perturbation operator, which in the dipole approximation has the form

$$V=-\mathbf{d}\cdot\mathbf{E}(t)=-\mathbf{d}\cdot\int\mathbf{E}_{\omega}e^{-i\omega t}d\omega.\tag{2}$$

Here $\mathbf{d}$ is the dipole moment operator, and $\mathbf{E}_\omega$ and ω are the amplitude and frequency of a spectral component of the light field.

Let us examine the case where the incident light $\mathbf{E}(t)$ is a linearly polarized, quasimonochromatic, stationary random field.[16] Then all the spectral components $\mathbf{E}_\omega$ are statistically independent:[16,17]

$$\langle \mathbf{E}_\omega \cdot \mathbf{E}_{\omega_1} \rangle = G(\omega)\delta(\omega+\omega_1). \tag{3}$$

Here $G(\omega)$ is the spectral density of the light field, which for a quasimonochromatic signal can be written as[16]

$$G(\omega) - Ig(|\omega| - \omega_0), \tag{4}$$

where ω_0 is the carrier frequency, and $g(x)$ is an even nonnegative bell-shaped function with its maximum at $x=0$, with the normalization condition $\int g(x)dx=1$. The halfwidth $\Delta\omega$ of the spectral density $G(\omega)$ satisfies $\Delta\omega \ll \omega_0$. $I=\int G(\omega)d\omega/2$ is the intensity of the light field (in Gaussian units, to within a factor $c\sqrt{\varepsilon}/2\pi$, where c is the speed of light and $\sqrt{\varepsilon}$ is the medium's refractive index).

Combining Eq. (1) with (2) and (3), we arrive at an equation for the diagonal elements ρ_{kk} of the density matrix ρ of the electron subsystem in second-order perturbation theory:

$$\frac{\partial \rho_{kk}}{\partial t} = \frac{2\pi}{\hbar^2} \sum_s |\mathbf{d}_{ks}|^2 G\left(\frac{\varepsilon_s - \varepsilon_k}{\hbar} \right)(\rho_{ss} - \rho_{kk}). \tag{5}$$

Here $\mathbf{d}_{ks}$ is the matrix element of the dipole moment operator, ε_s is the spectrum of the operator H, and $s=(\mathbf{s},j)$ is a parameter characterizing a single-electron quantum state with quasiwavevector $\mathbf{s}$ and band number j. In the special case of a monochromatic light field

$$E(t) = E_0 \cos(\omega_0 t + \varphi)$$

with a uniformly distributed random phase φ, the spectral density is given by

$$G(\omega) = \frac{E_0^2}{4} [\delta(\omega - \omega_0) + \delta(\omega + \omega_0)].$$

Then Eq. (5) becomes the well-known Fermi Golden Rule for the probability of stimulated transitions.[17]

Let us examine the case where only direct transitions are allowed between the valence band $(j=1)$ and the conduction band $(j=2)$. Then the dipole-moment matrix element $\mathbf{d}_{ks}$, where $k=(\mathbf{k},1)$ and $s=(\mathbf{s},2)$, assumes the form

$$|\mathbf{d}_{ks}| = d_{\mathbf{k}} \delta_{\mathbf{k},\mathbf{s}}, \tag{6}$$

where $\delta_{\mathbf{k},\mathbf{s}}$ is the Kronecker delta. By combining Eq. (5) with (6) we arrive at a formula describing the variation of the population $n_{\mathbf{k}} = \rho_{(\mathbf{k},2);(\mathbf{k},2)}$ of the $\mathbf{k}$th level in the conduction band due to stimulated optical transitions:

$$\frac{\partial n_{\mathbf{k}}}{\partial t} = \frac{2\pi}{\hbar^2} d_{\mathbf{k}}^2 G(\omega_{\mathbf{k}}), \tag{7}$$

where $\omega_{\mathbf{k}} = (\varepsilon_{\mathbf{k},2} - \varepsilon_{\mathbf{k},1})/\hbar$ is the electron excitation spectrum for direct optical band-to-band transitions. In writing (7) we also assumed that optical transitions at the frequency $\omega_{\mathbf{k}}$ are far from being saturated:

$$\rho_{(\mathbf{k},1);(\mathbf{k},1)} - \rho_{(\mathbf{k},2);(\mathbf{k},2)} = 1. \tag{8}$$

Plugging (4) into (7), we arrive at a kinetic equation for the electron concentration $n = \Sigma_{\mathbf{k}} n_{\mathbf{k}}$ in the conduction band:

$$\frac{\partial n}{\partial t} = \frac{2\pi I}{\hbar^2} \sum_{\mathbf{k}} d_{\mathbf{k}}^2 g(|\omega_{\mathbf{k}}| - \omega_0) - \frac{n}{\tau}. \tag{9}$$

The second term on the right-hand side of Eq. (9), which was introduced phenomenologically, describes band-to-band relaxation to the equilibrium value $n=0$ with a relaxation time τ.

If we ignore the variation in $d_{\mathbf{k}}$ near $\omega_{\mathbf{k}} = \omega_0$ $(d_{\mathbf{k}} \simeq d)$, then in the stationary mode $(\partial n/\partial t = 0)$ Eq. (9) yields

$$I = \frac{\hbar n}{2\pi\tau d^2 S(\omega_0, n)}, \tag{10}$$

where

$$S = S(\omega_0, n) = \int g(|\omega| - \omega_0)\nu(\omega, n)d\omega, \tag{11}$$

with $\nu(\omega,n)$ the combined density of states corresponding to the spectrum $\omega_{\mathbf{k}}$ of direct electron excitations.[18]

Equation (10) gives the inverse dependence of the electron concentration n on the intensity I of the light field in a stationary state of equilibrium. Analysis of (9) shows that the state is stable if $\partial I/\partial n > 0$ in (10). The equation for the critical points at which stability is lost has the form $\partial I/\partial n = 0$. Combining this with (10) yields

$$S(\omega_0, n) - n \frac{\partial S(\omega_0, n)}{\partial n} = 0. \tag{12}$$

Thus, if there are points (ω_0, n) that satisfy Eq. (12), there can be photoinduced phase transitions in the electron system.

At the bifurcation point (ω_b, n_b), where the direct and reverse phase transitions are second-order and occur at the same value of I, in addition to (12), the condition $\partial^2 I/\partial n^2 = 0$ must be met. This together with (10) yields

$$\frac{\partial^2 S(\omega_0, n)}{\partial n^2} = 0. \tag{13}$$

Equations (10)–(13) are the principal relationships describing the behavior of an electron system with a transformable spectrum in a light field.

3. RELATIONSHIP BETWEEN LIGHT INTENSITY AND ELECTRON CONCENTRATION IN THE CONDUCTION BAND

We now discuss a composite model for the density of electron states,

$$\nu(\omega, n) = \frac{\nu_0}{1 + e^x}, \tag{14}$$

where

$$x = x(\omega, n) = \alpha(\omega_1 - \beta n - \omega); \tag{15}$$

x is a dimensionless variable characterizing the frequency offset $\omega_1 - \beta n$ of the lower edge of the optical transition from the frequency ω of a spectral component of the light field, ω_1 is the frequency of the lower edge of the optical transition in the absence of electron excitations (at $n=0$), ν_0 is the density of states within the band, $\alpha>0$ is the reciprocal bandwidth of band smearing near the lower edge of the transition, and $\beta>0$ is the coefficient of the red shift of the lower edge of the transition due to an increase in the concentration n of electrons in the conduction band. Both expressions, (14) and (15), are selected with due regard for the following experimental facts: 1) for the majority of amorphous and some crystalline (e.g., CdS) substances, the frequency dependence of the optical absorption coefficient near the lower edge of the transition is described by the Urbach rule and is an exponential function;[19] 2) in many semiconductors, both amorphous[20] and crystalline,[11] e.g., CdS (Refs. 9 and 10), an increase in the concentration n of electrons in the conduction band brings about a red shift in the frequency of the lower edge of the transition that is close to linear over a wide range of n ($n\simeq 10^{17}$–10^{19} cm^{-3} for CdS; Ref. 9).

To calculate the integral (11) with the composite density of states ν given by (14) and (15), we assume that the incident spectrum, $G(\omega)=Ig(|\omega|-\omega_0)$ has a rectangular shape:[16]

$$g(y)=\begin{cases} 1/2\Delta\omega, & |y|\leqslant\Delta\omega, \\ 0, & |y|>\Delta\omega, \end{cases} \tag{16}$$

where $\Delta\omega$ is the halfwidth of the spectrum. Plugging (14)–(16) into (11), we get

$$S(\omega_0,n)=\nu_0-\frac{\nu_0}{2\alpha\Delta\omega}\ln\frac{1+\exp[x(\omega_0,n)+\alpha\Delta\omega]}{1+\exp[x(\omega_0,n)-\alpha\Delta\omega]}. \tag{17}$$

Equations (12) (13), and (17) imply that at the bifurcation point $(\omega_0,n)=(\omega_b,n_b)$, the parameters of the system and the light field obey the following conditions:

$$x=x(\omega_b,n_b)=0,$$

$$\alpha(\omega_1-\omega_b)=\alpha\Delta\omega\coth\frac{\alpha\Delta\omega}{2}\equiv x_{0b}. \tag{18}$$

In the second condition we introduced the dimensionless notation $x_{0b}=\alpha(\omega_1-\omega_b)$, which corresponds to the offset of the initial (at $n=0$) frequency ω_1 of the transition's lower edge from the light-field carrier frequency $\omega_0=\omega_b$ at the bifurcation point.

Near the bifurcation point at $|x|\leqslant 1$, Eq. (10), with (17) taken into account, can be expanded in a Taylor series in x, with only terms that are no higher than the cubic left in the series:

$$I=\frac{\hbar x_0 z(x)}{\pi d^2\tau\alpha\beta\nu_0}, \tag{19}$$

$$z(x)=1+a_1x+a_2x^2+a_3x^3,$$

$$x=\alpha(\omega_1-\beta n-\omega_0), \tag{20}$$

where $x_0=\alpha(\omega_1-\omega_0)$ is the dimensionless parameter characterizing the initial offset of the frequency ω_1 of the transition's lower edge from the light-field carrier frequency ω_0; a_1, a_2, a_3 and

$$a_1=\frac{1}{x_{0b}}-\frac{1}{x_0},\quad a_2=\frac{1}{x_{0b}}\left(\frac{1}{x_{0b}}-\frac{1}{x_0}\right),$$

$$a_3=\frac{1}{x_{0b}}\left[\frac{1}{x_{0b}}\left(\frac{1}{x_{0b}}-\frac{1}{x_0}\right)-\frac{1}{12\cosh^2(\alpha\Delta\omega/2)}\right], \tag{21}$$

are the expansion coefficients. Equations (19) and (20) represent the inverse dependence on the light-field intensity I of the concentration n of electrons in the conduction band.

4. EXISTENCE CRITERIA FOR A PHOTOINDUCED PHASE TRANSITION

The points x_1 and x_2 corresponding to loss of system stability can be obtained from the equation $\partial z/\partial x=0$. Combining this equation with (20) yields

$$x_{1,2}=-\frac{a_2\pm\sqrt{a_2^2-3a_1a_3}}{3a_3}. \tag{22}$$

We see that the system behaves in a critical way if $a_2^2-3a_1a_3>0$. If the expansion coefficients (21) are plugged into this inequality, we arrive at the necessary condition for a phase transition, a condition that imposes a restriction on the initial offset of the light-field carrier frequency from the frequency of the lower edge of the transition: $x_{0b}\leqslant x_0$ or, in the initial notation,

$$\Delta\omega\coth\frac{\alpha\Delta\omega}{2}\leqslant\omega_1-\omega_0. \tag{23}$$

This shows that as the halfwidth $\Delta\omega$ of the incident spectrum grows, the lower limit of allowed initial offsets $\omega_1-\omega_0$ increases.

The condition (23), which determines the existence of a phase transition, bounds the light-field frequency ω_0 from above. We now derive a relationship that bounds this frequency from below. The physical reason for such a restriction is the saturation of the optical transition if the light-field frequency is sufficiently low.

A transition is not saturated for any frequency of the incident light (at absolute zero) if

$$\hbar\int_{-\infty}^{\omega_0-\Delta\omega}\nu(\omega,n)d\omega>n. \tag{24}$$

Plugging (14) and (15) into (24), we find that

$$\omega_1-\omega_0<\beta n-\Delta\omega-\ln\left[\exp\frac{\alpha n}{\hbar\nu_0}-1\right]. \tag{25}$$

The inequality (25) must be true for all n up to the phase transition point. An analysis of (25) with allowance for (23) finally yields

$$\begin{aligned}\alpha\Delta\omega\coth\frac{\alpha\Delta\omega}{2}&\leqslant\alpha(\omega_1-\omega_0)\\ &<\beta\hbar\nu_0\ln\beta\hbar\nu_0-(\beta\hbar\nu_0-1)\\ &\quad\times\ln(\beta\hbar\nu_0-1)-\alpha\Delta\omega,\end{aligned} \tag{26}$$

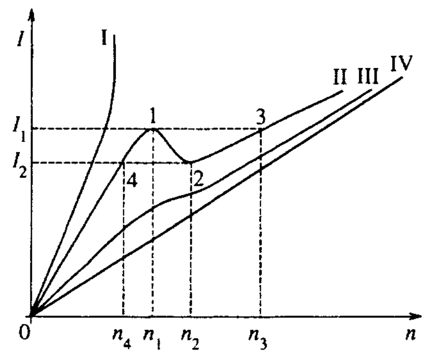


FIG. 1. Inverse dependence of the concentration n of the nonequilibrium electrons in the conduction band on the intensity I of the light field with a carrier frequency $\omega_0=\omega_j$ ($j=$I–IV, $\omega_{\mathrm{I}}<\omega_{\mathrm{II}}<\omega_{\mathrm{III}}<\omega_{\mathrm{IV}}$). The frequency $\omega_0=\omega_{\mathrm{II}}$ meets the criterion (26).

$$\alpha\Delta\omega\left(1+\coth\frac{\alpha\Delta\omega}{2}\right)<\beta\hbar\nu_0\ln\beta\hbar\nu_0-(\beta\hbar\nu_0-1)\ln(\beta\hbar\nu_0-1). \tag{27}$$

Clearly, saturation is absent for all values of n if conditions (26) and (27) are met.

The existence criterion (27) for a phase transition, which constrains the system parameters, is a corollary of the criterion (26) for the light−field frequency. The stronger the inequality (27), the greater the range of optical frequencies that can be used to observe a photoinduced phase transition and cavityless optical bistability.

5. PRINCIPAL CHARACTERISTICS OF PHOTOINDUCED PHASE TRANSITIONS AND CAVITYLESS OPTICAL BISTABILITY

When the criterion (27) is met, analysis of Eqs. (10), (17), (19) and (20) shows that schematically the diagrams representing the I vs. n dependence resemble those in Fig. 1. Curve I illustrates the saturation of band-to-band optical transitions when the light-field carrier frequency $\omega_0=\omega_1$ is below the lower bound on frequency values at which cavityless optical bistability can be observed, values determined by the condition (26). At frequencies $\omega_0=\omega_{\mathrm{III}},\omega_{\mathrm{IV}}$ higher than the upper bound on frequency values allowed by (26), there is no optical bistability either (curves III and IV). Curve IV corresponds to optical transitions into the band: $\omega_{\mathrm{IV}}-\Delta\omega-\omega_1\gg 1/\alpha$. At frequencies $\omega_0=\omega_{\mathrm{II}}$ satisfying the inequality (26), the I vs. n curve may have a negative slope (curve II). As the light intensity grows, a point on the curve moves along the curve 0→4→1→3, but as the intensity decreases the point moves along the curve 3→2→4→0. The discontinuous variation of n on the 1→3 and 2→4 sections corresponds to direct and reverse phase transitions. The section $I_2<I<I_1$ is the bistability region.

By analyzing Eq. (20) we find the points x_3 and x_4 corresponding to new positions of equilibrium for the direct and reverse phase transitions, respectively,

$$x_{3,4}=-\frac{a_2\mp 2\sqrt{a_2^2-3a_1a_3}}{3a_3}, \tag{28}$$

and the dimensionless value Δx of the jump in electron concentration in the conduction band, $\Delta n=\Delta x/\alpha\beta$:

$$\Delta x=-\frac{\sqrt{a_2^2-3a_1a_3}}{a_3}. \tag{29}$$

Here it turns out that the jump Δx is the same for direct and reverse phase transitions.

Using (20), we can also derive an expression for the width of the hysteresis loop (the size of the region of cavityless optical bistability in terms of the dimensionless light-field intensity):

$$\Delta z=z(x_1)-z(x_2)=\frac{8(a_2^2-3a_1a_3)^{3/2}}{27a_3^2}. \tag{30}$$

Note that in accordance with the approximate nature of the initial formulas (19) and (20), Eqs. (22) and (28)–(30) are valid if $|x|\leqslant 1$. This imposes additional constraints on the value of x_0 from above. In particular, as (22) and (21) clearly show, for $|x_1|\leqslant 1$ not to exceed unity the following inequality must hold:

$$x_0\leqslant\frac{x_{0b}}{1-3x_{0b}^2/12(x_{0b}^2+2x_{0b}+3)\cosh^2(\alpha\Delta\omega/2)}. \tag{31}$$

Clearly, (26) and (31) allow for a simultaneous solution near the bifurcation point: $x_0\geqslant x_{0b}$. A similar situation exists for the other points specified by (22) and (28), x_2, x_3, and x_4, and for Eqs. (29) and (30). Thus, the adopted approximation makes it possible to describe a photoinduced phase transition and cavityless optical bistability near a bifurcation point $x_0=x_{0b}$, $x=0$.

6. COMPARISON WITH EXPERIMENT

As is known, photoinduced phase transitions in CdS and amorphous GeS_2 occur in a characteristic time τ_0 approximately equal to 10^{-10} s (Ref. 8) and 1 s (Ref. 13), respectively. Thus, there is no way in which this phenomenon can be explained by the thermal model, in which the switching time $\tau_0\sim 10^{-5}$ s (Refs. 8, 13, and 21). Additional corroboration of this fact is the small temperature variation, $\Delta T=0-15$ K, recorded in experiments in irradiating CdS (Ref. 8). For GeS_2 the temperature variation ΔT was not registered in the experiment of Lyubin and Tikhomirov.[13] It has been established, however, that the intensity of the light required for observing cavityless optical bistability in GeS_2 is lower than that for observing cavityless optical bistability in CdS by a factor of 10^4 (Refs. 1, 8 and 13).

We now interpret the phenomena of photoinduced phase transitions and cavityless optical bistability in CdS and amorphous GeS_2 irradiated by monochromatic light using the electron theory developed in the present paper. We start with CdS. For this substance we put $d^2\approx 10^{-39}$ esu (Ref. 9), $\tau\approx 10^{-9}$ s (Ref. 9), $\beta\approx 10^{-5}$ cm^3 s^{-1} (Ref. 9), $\alpha\approx 10^{-12}$ s, (Ref. 10), and $\nu_0\approx 0.5$ eV^{-1} per atom $\sim 10^{23}$ eV^{-1} cm^{-3} (Ref. 10). This yields $\hbar\beta\nu_0\approx 10^2$, which implies that the criterion (27) is sure to be met (for a monochromatic light field the inequality (27) is transformed into (36)). Using (19) and (20), we can estimate I at the bifurcation point (where, according to (18), $x=0$ and $x_0=2$): $I\sim 10^4$ esu, which corresponds to a light field with an intensity $cI/2\pi \sim 10^{13}$ esu$\sim 10^6$ W/cm^{-2}. This result agrees with the experimental value of intensity being roughly 10^6 W/cm^2 (Ref. 1).

Now we turn to amorphous GeS_2. We calculate the optical absorption coefficient γ and the jump in the intensity of the output signal brought on by the phase transition in the case where the incident light field is monochromatic. We allow for the fact that the photon flux F in the stationary mode with light propagating along the y axis obeys the following equation:

$$\frac{dF}{dy} = -\left(\frac{\partial n}{\partial t}\right)_i = -\frac{\pi E_0^2 d^2}{2\hbar}\,\nu(\omega_0,n). \tag{32}$$

Here $(\partial n/\partial t)_i$ is the variation in electron concentration in the conduction band due to stimulated transitions in the monochromatic field $E(t)=E_0\cos(\omega_0 t+\varphi)$. The second equality in (32) is written in accordance with Eq. (9), where the first term on the right-hand side corresponds to stimulated transitions. Since the intensity $\hbar\omega_0 F = c\sqrt{\varepsilon}E_0^2/8\pi$ (Ref. 17), Eq. (32) yields

$$\frac{dF}{dy} = -\gamma F, \quad \gamma = \frac{4\pi^2 d^2 \omega_0}{c\sqrt{\varepsilon}}\,\nu(\omega_0,n), \tag{33}$$

where $\sqrt{\varepsilon}$ is the refractive index at frequency ω_0. Using Eqs. (33), (14), (15), (29), and (21), we arrive at the following expression for the output intensity ratio $(I_1/I_3)_{\text{out}}$ before and after the phase transition (see Fig. 1):

$$\left(\frac{I_1}{I_3}\right)_{\text{out}} = \exp\left[\frac{3\pi^2 d^2 \omega_0 \nu_0 L\sqrt{(x_0-2)(4-x_0)}}{c\sqrt{\varepsilon}(3-x_0)}\right], \tag{34}$$

where L is the sample thickness in the direction of light propagation.

In accordance with Eqs. (30) and (20), the formula for the relative width of the hysteresis loop, $\Delta I/I_b$, where $\Delta I = I_1 - I_2$, and $I_b = (I_1+I_2)/2$ the input intensity at the bifurcation point ($I_1 \approx I_2 \approx I_b$ near the bifurcation point), is

$$\frac{\Delta I}{I_b} \simeq \Delta z = \frac{2[(x_0-2)(4-x_0)]^{3/2}}{3x_0(3-x_0)}. \tag{35}$$

We now make estimates for amorphous GeS_2. As is known, in amorphous substances the width ΔE of the region within which the lower edge of the optical transition is smeared amounts to 0.1–0.3 eV (Ref. 19), and for GeS_2 this value is close to the upper boundary.[20] For this reason we can put $\Delta E = 2\hbar/\alpha \simeq 0.25$ eV. Bearing in mind that for monochromatic light $x_{0b} = \alpha(\omega_1-\omega_b) = 2$ (see Eq. (18)), we arrive at an estimate for the offset of the frequency ω_1 of the transition's lower edge from the light-field frequency ω_b at the bifurcation point, $\hbar(\omega_1-\omega_b) = 0.25$ eV. Thus, in accordance with criterion (23), a photoinduced phase transition should be observed at offsets greater than 0.25 eV. This was observed in the experiments of Lyubin and Tikhomirov,[13] where cavityless optical bistability was detected at 0.29 eV offset but was absent at 0.16 eV.

Using the offset $\hbar(\omega_1-\omega_0) = 0.29$ eV of Lyubin and Tikhomirov's experiment[13] and taking $\hbar(\omega_1-\omega_b) = 0.25$ eV and $\alpha(\omega_1-\omega_b) = 2$, we find that $x_0 = \alpha(\omega_1-\omega_0) \simeq 2.3$. Whence, with (35), we find that $\Delta I/I_b \approx 0.22$, which is close to the experimental value $(\Delta I/I_b)_{\text{expt}} \approx 0.2$ (Ref. 13). If we put $d^2 \approx 10^{39}$ esu (Ref. 9), $\omega_0 \approx 10^{15}$ s^{-1} (Ref. 13), $\nu_0 \approx 0.5$ eV^{-1} per atom^{-1} $\sim 10^{23}$ eV$^{-1}\cdot$cm^{-3} (Ref. 19), $L \approx 0.5$ cm (Ref. 13), and $\sqrt{\varepsilon} \approx 1$, from Eqs. (33) and (34) we find that $\gamma_1 \approx 10$ cm^{-1} and $\ln(I_1/I_3)_{\text{out}} \approx 2$. The resulting theoretical values also correspond to the experiment data of Lyubin and Tikhomirov:[13] $\gamma_1 \approx 10$ cm^{-1} and $\ln(I_1/I_3)_{\text{out}} \approx 2.1$.

The present author knows of no value of the coefficient β for amorphous GeS_2. However, Feltz noted[20] that the red shift of the lower-edge optical-transition frequency is a general property of many amorphous substances, and by irradiating As_2S_3, the lower edge of the transition was shifted by 0.2 eV. In CdS under maximum light intensity ($I \approx 10^7$ W/cm^2 and $n \approx 10^{19}$ cm^{-3}), the value of the same quantity amounts to 0.05 eV (Ref. 9). Hence for some amorphous substances the characteristic value of β is close to, or even greater than, the value $\beta \approx 10^{-5}$ cm^3/s for CdS. Note that the same value, $\hbar\beta \approx 5\times 10^{-21}$ eV/cm^3, was used by Emel'yanov and Uvarova[22] as the characteristic value of the coefficient of the red shift of the band gap for semiconducting substances. In this case criterion (27) is valid with a larger margin than it is for CdS.

We estimate the characteristic value of I_1 at the phase transition by using Eqs. (19) and (20). Note that in the experiment $\Delta I/I_b \approx 0.2$, so that $I_1 \approx I_b$. Then, assuming that $x_0 = 2$, $x = 0$, $d^2 \approx 10^{-39}$ esu (Ref. 9), $\Delta E = 2\hbar/\alpha \approx 0.25$ eV (Ref. 19 and 20), $\beta \approx 10^{-5}$ cm^3/s (Refs. 9 and 22), $\nu_0 \approx 0.5$ eV^{-1} per atom$\sim 10^{23}$ eV^{-1} cm^{-3} (Ref. 19), and $\tau \sim 10^{-3}$ s (Ref. 20), from Eqs. (19) and (20) we find that $cI/2\pi \sim 10^2$ W/cm^2, which corresponds to the experimental value 60 W/cm^2 of the intensity.[13]

7. CONCLUSION

We have built an electron-type theory of photoinduced phase transitions and cavityless optical bistability that yields a satisfactory description of experimental data on CdS and amorphous GeS_2. A necessary condition for observing optical bistability is the criterion (27), which bounds the half-width $\Delta\omega$ of the impinging-light spectrum from above and the material parameter $\beta\nu_0$ from below. In particular, at $\Delta\omega = 0$ the inequality (27) can be approximately written as

$$\hbar\beta\nu_0 > 3.2. \tag{36}$$

As criterion (27) shows, the lower bound on the allowed values of $\beta\nu_0$ rises as $\Delta\omega$ grows.

The second necessary condition for observing optical bistability is (26), which bounds the light-field carrier frequency ω_0 from above and below. As $\Delta\omega$ grows, the lower boundary of the allowed values of ω_0 rises and the upper boundary decreases. Thus, the broadening of the optical spectrum of the incident light reduces the possibility of observing photoinduced phase transitions and cavityless optical bistability in the system.

The inverse dependence of the electron concentration in the conduction band on the intensity of the incident light field is described by Eqs. (19)–(21), the characteristic points x_j of the phase transition, where $j=1$–4 (see Fig. 1) by Eqs. (22), (28) and (21), and the optical bifurcation parameters by Eqs. (29), (30) and (21). Analysis shows that as ω_0 approaches the bifurcation point, corresponding to the lower boundary of the values allowed by (26), the intensity I_1 and I_2 and the parameters Δx and Δz decrease. In the limit $\omega_0=\omega_b$, the photoinduced phase transition becomes a second-order transition.

1) E-mail: semenov@quant.univ.simbirsk.su

[1] H. Gibbs, *Optical Bistability: Controlling Light with Light*, Academic Press, New York (1985).

[2] V. I. Emel'yanov, N. L. Levshin, and A. L. Semenov, Vestnik Moskov. Univ. Ser. III Fiz. Astronom. **30**, 52 (1989).

[3] V. I. Emel'yanov and A. L. Semenov, Fiz. Tverd. Tela (Leningrad) **32**, 3083 (1990) [Sov. Phys. Solid State **32**, 1790 (1990)].

[4] V. I. Emel'yanov, N. L. Levshin, and A. L. Semenov, Fiz. Tverd. Tela (Leningrad) **31**, No. 10, 261 (1989) [Sov. Phys. Solid State **31**, 1803 (1989)].

[5] V. I. Emel'yanov, N. L. Levshin, and A. L. Semenov, Vestnik Moskov. Univ. Ser. III Fiz. Astronom. **31**, 99 (1990).

[6] A. L. Semenov, Fiz. Tverd. Tela (St. Petersburg) **36**, 1974 (1994) [Phys. Solid State **36**, 1079 (1994)].

[7] V. I. Emel'yanov, N. L. Levshin, S. Yu. Poroĭkov, and A. L. Semenov, Vestnik Moskov. Univ. Ser. III Fiz. Astronom. **32**, 63 (1991).

[8] P. I. Khadzhi, G. D. Shibarshina, and A. Kh. Rotaru, *Optical Bistability in a System of Excitons and Biexcitons in Semiconductors* [in Russian], Shtiintsa, Kishinev (1988), p. 120.

[9] V. G. Lysenko, V. I. Revenko, T. G. Tratas, and V. B. Timofeev, Zh. Éksp. Teor. Fiz. **68**, 335 (1975) [Sov. Phys. JETP **41**, 163 (1975)].

[10] V. V. Korshunov, M. V. Lebedev, and V. G. Lysenko, Fiz. Tverd. Tela (Leningrad) **27**, 1518 (1985) [Sov. Phys. Solid State **27**, 913 (1985)].

[11] A. A. Bugaev, B. P. Zakharchenya, and F. A. Chudnovskiĭ, *The Metal–Insulator Phase Transition and Its Application* [in Russian], Nauka, Leningrad (1979), p. 28.

[12] A. A. Bugaev, V. V. Gudyalis, B. P. Zakharchenya *et al.*, JETP Lett. **34**, 430 (1981).

[13] V. M. Lyubin and V. K. Tikhomirov, JETP Lett. **55**, 23 (1992).

[14] A. V. Andreev, V. I. Emel'yanov, and Yu. A. Il'inskiĭ, *Cooperative Phenomena in Optics: Superradiation, Bistability, and Phase Transitions* [in Russian], Nauka, Moscow (1988), p. 256.

[15] A. S. Davydov, *Solid-State Theory* [in Russian], Nauka, Moscow (1976), p. 296.

[16] S. A. Akhmanov, Yu. E. D'yakov, and A. S. Chirkin, *Introduction to Statistical Radiophysics and Optics* [in Russian], Nauka, Moscow (1981), p. 42.

[17] D. N. Klyshko, *Physical Bases of Quantum Electronics* [in Russian], Nauka, Moscow (1986), p. 22.

[18] V. L. Bonch-Bruevich and S. G. Kalashnikov, *Semiconductor Physics* [in Russian], Nauka, Moscow (1977), p. 608.

[19] N. F. Mott and E. A Davis, *Electronic Processes in Non-crystalline Materials*, Clarendon Press, Oxford (1979), pp. 274–5.

[20] A. Feltz, *Amorphous and Glassy Inorganic Solids* [in Russian], Mir Publishers, Moscow (1986), p. 518 [original German edition: *Amorphe und glasartige anorganische Festkörper*, Akademie-Verlag, Berlin (1983)].

[21] V. A. Stadnik, Fiz. Tverd. Tela (Leningrad) **29**, 3594 (1987) [Sov. Phys. Solid State **29**, 2059 (1987)].

[22] V. I. Emel'yanov and I. F. Uvarova, Zh. Éksp. Teor. Fiz. **94**, No. 8, 255 (1988) [Sov. Phys. JETP **67**, 1662 (1988)].

Translated by Eugene Yankovsky

Pinning by twin boundaries and peak effect in YBaCuO high-T_c superconductors

I. F. Voloshin, A. V. Kalinov and L. M. Fisher

State Research Center ''All-Russian Electrical Engineering Institute,'' 111250 Moscow, Russia

K. I. Kugel' and A. L. Rakhmanov

Scientific Center for Applied Problems in Electrodynamics, Russian Academy of Sciences, 127412 Moscow, Russia

(Submitted 8 October 1996)
Zh. Éksp. Teor. Fiz. **111**, 2158–2174 (June 1997)

Measurements of the imaginary part of the ac magnetic susceptibility of single crystals and melt-textured samples of $YBa_2Cu_3O_x$ (YBCO) at $T=77$ K in a magnetic field ranging between 1 and 20 kOe are reported. If the dc magnetic field $\mathbf{H}_{dc}$ is rotated in the **ab** plane of the sample, the magnetic susceptibility and critical current density j_c have peaks corresponding to the magnetic field aligned with twin boundaries. Peaks in the curve of j_c versus magnetic field are observed at angles corresponding to these peaks, where $j_c \propto \sqrt{H_{dc}}$ in a wide range of magnetic fields. The results have been interpreted in terms of the theory describing twin boundaries as a system of quasi-planar pinning sites. The pinning is strong if the elastic displacements of flux lines are of the order of the vortex lattice constant d_f. These displacements decrease with the magnetic field because of the decrease in d_f, and the contribution of the elastic energy to the Gibbs potential is reduced accordingly, which is the cause of the peak effect. © *1997 American Institute of Physics.* [S1063-7761(97)01706-X]

1. INTRODUCTION

Electromagnetic properties of high-temperature superconductors (HTSC) in a mixed state have some remarkable features.[1] Thus the critical current density j_c in many HTSC materials is strongly anisotropic, and this anisotropy is related not only to the direction of the current density **j**, but also to the alignment of applied magnetic field **H**. In other words, at a fixed orientation of **j**, the critical current j_c strongly depends on the magnetic field direction, and this anisotropy is sometimes not directly correlated with of the crystal lattice symmetry. Another interesting feature of HTSC is the nonmonotonic character of the function $j_c(H)$, which has a notable peak in the range of magnetic fields $H \ll H_{c2}$, where H_{c2} is the upper critical field.

The YBCO crystal structure is such that even high-quality single crystals contain a lot of twin boundaries parallel to the crystal axis [110], in other words, at an angle of 45° with respect to the **a** and **b** axes and parallel to the **c** axis. The effect of twins on properties of HTSC in a mixed state has been studied both theoretically and experimentally.[2–16] The effect of twin boundaries on the microscopic characteristics of HTSC, such as j_c, however, is not clear yet, partly because it is difficult to manufacture single crystals without twins. Usually a crystal is first grown, then twin boundaries are eliminated through special treatment.[17] It is not clear whether defects localized near twin boundaries are also eliminated. The question of whether twin boundaries are strong pinning sites preventing motion of magnetic flux lines in a superconductor or, on the contrary, they are weak links through which vortices penetrate into the sample has remained unanswered. Measurements by different techniques contradict one another. For instance, magneto-optical (Faraday effect) measurements,[4,5] torque measurements,[6] and measurements of magnetization[7] indicate that twin boundaries are efficient pinning sites if the magnetic field is aligned with the twinning plane, whereas similar measurements performed by other authors (measurements of the Faraday effect[8] and magnetization[9,10]) lead to a different conclusion. The data concerning the relation between the peak effect and twin boundaries in HTSC obtained by different researchers are also contradictory (compare the data obtained by Fisher *et al.*,[11] who claimed to have detected this relation, and by Zhukov *et al.*,[12] who did not observe it).

In our opinion, the contradictions between results by different authors arise for the following reasons. First, a twin can be either a pinning center or a channel for penetration of flux lines depending on the properties of the sample and magnetic field value. Second, in many experiments on the effect of twin boundaries on properties of HTSC, the critical current anisotropy was measured by varying the angle between the dc component of magnetic field, $\mathbf{H}_{dc}$, and the **c**-axis.[7,9,10,12] In this configuration of the experiment, the large contribution of the crystal anisotropy along the **c**-axis and in the **ab**-plane,[11] and the large demagnetization factor of a sample, whose dimension along the **c**-axis is usually much smaller than along the **a**- and **b**-axes, can mask the anisotropy due to twinning planes.

The present paper describes measurements demonstrating strong anisotropy of the critical current in the case when $\mathbf{H}_{dc}$ is rotated in the **ab**-plane and the angle between the **c**-axis and $\mathbf{H}_{dc}$ is maintained constant at 90°. In some samples, the critical current was measured as a function of the angle between the **c**-axis and $\mathbf{H}_{dc}$. We measured the low-frequency ac magnetic susceptibility of samples, and the critical current density was derived from the experimental data using the critical state model (the experimental techniques and calculations were described elsewhere[18]).

Theoretical studies dedicated to pinning on twins can be classified with two groups. In the first group, the magnitude of the elementary pinning force due to twin boundaries was studied.[13] In the second group, capture of tilted vortices by twin boundaries (so-called lock-in transition) was investigated.[14,15] These papers studied pinning of isolated vortices, and the role of interaction among vortices was discussed only cursorily. The paper by Larkin *et al.*,[16] where a twin is treated as an assembly of point defects and the critical current is calculated for the case of the magnetic field aligned with the twins, cannot be attached to either group.

In this paper, we use a theoretical approach somewhat different from those proposed previously. We take into account the shear and tilt deformations of the vortex lattice and pinning on twin boundaries at the same time. The theory predicts a peak on the curve of $j_c(H_{\mathrm{dc}})$ for pinning on twins at small angles between the twin boundaries and the magnetic field. Over a wide range of parameters we have $j_c \propto \sqrt{H_{\mathrm{dc}}}$, which is in good agreement with our experimental data.

Section 2 describes the parameters of the samples studied. The experimental techniques and calculations of critical current density are briefly described in Sec. 3, along with the experimental results. The theory is presented in Sec. 4, and the results are discussed in Sec. 5.

2. SAMPLES

Measurements were performed using both single crystals and highly melt-textured samples. In the first series of measurements, three $\mathrm{YBa_2Cu_3O}_x$ single crystals with dimensions of $2.4\times1.2\times0.06\ \mathrm{mm}^3$, $1.4\times0.8\times0.04\ \mathrm{mm}^3$, and $1.1\times0.6\times0.04\ \mathrm{mm}^3$ were used. In this paper, these samples are labeled as Y1, Y2, and Y6, respectively. The temperature T_{c0} of the onset of the superconducting transition derived from the real component of the ac magnetic susceptibility is 91.3 K for Y1 and 91.5 K for Y2 and Y6. Note that the T_{c0} values obtained by this technique are about 1 K lower than those measured by the four-terminal method. The transition width ΔT_c was defined as the difference between T_{c0} and the temperature of the peak in the imaginary component of the ac magnetic susceptibility, as usual in magnetic measurements. It was about 0.3 K for all the samples. The small transition width indicates that the sample quality was fairly high.

The tested samples had radically different twin patterns. Usually YBCO single crystals are composed of small regions (domains) in which the twin boundaries are parallel to each other.[4,19] The twin boundaries in neighboring domains are orthogonal (aligned with the [110] and [110] directions). The twin patterns of our samples could be clearly seen through a conventional optical microscope: for light at certain angles of incidence, parallel lines at an angle of 45° with respect to the sample faces are observed. Sample Y1 contained approximately equal numbers of domains with different orientations of twin boundaries, whereas samples Y2 and Y6 were almost single-domain, and the fraction of orthogonally oriented domains was less than 5%. The density of twin boundaries was determined using a transmission electron microscope. The typical separation between twins was about 0.5 μm.

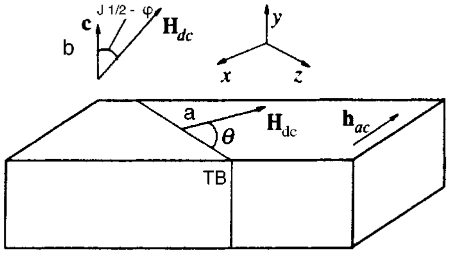


FIG. 1. Experimental configuration: (a) dc magnetic field is rotated in the **ab**-plane; (b) magnetic field is rotated from the **ab**-plane towards **c**-axis (TB denotes the twin boundary).

The melt-textured sample Sp1 was cut from a bulk $\mathrm{YBa_2Cu_3O}_x$ workpiece with a diameter of 8 mm and a length of 3 cm and had dimensions of $6\times2.3\times0.55\ \mathrm{mm}^3$. The **ab**-plane was parallel to the larger sample face. The sample was tested with a view to detecting weak links through ac magnetic susceptibility as a function of the amplitude of the alternating magnetic field at $H_{\mathrm{dc}}=0$. These measurements indicated that sample Sp1 did not contain many weak links.

3. EXPERIMENT

3.1. Experimental technique and measurements of susceptibility

The samples were tested using a contactless inductive technique by measuring the low-frequency magnetic susceptibility $\chi=\chi'+i\chi''$. The susceptibility χ can be determined from the general expression for the sample magnetization M induced by an ac magnetic field $h=h_0\cos\omega t$:

$$M(t)=h_0\sum_{n=1}^{\infty}\mathrm{Re}(\chi_n e^{in\omega t})=h_0\sum_{n=1}^{\infty}[\chi_n'\cos(n\omega t)+\chi_n''\sin(n\omega t)]. \tag{1}$$

In what follows χ' and χ'' denote the fundamental components χ_1' and χ_1'' of the ac magnetic susceptibility.

Technical details of the method are described elsewhere.[18] All measurements were performed at the liquid-nitrogen temperature $T=77$ K. The experimental configuration is shown in Fig. 1. A plate-shaped sample was placed in the middle of a solenoid generating the ac magnetic field $h=h_0\cos\omega t$ at a frequency $\omega/2\pi=130$ Hz. The magnetic field nonuniformity was within 0.5% in the sample volume. The field was within 1–2° of being parallel to the sample plane. The pickup coil was wound on the middle section of the sample as close as possible to its surface in order to minimize the magnetic flux between the sample and coil. The magnetic flux in this gap, which was not related directly to the magnetic flux in the sample, was compensated for by an auxiliary assembly of axial coils with a tunable mutual inductance. One of the coils was connected in series with the solenoid generating the ac magnetic field, and the other in series with the pickup coil. In addition to compensating for the spurious magnetic flux, this assembly allowed us to minimize the signal due to χ' without a change in the signal due

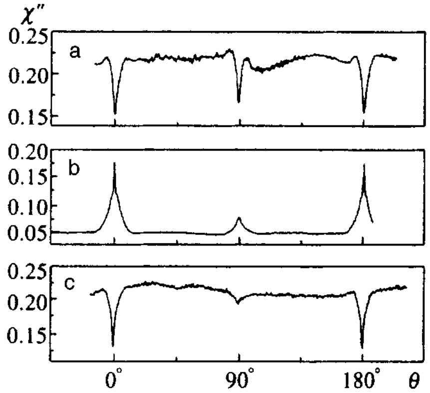


FIG. 2. Typical curves of $\chi''(\theta)$: (a) sample Y1, $H_{dc}=12$ kOe, $h_0=150$ Oe; (b) sample Y2, $H_{dc}=8$ kOe, $h_0=400$ Oe; (b) sample Y6, $H_{dc}=2$ kOe, $h_0=10$ Oe.

to the shift by $\pi/2$ in the χ'' phase (this signal was later used to calculate j_c). Thus, the measurement error related to the phase-locked detection was reduced considerably.

The sample, pickup coil, and solenoid generating the ac magnetic field were placed in the dc magnetic field H_{dc} generated by an electromagnet. This field was uniform to within 0.1% in the measurement zone. The electromagnet could be rotated independently around three orthogonal axes. The angle of rotation around the vertical axis was 360°. The rotation around the horizontal axes was used to align the magnetic field with the **ab** crystal plane to a high accuracy (about 1/6°).

The sample was aligned with the dc magnetic field as follows. First, it was placed in the electromagnet so that its **ab**-plane coincided approximately with the horizontal plane (rough alignment). Then the fine alignment was performed using the fact that the magnetic susceptibility and critical current j_c have narrow peaks close to $\mathbf{H}_{dc}\|\mathbf{ab}$[11,20] (see Fig. 4 below). The corresponding narrow peak in the signal from the pickup coil was used to rotate the magnet (and its field of about 10 kOe) accurately through small angles around the horizontal axes to attain the signal maximum. The alignment accuracy was determined by the fineness of the magnet rotation gear and was about 1/6°.

In order to measure the function $\chi''(\theta)$, where θ is the angle between $\mathbf{H}_{dc}$ and the twin boundary, we rotated the electromagnet so that the angle θ changed continuously between 0° and 360°. The field H_{dc} was parallel to the **ab** plane to within 1/6°. Typical curves of $\chi''(\theta)$ for various H_{dc} and h_0 in samples Y1, Y2, and Y6 are given in Fig. 2. The curves $\chi''(\theta)$ have pronounced peaks at $\theta=0$, 90°, and 180° in sample Y1, and at $\theta=0$ and 180° in samples Y2 and Y6, which corresponds to the directions of twin boundaries in these samples. Thus, our experiments indicate that the curves of $\chi''(\theta)$ are determined by the macrostructure of twin boundaries in single crystals.

Note that in the critical-state model the magnetic susceptibility χ'' is inversely proportional to j_c for $h_0<h_p$ and increases with j_c for $h_0>h_p$, where h_p is the ac field amplitude at which the magnetic field penetrates to the sample center (i.e., the Bean penetration field, which was less than 100–150 Oe in our samples). The measurements shown in Figs. 2a and 2c were taken at $h_0<h_p$, and those in Fig. 2b at $h_0>h_p$. The minima of $\chi''(\theta)$ in Figs. 2a and 2c and the maxima in Fig. 2b correspond to the maxima in the critical current density. The minima in χ'' are replaced by maxima as h_0 increases, which means a transition from the $h_0<h_p$ regime to $h_0>h_p$.

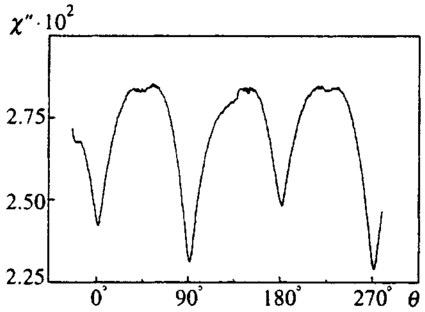


FIG. 3. Magnetic susceptibility χ'' versus angle θ between dc magnetic field and twin boundary for $H_{dc}=10$ kOe and $T=77$ K in the melt-textured sample Sp1.

Measurements of the magnetic susceptibility versus angle θ in the melt-textured sample Sp1 are given in Fig. 3 ($h_0<h_p$). At first sight, one could hardly expect that the twin boundaries in different crystallites would demonstrate a predominant orientation. Figure 3, however, shows that there is some correlation. The halfwidth of the peaks plotted against the rotation angle is about 10° (in single crystals it is about 1°), which indicates, apparently, that the misalignment among twin boundaries in different crystallites is relatively small throughout the sample.

In addition to measuring $\chi''(\theta)$ in the single crystal Y1, we have investigated the susceptibility as a function of the angle φ between the magnetic field $\mathbf{H}_{dc}$ and the **ab**-plane of the single crystal at $\theta\sim45°$ when $\varphi=0$. Figure 4 shows the susceptibility χ'' as a function of φ. Measurements were performed at $H_{dc}=6$ kOe and $H_{dc}=16$ kOe, and a temperature of 77 K. The curves have well defined peaks corresponding to the field aligned with the **ab**-plane ($\varphi=0$) and **c**-axis ($\varphi=90°$). One should bear in mind that the measurements

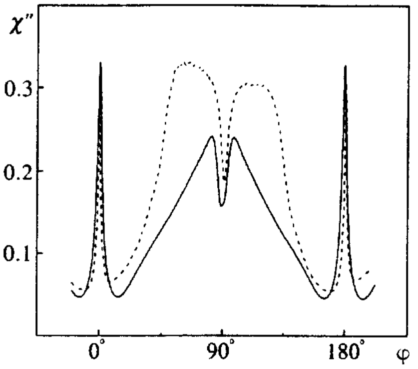


FIG. 4. Magnetic susceptibility χ'' as a function of angle φ between the dc magnetic field and **ab**-plane in the single crystal Y1 for $h_0=200$ Oe and two values of the dc magnetic field: $H_{dc}=16$ kOe (dashed line) and $H_{dc}=6$ kOe (solid line).

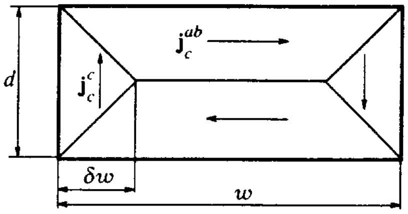


FIG. 5. Superconducting plate in a magnetic field.

of the effects of twin boundaries through variation of φ are more difficult to interpret than measurements in which θ varies, because the rotation of the magnetic field between the **c**-axis and the **ab**-plane also changes the angle between the dc magnetic field and the twin boundaries. Note that when φ was varied between 0 and 90°, the condition $h_0 > h_p$ was replaced with $h_0 < h_p$ owing to the notable increase in the critical current density near $\varphi = 90°$.

Our results were independent of the magnetic and thermal history of the samples if the ac field amplitude was larger than h_p. We have carried out sets of measurements with different magnetic, thermal, and angular histories of the samples and have not noticed any difference among the results for $h_0 > h_p$, which is not surprising if this fact is interpreted within the critical-state model.

3.2. Calculation of critical current density

Measurements of the imaginary part of the ac magnetic susceptibility can be used in deriving the critical current density based on using the critical-state model in the case when $\mathbf{H}_{\mathrm{dc}} \| \mathbf{ab}$.[18] Calculations for a sample shaped as a parallelepiped in a parallel magnetic field in a fairly general form taking into account the contributions of the current components j_c^{ab} and j_c^c to the susceptibility are reported elsewhere.[21] This technique can yield elaborate formulas for $\chi''(j_c^{ab}, j_c^c)$. Here we use, however, simpler formulas for the studied samples.

The relative contribution of j_c^{ab} and j_c^c to the sample magnetization M is determined by the equation[21]

$$\psi = \frac{j_c^{ab} d}{j_c^c w}, \tag{2}$$

where d is the sample thickness and w is its length (Fig. 5). The ratio j_c^{ab}/j_c^c for YBCO is usually less than 10 at $T = 77$ K.[22] On the other hand, our samples have $w/d = 30$–40, hence $\psi \ll 1$. It follows from the general formulas and qualitative considerations that the main contribution to M is due to j_c^{ab}. Specifically, if the ac magnetic field completely penetrates into the sample ($h_0 > h_p$), it follows from the current conservation (Fig. 5) that for $\psi \ll 1$ the total current $j_c^{ab} da$ in the **ab**-plane (here a is the sample width) equals the total current $j_c^c \delta w a$ along the **c**-axis ($\delta w a$ is the cross section through which the current flows along the **c**-axis). Then we have $\delta w / w \sim \psi \ll 1$, and the critical current can be calculated to within $\sim \psi$ using the expression for a thin plate neglecting edge effects. For $h_0 > h_p(H_{\mathrm{dc}})$ we have[18]

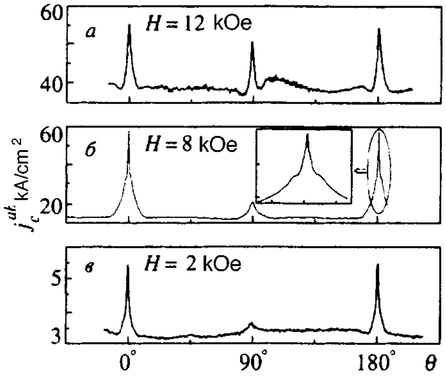


FIG. 6. Curves of $j_c^{ab}(\theta)$ for samples (a) Y1, (b) Y2, and (c) Y6. The insert shows the peak at $H_{\mathrm{dc}} = 8$ kOe for sample Y2 on the enhanced scale.

$$j_c^{ab}(H_{\mathrm{dc}}) = \frac{3ch_0}{8\pi d}\left[1 - \sqrt{1 - \frac{4\pi\chi''(H_{\mathrm{dc}})}{3}}\right], \tag{3}$$

where $h_p = 2\pi d j_c^{ab}/c$.

The range of magnetic fields to which Eq. (3) applies is determined by the inequalities

$$H_{c1}, h_p < h_0 \ll H_{\mathrm{dc}}, \tag{4}$$

where H_{c1} is the first critical field, which is no higher than 50–80 Oe in YBCO at liquid nitrogen temperature. Since $h_p \leq 100$–150 Oe in the studied samples, the magnetic fields used in our experiments are $H_{\mathrm{dc}} > 1$ kOe and $h_0 = 100$–200 Oe.

The curves of $j_c^{ab}(\theta)$ for samples Y1, Y2, and Y6 are shown in Fig. 6. One can see that these curves, like those of $\chi''(\theta)$, have clearly pronounced narrow peaks at angles corresponding to the twin boundaries (at $\theta = 0$, 90, and 180° for sample Y1 and $\theta = 0$ and 180° for samples Y2 and Y6). The strong correlation between the peak positions and the orientations of the twin boundaries leads to the conclusion that the twins are strong pinning sites when the field $\mathbf{H}_{\mathrm{dc}}$ is parallel to the twin boundaries. Figures 2 and 6 show that $\chi''(\theta)$ and $j_c^{ab}(\theta)$ vary only slightly between the peaks. This means, in particular, that in our experiments the magnetic field $\mathbf{H}_{\mathrm{dc}}$ remained in the **ab**-plane.

Figure 7 shows curves of $j_c^{ab}(H_{\mathrm{dc}})$ at various θ for the sample Y2. It is clear that, in accordance with experimental data reported by other authors, the critical current density decreases monotonically with the magnetic field $H_{\mathrm{dc}} \| \mathbf{ab}$ at almost all angles θ. The new feature is that a strong peak effect occurs in the narrow range of angles $\Delta\theta \leq 5°$ near the peak of $j_c^{ab}(\theta)$ when the magnetic field is aligned with the twin boundaries. The range of angles in which the peak effect is observed almost coincides with the width of peaks in the curves of $j_c^{ab}(\theta)$. The peak amplitude in the curves of critical current versus magnetic field is the higher, the smaller the deviation of $\mathbf{H}_{\mathrm{dc}}$ from the twin boundary. At $\theta = 0$, we could not attain maxima in the curves of $j_c^{ab}(H_{\mathrm{dc}})$ for magnetic fields ranging up to 20 kOe. These results demonstrate that there is a direct relation between the peak effect in YBCO single crystals and the pinning by the twin boundaries.

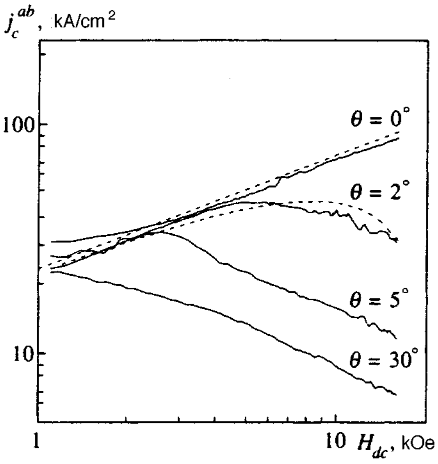


FIG. 7. Curves of $j_c^{ab}(H_{\mathrm{dc}})$ for sample Y2 at various θ.

Both the insert in Fig. 6 and Fig. 8 show the shapes of the peak in the curves of $j_c^{ab}(\theta)$ for different dc magnetic fields (sample Y2). One can see in Fig. 8 that the peak becomes narrower and higher as the magnetic field H_{dc} rises. Note that the peak has a specific shape: it has a narrow central section with a width of about 1° and a wider and lower pedestal (about an order of magnitude as low).

The function $j_c(H_{\mathrm{dc}},\varphi)$ at fixed $\theta = 45°$ was calculated for sample Y1. It turned out that the peak effect occurred when $\mathbf{H}_{\mathrm{dc}}$ was oriented around the **c**-axis in the angular range $\Delta\varphi \leqslant 40°$. This also indicates the correlation between the peak effect and pinning by the twin boundaries.

4. THEORY

In order to determine the critical current density, we need the Gibbs free energy G of a vortex lattice interacting with a system of quasi-planar defects imitating the twin boundaries. This energy contains three main components, namely, the energy G_p directly related to the vortex pinning by the defects, energy G_e of elastic deformation of the vortex lattice, and magnetic energy G_m. We assume that the twins are in planes parallel to **yz** (Fig. 1). The system of twins is not exactly periodic, but the deviation of the separation between the twins from its mean value is not large.[19] For simplicity, we ignore the anisotropy of YBCO crystals, because it should not radically change the results for the discussed configuration (when the magnetic field is in the **ab** plane), but would only change some numerical constants.

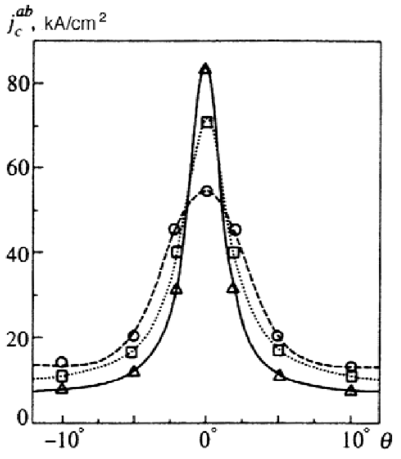


FIG. 8. Curves of $j_c(\theta)$ near the peak at various dc magnetic fields: circles) $H_{\mathrm{dc}} = 6$ kOe; squares) $H_{\mathrm{dc}} = 10$ kOe; triangles) $H_{\mathrm{dc}} = 16$ kOe (the curves are guides for the eye).

4.1. Twin model

Following Larkin *et al.*,[16] we assume that various defects can be localized near twin boundaries. Consequently, superconducting properties of the sample can vary notably near the twin boundaries. We assume also that the thickness of the defect region satisfies $l_{tx} \ll L_t$. A vortex outside the defect region near a twin interacts with defects in the sample volume, and the pinning potential per unit length of the vortex can be conveniently expressed as

$$U_p = -\frac{H_c^2 \xi^2}{8\pi} f_p, \tag{5}$$

where H_c is the thermodynamic critical field, ξ is the coherence length, and $f_p \ll 1$ is the dimensionless pinning potential. If a vortex is shifted to the region near the twin boundary, it is natural to assume that the magnitude of the pinning potential should increase, and f_p should be replaced with $f_p + \delta f_{px}$ in Eq. (5). In the general case, the pinning potential also varies in the direction parallel to the twinning plane. Let us denote the typical spread of f_p along the y-axis by δf_{py} and the typical spatial scale of this nonuniformity by l_{ty}. Then the pinning force per unit length of the vortex along the twinning plane is

$$F_{py} = \frac{H_c^2 \xi^2}{8\pi} \frac{\delta f_{py}}{l_{ty}}. \tag{6}$$

In what follows, we consider the range of fairly strong magnetic fields, when the magnetic induction B inside the sample is much larger than the characteristic field $B_0 = \Phi_0 / L_t^2$, where Φ_0 is the magnetic flux quantum. This means that the vortex lattice constant satisfies $d_f \sim \sqrt{\Phi_0/B_0} \ll L_t$ (in our samples $B_0 \approx 100$ G). Moreover, we assume that the scale of the pinning potential nonuniformity is small and

$$l_{tx,y} \ll d_f \ll L_t. \tag{7}$$

The first inequality in (7) allows us to consider the defects in question as quasi-planar, and the second allows us to apply equations of the theory of elasticity to the vortex lattice with boundary conditions specified at the twin boundaries.

4.2. Critical current (the case of magnetic field parallel to twin boundaries)

In order to clarify the basic principles of our theoretical model, let us analyze in detail the simplest case of pinning by a system of twins, when the magnetic field $\mathbf{H}_{\mathrm{dc}}$ is parallel to the twin boundaries and the **ab**-plane, and the pinning in the sample volume is weak and can be neglected. In the absence of transport current, the equations of the theory of elasticity for the vortex lattice in regions between twins have the form[1]

$$(C_{11}-C_{66})\,\frac{\partial}{\partial x_i}\,(\nabla\cdot\mathbf{u})+C_{66}\Delta u_i+C_{44}\,\frac{\partial^2 u_i}{\partial z^2}=0, \tag{8}$$

where $i=x,\,y$; C_{11}, C_{44}, and C_{66} are the bulk, tilt, and shear elastic moduli; ∇ and Δ are two-dimensional operators acting in the **xy**-plane. It will be shown below that only long-range deformations with a typical wave vector q such that $(qL_t)^2\ll 1$ are essential. For the studied samples, the separation L_t between twin boundaries is larger than the magnetic field penetration depth λ, so $(q\lambda)^2\ll 1$ holds, and we will neglect a spatial dispersion of C_{11} and C_{44}. Then in the range $H_{c1}\ll B\ll H_{c2}$ of magnetic fields studied[1]

$$C_{66}=\frac{B\Phi_0}{(8\pi\lambda)^2}\ll C_{11}=C_{44}=\frac{B^2}{4\pi}, \tag{9}$$

and the elastic energy per unit volume can be expressed as

$$G_e=\frac{1}{2V}\int \mathrm{d}^3r\left\{(C_{11}-C_{66})(\nabla\cdot\mathbf{u})^2+C_{66}(\mathbf{e}\cdot\nabla\mathbf{u})^2 + C_{44}\left(\frac{\partial\mathbf{u}}{\partial z}\right)^2\right\}, \tag{10}$$

where the vector $\mathbf{e}=\{1,1\}$ is in the **xy**-plane.

In the unperturbed state, when vortices do not interact with twins, $G_e=0$. When the pinning comes into play, a fraction of vortices are pinned on twin boundaries, thus reducing the free energy per unit volume by

$$G_p=-\frac{H_c^2\xi^2}{8\pi}\,\delta f_{px}n_p, \tag{11}$$

where n_p is the density of pinned vortices. But the pinning also leads to deformations, hence $G_e>0$. At a given n_p, the lattice is configured so as to minimize its elastic energy. Various regimes of lattice pinning by twins are possible. In the simplest one-dimensional case, only displacements in the x-direction which are functions of x, $\mathbf{u}=\{u_x(x),0\}$, are generated in the lattice. But in this case unfavorable ($C_{11}\gg C_{66}$) compressing and tensile strains that increase the free energy also occur. The more favorable configuration of the lattice pinned by twins is a two-dimensional one, such that

$$\nabla\cdot\mathbf{u}=0. \tag{12}$$

Consider the part of the vortex lattice between two twin boundaries on which vortices are pinned. Let the separation L between the two twins be $L\gg L_t$. We represent a lattice displacement vector in the form of a Fourier series with respect to y:

$$u_j=\sum_q u_{qj}(x)e^{iqy},\quad j=x,\,y. \tag{13}$$

After substituting the expansion (13) into Eq. (8) and taking into account Eq. (12), we have

$$u_{qx}=C_{1q}e^{qx}+C_{2q}e^{-qx},$$

$$u_{qy}=-i(C_{1q}e^{qx}-C_{2q}e^{-qx}). \tag{14}$$

It is convenient to express the coefficients C_{jq} in terms of the component u_x of the lattice displacements at the neighboring twin boundaries $u_x(0)=u_x^{\alpha}$ and $u_x(L)=u_x^{\alpha+1}$:

$$u_{qx}=\frac{1}{\sinh(qL)}\,\{u_{qx}^{\alpha+1}\sinh(qx)+u_{qx}^{\alpha}\sinh[q(L-x)]\},$$

$$u_{qy}=\frac{i}{\sinh(qL)}\,\{u_{qx}^{\alpha+1}\cosh(qx)-u_{qx}^{\alpha}\cosh[q(L-x)]\}. \tag{15}$$

By substituting Eq. (15) into Eq. (10), we obtain the elastic energy per unit volume

$$G_e=\frac{2C_{66}}{L}\sum_q q\langle|u_{qx}^{\alpha}|^2\rangle\coth(qL), \tag{16}$$

where $\langle\ldots\rangle$ means averaging over twin boundaries, and all average values satisfy $\langle u_{qx}^{\alpha}u_{qx}^{\alpha+1}\rangle=0$. The latter condition is satisfied if the twins are distributed randomly, and deviations from the mean positions are approximately equal to or greater than d_f. Within our model, the parameter G_e can be calculated exactly, but, in order to simplify calculations, we will obtain our results with arbitrary constant factors.

In order to have a notable pinning energy, we need displacements comparable to the lattice constant d_f with a characteristic wave vector $q_0\sim 1/\Delta L$, where ΔL is the length of a vortex row pinned by a twin boundary. On the other hand, it follows from Eq. (16) that

$$G_e\approx\gamma\,\frac{C_{66}d_f^2}{L^2}\begin{cases}1, & q_0L\ll 1,\\ q_0L, & q_0L\gg 1,\end{cases} \tag{17}$$

where $\gamma=\mathrm{const}<1$. For example, if the displacements are supposed to be uniformly distributed over the interval $\pm d_f/2$, one can easily obtain $\gamma=1/6$. Therefore, in the case of approximately equal displacement amplitudes (hence, approximately equal n_p), the largest gain in the free energy is provided by long-wavelength deformations with $q_0L\ll 1$. This statement, however, is fairly obvious. Thus, the optimum structure of magnetic flux trapped by twin boundaries involves positions of pinned vortex rows with the typical length $\Delta L\sim 1/q_0\gg L$, and in an ideal case (zero temperature, absolutely parallel twin boundaries, no history effects, etc.) should be comparable to the sample dimension along the y-axis. Then we have an obvious estimate of the density of pinned vortices for our model:

$$n_p\approx n_0d_f/L\approx 1/Ld_f, \tag{18}$$

where $n_0=B/\Phi_0\approx 1/d_f^2$ is the average vortex density in the sample.

The characteristic separation L between the pinned rows is determined by the condition of the minimum Gibbs energy. Using Eqs. (9), (11), and (17) in the limit $q_0L\ll 1$ and the Ginzburg–Landau relationships, we obtain

$$G\approx-\frac{1}{4\gamma C_{66}}\left(\frac{H_c^2\xi^2}{8\pi}\,n_0\delta f_{px}\right)^2,$$

$$L=\frac{16\pi\gamma C_{66}d_f}{H_c^2\xi^2n_0\delta f_{px}}>L_t,$$

$$B<B_t=\frac{(2\pi\gamma)^2\Phi_0}{L_t^2\delta f_{px}^2}, \tag{19}$$

$$G\approx-\frac{H_c^2\xi^2 d_f}{8\pi L_t}\,n_0\delta f_{px}+\gamma C_{66}\frac{d_f^2}{L_t^2}<0,$$

$$L=L_t,\quad B>B_t. \tag{20}$$

Since we have $2\pi\gamma\sim 1$ and $\delta f_{px}\ll 1$, the magnetic field B_t in our samples should be at least several tesla. Since $B\approx H_{\rm dc}$ holds in the magnetic field range studied, $G\propto -H_{\rm dc}$, $L\propto 1/\sqrt{H_{\rm dc}}$ for $H_{\rm dc}<B_t$. Thus, the parameter G decreases with $H_{\rm dc}$ because the distance L between chains of pinned vortices decreases.

It was mentioned in Sec. 3 that we measured the current flowing in the **ab**-plane (or, using the notation of this section, in the **xz**-plane). Then the Lorentz force F_L acting on vortices has only a y-component and equals $F_L=j_xB/c$, where j_x is the x-component of the transport current density. Our task is to calculate the critical value $j_x=j_{cx}$ at which the vortex lattice is set in motion. Let us assume for simplicity that the critical current density component j_{cb} due to the volume pinning is simply added to the critical current density j_{ct} due to pinning by twin boundaries: $j_{cx}=j_{ct}+j_{cb}$. This approximation can be used in two limiting cases: $j_{cb}\gg j_{ct}$ and $j_{cb}\ll j_{ct}$.

Vortices can be depinned by one of three processes. First, vortices pinned by twins can move along them and entrain the deformed vortex lattice if

$$\frac{L}{d_f}\,\frac{(j_x-j_{cb})\Phi_0}{c}\geqslant\frac{H_c^2\xi^2\delta f_{py}}{8\pi l_{ty}},$$

where the ratio L/d_f is approximately equal to the number of vortices in the bulk per vortex pinned by twin boundaries. Then the critical current is given by

$$\frac{j_{ct}}{j_0}\sim\frac{\xi^2}{L\sqrt{h}}\,\frac{\delta f_{py}}{l_{ty}},\quad h=\frac{H_{\rm dc}}{H_{c2}}, \tag{21}$$

where $j_0\sim c\Phi_0/\lambda^2\xi$ is the depairing current density in the Ginzburg–Landau theory.

Second, the vortices can be depinned from twin boundaries if the additional elastic energy G_L due to vortex lattice deformation by the transport current is larger than the gain in free energy due to the pinning. In calculating G_L, we use Eqs. (8)–(12), adding to the right-hand side of Eq. (8) the term corresponding to F_L. Assuming that the transport current is lower than the critical value, the vortex lattice is pinned, and, as stated previously, $\langle u_x^\alpha\rangle=0$, we can calculate the total elastic energy of the pinned vortex lattice. One can easily check that it can be expressed as a sum of two components: G_e+G_L. The first of these is determined by Eqs. (16) and (17) and is due to the lattice pinning by twins. The second is due to the Lorentz force and equals

$$G_L=\frac{1}{6C_{66}}\left[\frac{H_{\rm dc}(j-j_{cb})L}{c}\right]^2. \tag{22}$$

The critical current for $H_{\rm dc}<B_t$ is determined by the equations

$$\frac{\partial}{\partial L}\,(G_p+G_e+G_L)=0,$$

$$G_p+G_e+G_L=0. \tag{23}$$

Thus we obtain the relations omitting the constant factors:

$$\frac{j_{ct}}{j_0}\sim\frac{\xi}{L}\,\delta f_{px}\sim\delta f_{px}^2\sqrt{h},\quad H_{\rm dc}<B_t. \tag{24}$$

The critical current increases with the magnetic field because the characteristic displacement ($\sim d_f$) of the vortex lattice required for the effective pinning of a vortex by a twin boundary drops with $H_{\rm dc}$. Accordingly, the density of pinned vortices n_p increases if the contribution G_e to the Gibbs energy is constant. When $H_{\rm dc}>B_t$ holds, the parameter n_p saturates, and the slope of the curve of $j_c(H_{\rm dc})$ changes its sign. Note, however, that this effect occurs in the range of high magnetic fields, which were not attained in our experiments, and so it will not be discussed below.

Finally, the third mechanism setting vortices in motion is plastic deformation of the vortex lattice near the twin boundaries, while the vortices pinned by the twin boundaries are immobile. A calculation similar to that in Ref. 23 [with L derived from Eq. (19)] yields

$$\frac{j_{ct}}{j_0}\sim\beta_L\,\frac{\xi}{L}\sim\beta_L\delta f_{px}\sqrt{h}, \tag{25}$$

where $\beta_L\sim 0.1-0.2$ is a constant similar to the Lindemann constant in the theory of melting. As in the previous case, Eq. (25) predicts a peak effect due to the pinning by the twin boundaries.

Naturally, the actual mechanism of transition to a resistive state corresponds to the lowest critical current. Thus, our theory predicts an increase in the component j_{ct} of the critical current density with the magnetic field in the range between B_0 and B_t, if the quasi-planar defect is highly nonuniform in the longitudinal direction so that

$$\frac{\delta f_{py}}{l_{ty}}>\min\{\delta f_{px}\beta_L\}\,\frac{\sqrt{h}}{\xi}. \tag{26}$$

In particular, the peak effect in the transverse current is absent if a defect related to a twin boundary is an ideal plane. To conclude this section, note that δf_{px} in the samples studied is of the order of several hundredths (see Sec. 5), hence $\delta f_{px}<\beta_L$. Then it follows from Eq. (26) that in our model the peak effect should occur only in the field range $h<h^*=(\xi\delta f_{py}/L_{ty}\delta f_{px})^2$. On the other hand, there is a range of growth for $j_c(H_{\rm dc})$ only if $B_t/H_{c2}<h^*$ or

$$\delta f_{py}>l_{ty}/L_t. \tag{27}$$

The latter inequality is the condition for existence of the peak effect in our model. Below we will assume that the twin is sufficiently nonuniform and the condition (27) is valid.

4.3. Angular dependence of critical current

Now let the field $\mathbf{H}_{\rm dc}$ be rotated through an angle θ in the **xz**-plane. In this case, the magnetic field induction $\mathbf{B}$ in the sample makes an angle $\beta(r)$ with the twin boundary gradually changing between 0 and θ as a function of position. Then an additional term should be added to the Gibbs free energy:

$$G_m = -\frac{1}{4\pi V}\int \mathrm{d}V\mathbf{B}\cdot\mathbf{H}, \tag{28}$$

and the derivatives $\partial\mathbf{u}/\partial z$ should be included in the equations for elastic vortex displacements. Here we have defined the undistorted (ground) state as a configuration of the vortex lattice with all vortices aligned with twin boundaries. In what follows, we limit our discussion to the range of small angles $\theta \ll 1$. Then we can use the elasticity equations in the form of Eq. (8), assuming that the z-axis is aligned with the twin plane. Moreover, for $B \gg H_{c1}$ we can take $B \approx H$. Then we find from Eq. (28)

$$G_m = \mathrm{const} + \frac{H_{\mathrm{dc}}^2}{8\pi}\langle(\theta-\beta)^2\rangle. \tag{29}$$

After substituting into Eq. (8) the Fourier series for the displacement,

$$u_j = \sum_q u_{qj}(x)e^{iqy+ikz}, \quad j = x,\, y,$$

we obtain an expression for the elastic energy using Eqs. (10) and (12) by analogy with Eq. (16):

$$G_e = \frac{C_{66}}{L} \times \sum_{q,k}\frac{\langle|u_{qkx}^{\alpha}|^2\rangle}{\sinh^2(pL)} \times \frac{(p+q)[(p^2+q^2)\sinh(2pL)-2pL(p-q)^2]}{4pq} + \frac{C_{44}}{2}\langle\beta^2\rangle, \tag{30}$$

where

$$p = \sqrt{q^2 + C_{44}k^2/C_{66}},$$

and we have taken into account that

$$\beta \approx \sqrt{\left(\frac{\partial u_x}{\partial z}\right)^2 + \left(\frac{\partial u_y}{\partial z}\right)^2}. \tag{31}$$

Given Eqs. (30) and (31), one can easily check that for the same displacement amplitude (assuming $\langle|u_i^{\alpha}|^2\rangle \sim d_f^2$, so that effective pinning occurs at twins), the maximum gain in the free energy is provided by thermodynamically favored long-wavelength deformations, such that $pL \ll 1$. This conclusion, however, is self-evident. Thus, the typical dimension L of the fraction of the vortex lattice captured by a twin along the z and y directions is much larger than L_t. Then the density of the pinning energy is still determined by Eqs. (11) and (18), and the elastic energy can be estimated, by analogy with Eq. (17), by the equation

$$G_e = \gamma\,\frac{C_{66}d_f^2}{L^2} + \frac{C_{44}}{2}\langle\beta^2\rangle. \tag{32}$$

Since the angle β has a constant sign and changes slowly, we have $\sqrt{\langle\beta^2\rangle} \sim \langle\beta\rangle$. By minimizing the sum $G = G_p + G_e + G_m$ with respect to L and β, we obtain an estimate $\langle\beta\rangle \sim \theta/2$ (omitting a numerical factor) and

$$G \approx -\frac{H_{\mathrm{dc}}\Phi_0\delta f_{px}^2}{4\gamma\lambda^2(8\pi^2)^2} + \frac{H_{\mathrm{dc}}^2\theta^2}{32\pi}, \quad B_0 < B < B_t, \tag{33}$$

where we have taken into account Eq. (9) and the approximate equality $B \approx H_{\mathrm{dc}}$. It follows from Eq. (33) that such structure of pinned vortices exists if

$$\theta < \theta_m \sim \frac{\delta f_{px}}{\kappa\sqrt{h}}, \tag{34}$$

where $\kappa = \lambda/\xi$ is the Ginzburg–Landau parameter.

Assuming that the condition (26) and $\delta f_{px} < \beta_L$ are satisfied, we determine the angular dependence of the critical current. To this end, we use equations like (23) with the additional Gibbs term G_m. In the limit $\theta \ll 1$, the energy of elastic deformation due to the Lorentz force is still expressed by Eq. (22), with the exception of minor corrections. Then we derive from Eqs. (32) and (34)

$$\frac{j_{ct}}{j_0} \sim \delta f_{px}^2 h^{1/2}\left[\sqrt{1-\left(\frac{\theta}{\theta_m}\right)^2} - \frac{1}{3}\right]^{1/2}\left[1+\sqrt{1-\left(\frac{\theta}{\theta_m}\right)^2}\right]^{3/2}, \quad \theta < \theta_m. \tag{35}$$

The width of the peak in the curve of $j_{ct}(\theta)$ is of the order of θ_m, and the peak effect should be observed at these parameters.

5. DISCUSSION

Theoretical and experimental curves of j_c versus H_{dc} are shown in Fig. 7. These plots demonstrate that the theory gives an adequate qualitative description of experimental curves. Moreover, the increase in $j_{ct} \propto \sqrt{h}$ around the peak at $\theta = 0$ is in good agreement with experimental curves. Assuming that $\lambda = 2000-3000$ Å and $\kappa = 60$, we obtain an estimate $j_0 \sim 10^9$ A/cm^2, $H_{c2} \sim 30$ T and check that δf_{px} should be several hundredths in order to obtain an approximately correct value of j_{ct}. In this case, the peak width θ_m is several tenths of one degree, in accordance with Eq. (35), and the peak should be an order of magnitude narrower than the experimental value.

There may be several reasons for broadening of the peak in the angular dependence of the critical current. First of all, these may be purely technical reasons. For instance, the direction of the total magnetic field in experiment is variable because the vectors $\mathbf{H}_{\mathrm{dc}}$ and $\mathbf{h}_{\mathrm{ac}}$ are not strictly parallel. Moreover, twin boundaries can also be slightly nonparallel. The peak broadening, however, can be caused by physical properties of the sample. First, the defect regions near the twin boundaries have a finite thickness, which has not been taken into account in the calculation of the elastic energy of the vortex lattice. Second, all our estimates of the current and peak width refer to the model of an isotropic sample with some effective parameters [see, for example, Eq. (9)]. In reality, YBCO crystals are essentially anisotropic. This anisotropy can lead to a softening of the tilt modulus C_{44} if the vortex direction is not exactly orthogonal to the **c**-axis, and the vortex direction always has a **c**-component due to the y-component of the vortex lattice displacement. In the present work we do not try to discover which of the mechanisms broadens the peak in the angular dependence of the critical current, because this problem is too complicated and deserves special investigation. In this study, we have fitted θ_m to experimental data.

As was noted in the Introduction, in some experiments twins behaved like weak links, i.e., channels for penetration of vortices into samples. The analysis of literature data[2–16] reveals a tendency for twin boundaries to act as penetration channels mostly in the range of high magnetic fields, probably because the superconducting order parameter is suppressed by the magnetic field in regions with a high concentration of defects.

In conclusion, we have measured the ac magnetic susceptibility of single crystals and melt-textured samples of YBCO. When the magnetic field is rotated in the **ab**-plane and from the **c**-axis to the **ab**-plane, the ac magnetic susceptibility and critical current density have peaks versus angle when the field is aligned with twin boundaries. At these peaks, a peak effect is observed in the magnetic-field dependence of j_c. Our results indicate that twins are strong pinning sites in the studied range of magnetic fields between 1 and 20 kOe at liquid-nitrogen temperature. The features of our measurements can be interpreted in terms of pinning of the vortex lattice by a system of planar defects.

This work was partly supported by the Russian State Program on Superconductivity (Projects Nos. 93027, 93087, and 95046) and by the Russian Fund for Fundamental Research (Grant No. 96-02-18949).

[1] G. Blatter, M. V. Feigel'man, V. B. Geshkenbein *et al.*, Rev. Mod. Phys. **66**, 1125 (1994).

[2] V. K. Vlasko-Vlasov, L. A. Dorosinskii, A. A. Polyanskii *et al.*, Phys. Rev. Lett. **72**, 3246 (1994).

[3] W. K. Kwok, J. A. Fendrich, C. J. van der Beek, and G. W. Grabtree, Phys. Rev. Lett. **73**, 2614 (1994).

[4] M. Turchinskaya, D. L. Kaiser, F. W. Gayle *et al.*, Physica C **216**, 205 (1993).

[5] U. Welp, T. Gardiner, D. O. Gunter *et al.*, Phys. Rev. Lett. **74**, 3713 (1995).

[6] E. M. Gyorgy, R. B. van Dover, L. F. Schneemeyer *et al.*, Appl. Phys. Lett. **56**, 2465 (1990).

[7] R. B. Flippen, T. R. Askew, and Ruixing Liang, Physica C **231**, 352 (1994).

[8] C. A. Duran, P. L. Gammel, D. J. Bishop *et al.*, Phys. Rev. Lett. **74**, 3712 (1995).

[9] M. Oussena, P. A. J. de Groot, S. J. Porter *et al.*, Phys. Rev. B **51**, 1389 (1995).

[10] M. Oussena, P. A. J. de Groot, A. V. Volkozub *et al.*, Phys. Rev. Lett. **76**, 2559 (1996).

[11] L. M. Fisher, A. V. Kalinov, J. Mirkovič *et al.*, Appl. Suppercond. **2**, 639 (1994).

[12] A. A. Zhukov, H. Kupfer, M. Klaser *et al.*, Int. Phys. Conf. Ser. No. 148, IOP Publishing Ltd. (ed. by D. Dew-Huges), Bristol & Philadelphia (1995), p. 275.

[13] P. H. Kes, A. Pruimboom, J. van der Berg, and J. A. Mydosh, Cryogenics **29**, 228 (1989).

[14] G. Blatter, J. Rhyner, and V. M. Vinokur, Phys. Rev. B **43**, 7826 (1991).

[15] E. B. Sonin, Phys. Rev. B **48**, 10487 (1993).

[16] A. I. Larkin, M. C. Marchetti, and V. M. Vinokur, Phys. Rev. Lett. **16**, 2992 (1995).

[17] V. I. Voronkova and Th. Wolf, Physica C **218**, 175 (1993).

[18] L. M. Fisher, V. S. Gorbachev, N. V. Il'in *et al.*, Phys. Rev. B **46**, 10986 (1992).

[19] T. Roy and T. E. Mitchel, Phil. Mag. A **63**, 225 (1991).

[20] M. Tachiki and S. Takahashi, Solid State Commun. **72**, 1083 (1989).

[21] L. M. Fisher, A. V. Kalinov, S. E. Savel'ev, and V. A. Yampol'skii, in *Proceedings of the 8th Internat. Workshop on Critical Currents in Superconductors* (8th IWCC), Japan, Kitakyushu, May 27–29 (1996), p. 213.

[22] P. H. Kes, *Proceedings of the 8th Internat. Workshop on Critical Currents in Superconductors* (8th IWCC), Japan, Kitakyushu, May 27–29 (1996), p. 23.

[23] M. C. Marchetti and D. R. Nelson, Phys. Rev. B **42**, 9938 (1990).

Translation provided by the Russian Editorial office.

Magnetic vortices and thermoelectric effect in a hollow superconducting cylinder

R. M. Arutyunyan, V. L. Ginzburg, and G. F. Zharkov

P. N. Lebedev Institute of Physics, Russian Academy of Sciences, 117924 Moscow, Russia

(Submitted 8 October 1996)
Zh. Éksp. Teor. Fiz. **111**, 2175–2193 (June 1997)

The question of a surface barrier which determines the behavior of a vortex in a hollow superconducting cylinder of finite thickness in an external magnetic field is discussed, taking into account magnetic flux quantization in the cavity. The behavior of magnetic vortices in a hollow superconductor in the presence of a thermoelectric current is also considered. Pairs of magnetic vortices with opposite magnetic field orientations (vortex–antivortex pairs) are generated by this current near T_c. The thermoelectric current drives the antivortex (the vortex with oppositely directed field) out of the cylinder, whereas the vortex is ejected into the cavity and remains on the inside cylinder surface as a current. The number of magnetic flux quanta trapped inside the cylinder increases by one. The relation of this mechanism to the ‘‘giant’’ thermoelectric effect in hollow superconductors is discussed. © *1997 American Institute of Physics.* [S1063-7761(97)01806-4]

1. INTRODUCTION

The existence of a surface barrier impeding penetration of magnetic vortices into type-II superconductors in an external magnetic field was first analyzed by Bean and Livingston,[1] and then investigated in many experimental and theoretical works.[2–6] Behavior of a magnetic flux line near the plane surface of a semi-infinite superconductor in an external magnetic field $\mathbf{H}_e$ parallel to the interface has been studied, and, in particular, an expression for the free energy of a superconductor containing a vortex has been obtained with a view to describe a plane surface barrier. (The problem of a surface barrier for circular flux lines in a hollow superconductor with applied azimuthal magnetic field has also been studied recently.[7,8])

Section 2 of this paper gives a general expression describing the surface barrier in the case of a hollow superconducting cylinder of a finite thickness in an external magnetic field parallel to the cylinder axis with due account of flux quantization in the cavity. Using this expression, various limiting cases have been analyzed. The behavior of a vortex in a plate of a finite thickness driven by an external field is discussed in Sec. 3. The behavior of a vortex in a hollow cylinder at different temperatures is studied in Sec. 4, and the behavior of a vortex driven by thermoelectric current in Sec. 5. In Sec. 6 we study the possibility of generating a vortex–antivortex pair by thermocurrent, and in Sec. 7 the relation between this mechanism and the so-called ‘‘giant’’ thermoelectric effect detected in hollow superconducting structures is discussed.

2. THERMODYNAMIC POTENTIAL OF THE SYSTEM

Let us consider a hollow superconducting cylinder (Fig. 1) with inside radius r_1 and outside radius r_2, the field outside the cylinder being $\mathbf{H}_e$ and the inside field $\mathbf{H}_i$. A vortex filament is inside the superconductor carrying one flux quantum $\boldsymbol{\mu}\phi_0$, where $\phi_0=hc/2e$ (the vector $\boldsymbol{\mu}$ is aligned with the z-axis; its z-projection assumes values $\mu=\pm 1$ and 0, and the fields $\mathbf{H}_i$ and $\mathbf{H}_e$ are also parallel to the z-axis). The vortex filament $\boldsymbol{\mu}\phi_0$ is at distance x from the cavity surface, whose radius is r_1. The field $\mathbf{H}_i$ corresponds to m flux quanta (with direction $\mathbf{m}\phi_0$) trapped in the cavity. In order to solve the problem, we need to find the free energy (more exactly, the thermodynamic potential) of the system given in Fig. 1.

We start with the energy conservation in the superconductor:

$$\Delta\mathscr{E}=\Delta Q+\frac{c\Delta t}{4\pi}\int_{\sigma_e}\mathbf{E}\times\mathbf{H}_e\cdot d\boldsymbol{\sigma}+\frac{c\Delta t}{4\pi}\int_{\sigma_i}\mathbf{E}\times\mathbf{H}_i\cdot d\boldsymbol{\sigma}+\frac{c\Delta t}{4\pi}\int_{\sigma_\mu}\mathbf{E}\times\mathbf{H}_\mu\cdot d\boldsymbol{\sigma}, \tag{1}$$

when the change in the energy during time Δt is due to the heat ΔQ dissipated in the superconductor and electromagnetic energy transmitted through the inside (σ_i) and outside (σ_e) surfaces of the sample. The vortex has a normal core which is modeled by a void σ_μ with radius $r_\mu\to 0$ containing one flux quantum $\boldsymbol{\mu}\phi_0$. The field on the vortex axis is denoted by $\mathbf{H}_\mu(0)$. In the limit $r_\mu\to 0$ the field $\mathbf{H}_\mu(0)$ coincides with the magnetic induction $\mathbf{B}_\mu(0)$, i.e., the real field on the vortex axis.

Using the formulas like

$$\mathbf{a}\ \mathrm{curl}\ \mathbf{b}-\mathbf{b}\ \mathrm{curl}\ \mathbf{a}+\mathrm{div}(\mathbf{a}\times\mathbf{b})=0 \tag{2}$$

and the Gauss theorem

$$\int_V \mathrm{div}(\mathbf{F}(\mathbf{r}))dv=\int_S \mathbf{F}(\mathbf{r})\cdot d\mathbf{S}, \tag{3}$$

one can replace the surface integrals in Eq. (1) with volume integrals. We also use the Maxwell equations

$$\mathrm{curl}\ \mathbf{E}=-\frac{1}{c}\frac{\partial\mathbf{B}}{\partial t},\quad \mathrm{curl}\ \mathbf{H}=0, \tag{4}$$

where $\mathbf{E}$ and $\mathbf{B}$ are the electric and magnetic fields in the superconductor, and $\mathbf{H}$ (i.e., $\mathbf{H}_e$, $\mathbf{H}_i$, and $\mathbf{H}_\mu$) are the unscreened (unlike the magnetic induction $\mathbf{B}$) magnetic fields

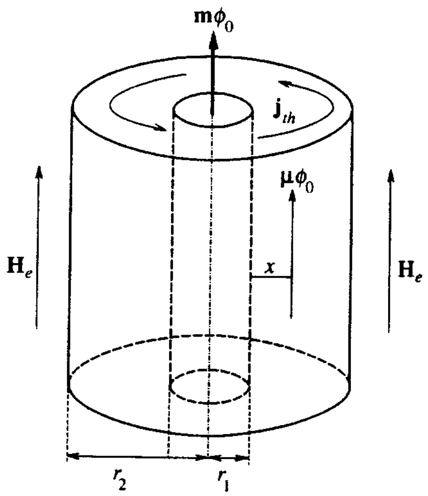


FIG. 1. Cylinder with radii r_1 and r_2 in the external magnetic field $\mathbf{H}_e$, with internal field $\mathbf{H}_i=\mathbf{m}\phi_0/\pi r_1^2$. The vortex $\boldsymbol{\mu}\phi_0$ is at a distance x from the cavity. Thermocurrent $\mathbf{j}_{\rm th}$ can flow around the inside surface (see Sec. 5).

in vacuum. The fields $\mathbf{H}$ in Eq. (1) are constant on the surfaces σ, which applies to all cylindrical surfaces whose cross sections are defined by curves of the second order, not just to the circular cylinder shown in Fig. 1.

Using the inequality $\Delta Q\leqslant T\Delta\mathscr{S}$, where $T=\text{const}$ is the sample temperature and $\mathscr{S}$ is its entropy, and introducing the superconductor free energy $\mathscr{F}_s=\mathscr{E}-T\mathscr{S}$, we transform the right-hand side of Eq. (1) to $\Delta\mathfrak{G}\leqslant 0$, where the Gibbs free energy $\mathfrak{G}$ is expressed by (compare to Refs. 9–11)

$$\mathfrak{G}=\mathscr{F}_s-\frac{1}{4\pi}\int_V \mathbf{H}_e\cdot\mathbf{B}dv+\frac{H_i^2}{8\pi}V_i. \tag{5}$$

Here $V=V_i+V_s$ is the total sample volume, V_i is the interior cavity volume, V_s is the superconductor volume, while the volume of the normal core of a vortex is $V_\mu=0$. In deriving Eq. (5), we have assumed that the external field H_e is fixed, $H_e=\text{const}$ $(\partial H_e/\partial t=0)$, but the field in the cavity is variable (it can change, for example with the temperature or the vortex distance x to the surface; only the number m of trapped magnetic flux quanta in the cavity is fixed, i.e., the magnetic fluxoid is quantized). Thus, any changes in the system lead to a decrease in the Gibbs free energy $(\Delta\mathfrak{G}\leqslant 0)$, provided that T and H_e are maintained constant with time.

The superconductor free energy $\mathscr{F}_s$ is expressed through the standard Ginzburg–Landau functional:

$$\mathscr{F}_s=\mathscr{F}_{n0}+\int_{V_s}\frac{\mathbf{B}^2}{8\pi}dv+\int_{V_s}\left\{-\alpha|\Psi|^2+\frac{\beta}{2}|\Psi|^4+\frac{1}{2m_*}\left|i\hbar\nabla\Psi+\frac{e_*}{c}\mathbf{A}\Psi\right|^2\right\}dv, \tag{6}$$

where $\mathscr{F}_{n0}$ is the free energy of a normal metal in zero magnetic field, α and β are the temperature-dependent parameters which determine the thermodynamic critical magnetic field of a massive superconductor, $H_c^2=4\pi\alpha^2/\beta$; $e_*=2e$ and $m_*=2m_e$ are the charge and mass of a Cooper pair, $\Psi=|\Psi|e^{i\Theta}$ is the complex order parameter, Θ is its phase, and $\mathbf{A}$ is the vector potential of the electromagnetic field.

The variational equation $\delta\mathfrak{G}/\delta\mathbf{A}=0$ yields, as usual, the magnetic field in the system:

$$\text{curl }\mathbf{B}=\frac{4\pi}{c}\mathbf{j}_s,\quad \mathbf{B}=\text{curl }\mathbf{A}, \tag{7}$$

$$\mathbf{j}_s=\frac{c}{4\pi\lambda^2}\left(\frac{\hbar c}{e_*}\nabla\Theta-\mathbf{A}\right),\quad \frac{1}{\lambda^2}=\frac{4\pi e_*^2|\Psi|^2}{m_*c^2}. \tag{8}$$

The boundary conditions on the inside and outside surfaces of the superconductor are:

$$\mathbf{B}|_{\sigma_i}=\mathbf{H}_i,\quad \mathbf{B}|_{\sigma_e}=\mathbf{H}_e,\quad \mathbf{B}|_{\sigma_\mu}=\mathbf{H}_\mu(0), \tag{9}$$

where the field $\mathbf{H}_e$ is fixed, and the fields in the cavity, $\mathbf{H}_i$, and on the vortex axis, $\mathbf{H}_\mu(0)$, must be calculated as a function both of the number m of magnetic flux quanta in the cavity and the distance x to the axis of the vortex $\boldsymbol{\mu}$.

The system energy is measured with respect to its value in the normal state (when $|\Psi|=0$). Let us express the free energy as $\mathfrak{G}=\mathfrak{G}_s+\mathfrak{G}_n$, where $\mathfrak{G}_n$ is the free energy of the normal state, besides, $\mathfrak{G}_s(|\Psi|=0)=0$ and $\mathbf{B}(|\Psi|=0)=\mathbf{H}_e$. Equation (5) can be transformed, using Eqs. (6)–(8) (see similar calculations in Ref. 11) to

$$\mathfrak{G}_s=\mathscr{F}_{s0}-\frac{\mathbf{H}_e}{8\pi}\int_V(\mathbf{B}-\mathbf{H}_e)dv+\frac{\hbar}{2e_*}\int_{V_s}\mathbf{j}_s\nabla\Theta dv, \tag{10}$$

$$\mathscr{F}_{s0}=\int_{V_s}\left\{-\alpha|\Psi|^2+\frac{\beta}{2}|\Psi|^4+\frac{\hbar^2}{2m_*}(\nabla|\Psi|)^2\right\}dv, \tag{11}$$

where $\mathscr{F}_{s0}$ corresponds to the condensation energy of the system.

The phase $\Theta(\mathbf{r})$ in Eq. (10) (unlike that in Ref. 11) has not one but two topological properties due to the two doubly connected regions σ_i and σ_μ shown in Fig. 1. In this connection, the phase as a function of coordinate can be expressed as

$$\Theta(\mathbf{r})=\theta_1(\boldsymbol{\rho}_1)+\theta_2(\boldsymbol{\rho}_2),\quad \nabla_{\mathbf{r}}\Theta=\nabla_{\boldsymbol{\rho}_1}\theta_1+\nabla_{\boldsymbol{\rho}_2}\theta_2, \tag{12}$$

where $\mathbf{r}$ is the integration point in Eq. (10) (we assume that $z=0$ holds); $\boldsymbol{\rho}_1$ is the radius vector connecting the center of the cavity σ_i with the point $\mathbf{r}$; $\boldsymbol{\rho}_2$ is the radius vector connecting the vortex center with the point $\mathbf{r}$; θ_1 is the angle at which point $\mathbf{r}$ is seen from the cavity center; and θ_2 is the angle at which this point is seen from the vortex center. The phases in Eq. (12) satisfy the following conditions:

$$\oint_{C_1}\nabla\theta_1 dl=2\pi m,\quad \oint_{C_2}\nabla\theta_2 dl=2\pi\mu,$$

$$\oint_{C_3}\nabla\Theta dl=2\pi(m+\mu), \tag{13}$$

where m is an integer; $\mu=0,\pm 1$; C_1 is an arbitrary closed path around the cavity σ_1, but not encircling the vortex σ_μ; C_2 is an arbitrary closed path around the vortex, but not encircling the cavity; and C_3 is a closed path encircling both σ_i and σ_μ.

From $\mathbf{j}_s=(c/4\pi)\mathrm{curl}\,\mathbf{B}$ and the formula $\mathrm{curl}\nabla\theta=0$, which applies to all functions $\theta(\mathbf{r})$, the second integral in Eq. (10) can be transformed, using Eqs. (2), (3), and (12), to

$$\frac{\hbar}{2e}\int_{V_s}\mathbf{j}_s\nabla\Theta\,dv=\frac{\phi_0}{8\pi}\,L_z\left\{\oint_{C_i}\mathbf{H}_i\nabla\theta_1 d\mathbf{l}\right.$$
$$\left.+\oint_{C_\mu}\mathbf{H}_\mu(0)\nabla\theta_2 d\mathbf{l}-\oint_{C_e}\mathbf{H}_e\nabla d\mathbf{l}\right\},$$

where L_z is the cylinder height (in what follows, $L_z=1$); C_i is a closed path on the inside surface σ_i; C_μ is a path encircling the vortex; and C_e is a closed path on the outside cylinder surface.

The field $\mathbf{B}(\mathbf{r})$ in the superconductor, the field $\mathbf{H}_\mu(0)$ on the vortex axis, and the field $\mathbf{H}_i$ in the cavity are functions of the vortex coordinate x. This dependence will be made explicit: $\mathbf{B}(\mathbf{r};x)$, $\mathbf{H}_\mu(0;x)$, and $\mathbf{H}_i(x)$. As a result, the Gibbs free energy (10) is transformed, taking into account Eq. (13), to

$$\mathfrak{G}_s=\mathscr{F}_{s0}-\frac{1}{8\pi}\int_V\mathbf{H}_e\cdot(\mathbf{B}(\mathbf{r};x)-\mathbf{H}_e)dv$$
$$+\frac{\phi_0}{8\pi}\{\mathbf{m}\cdot(\mathbf{H}_i(x)-\mathbf{H}_e)$$
$$+\boldsymbol{\mu}\cdot(\mathbf{H}_\mu(0;x)-\mathbf{H}_e)\}. \tag{14}$$

This equation is universal, since it has been derived from Eqs. (1), (5), and (10) using identities. It is exact if applied to superconductors defined by cylindrical surfaces whose cross sections are second-order curves. Equations (11) and (14) do not demand that $|\Psi(\mathbf{r})|=\mathrm{const}$. If $\boldsymbol{\mu}=0$ holds, Eq. (14) yields the Gibbs free energy of circular[11] and elliptical[12] hollow cylinders, and at $\mathbf{m}=\boldsymbol{\mu}=0$ it yields the Gibbs free energy of a superconducting plate[9] and of a solid circular cylinder[10] in an external magnetic field $\mathbf{H}_e$.

3. A VORTEX IN A PLANAR PLATE

The exact equation (14) expresses the free energy of a hollow superconductor containing a flux line in terms of the fields $\mathbf{B}(\mathbf{r})$, $\mathbf{H}_i$, and $\mathbf{H}_\mu(0)$. The field $\mathbf{B}(\mathbf{r})$ in the superconductor, the field $\mathbf{H}_i$ in the cavity, and the field $\mathbf{H}_\mu(0)$ on the vortex axis are functions of the vortex position x with respect to the cavity surface. These fields are derived from Eqs. (7) and (8) with the boundary conditions (9). This derivation can be carried out only in the case $|\Psi(\mathbf{r})|=\mathrm{const}$, and this condition is assumed hereinafter. This approximation is feasible if the magnetic field is sufficiently weak (the region near the vortex axis is considered separately).

First of all, let us consider an isolated vortex $\boldsymbol{\mu}$ in an infinite superconductor. The field $\mathbf{B}(\boldsymbol{\rho})$ with respect to the vortex axis ($\boldsymbol{\rho}=0$) is derived from Eq. (7) and expressed by the equations (the field around an isolated vortex, $|\mathbf{B}(\boldsymbol{\rho})|$, is denoted by $h(\rho)$[2,3])

$$h(\rho)=\frac{\phi_0}{2\pi\lambda^2}\,K_0\left(\frac{\rho}{\lambda}\right),$$
$$\frac{dh}{d\rho}=-\frac{\phi_0}{2\pi\lambda^3}\,K_1\left(\frac{\rho}{\lambda}\right),\quad \rho\geqslant\xi, \tag{15}$$

where $K_0(y)$ and $K_1(y)=-dK_0/dy$ are modified Bessel functions of the second kind, λ is the magnetic field penetration depth, and $\lambda^2=m_*c^2/4\pi e_*^2|\Psi|^2$. For $y\gg1$ the function $K_0(y)$ drops exponentially. For $y\ll1$ the function $K_0(y)$ has a logarithmic singularity, so Eq. (15) does not apply to the field h near the vortex axis ($\rho=0$). A more accurate analysis taking into account the equation for the order parameter $|\Psi(\rho)|$ indicates[4,5] that Eq. (15) holds for distances down to $\rho\sim\xi$, where ξ is the coherence length, $\lambda(T)=\kappa\xi(T)$, and $\kappa\gg1$ is the Ginzburg–Landau parameter for the case of a type-II superconductor. In the range of distances $\rho<\xi(T)$, the amplitude of the order parameter drops, $|\Psi(\rho)|\to0$ as $\rho\to0$. As a result, the field $h(0)$ on the vortex axis is finite and approximately equals twice the first critical field H_{c1}, which corresponds to the onset of vortex penetration into the superconductor:

$$h(0)=2H_{c1}=\frac{\phi_0}{2\pi\lambda}\,K_0\left(\frac{1}{\kappa}\right),$$
$$\left.\frac{dh(\rho)}{d\rho}\right|_{\rho=0}=-\frac{\phi_0}{2\pi\lambda^3}\,K_1\left(\frac{1}{\kappa}\right),$$
$$K_1(y)=-\frac{dK_0(y)}{dy},\quad K_0\left(\frac{1}{\kappa}\right)\approx\ln\kappa,$$
$$K_1\left(\frac{1}{\kappa}\right)\approx\kappa,\quad \kappa\gg1. \tag{16}$$

The expression for H_{c1} in Eq. (16) holds for $\kappa>20$, whereas at smaller κ this formula is quite inaccurate.[4] Note that Eq. (16) has been derived taking into account the gradient of the order parameter near the vortex axis.[1–6] Below we will use Eq. (15) for $\rho\geqslant\xi$, assuming $\lambda=\mathrm{const}$ and taking into account Eq. (16) for $\rho<\xi$.

The field $\mathbf{B}(\boldsymbol{\rho};x)$ at some point $\boldsymbol{\rho}$ in a semi-infinite superconductor generated by a vortex at distance x from a plane boundary $x=0$ is calculated using the mirror-reflection technique as a sum of two solutions of Eq. (15) for the vortex $\boldsymbol{\mu}$ at point x and its mirror reflection [the antivortex $(-\boldsymbol{\mu})$] at $-x$ with respect to the boundary:

$$\mathbf{B}_\mu(\boldsymbol{\rho};x)=\boldsymbol{\mu}h(|\boldsymbol{\rho}_1|)-\boldsymbol{\mu}h(|\boldsymbol{\rho}_2|). \tag{17}$$

We have added the index μ to the field $\mathbf{B}_\mu$ to indicate that this field is generated by the vortex. Here $\boldsymbol{\rho}_1$ is the radius vector connecting the center of the vortex $\boldsymbol{\mu}$ with the observation point $\boldsymbol{\rho}$, and $\boldsymbol{\rho}_2$ is the radius vector connecting the center of the antivortex $(-\mu)$ with the observation point $\boldsymbol{\rho}$ ($\boldsymbol{\rho}_2=2\mathbf{x}+\boldsymbol{\rho}_1$, where $2\mathbf{x}$ is the vector connecting the centers of the vortex and antivortex and perpendicular to the superconductor boundary). It is obvious that at any point $\boldsymbol{\rho}$ on the interface between the superconductor and vacuum ($x=0$) the field satisfies $\mathbf{B}_\mu(\boldsymbol{\rho};x=0)=0$, as follows from Eq. (17).

If there is a certain external field $\mathbf{H}_e$ on the interface $x=0$, an exponentially decaying function $\mathbf{H}_e e^{-x/\lambda}$ should be added to Eq. (17). In this case, the condition $\mathbf{B}_\mu(0;x)=\mathbf{H}_e$ is satisfied. By putting $\boldsymbol{\rho}_1=0$ and $\boldsymbol{\rho}_2=2\mathbf{x}$ in Eq. (17) (the point $\boldsymbol{\rho}_1=0$ is on the vortex axis), we obtain $\mathbf{B}_\mu(0;x)$ on the axis of the vortex at distance x from the interface:

$$\mathbf{B}_\mu(0;x)=\boldsymbol{\mu}[h(0)-h(2x)]+\mathbf{H}_e e^{-x/\lambda}.$$

In the case of a superconducting plate of a finite thickness d in an external magnetic field $\mathbf{H}_e$, the solution $\mathbf{B}_\mu(0;x)$is expressed as a sum of repeated mirror reflections from the two interfaces. If $d\gg\lambda$, one can take into account only the nearest reflections:

$$\mathbf{H}_\mu(0;x)=\boldsymbol{\mu}[h(0)-h(2x)-h(2(d-x))]+\mathbf{H}_e e^{-x/\lambda}+\mathbf{H}_e e^{-(d-x)/\lambda}. \tag{18}$$

(Note that $\mathbf{H}_\mu(0;x)\equiv\mathbf{B}_\mu(0;x)$, since both these functions determine the field on the vortex axis.) The solution (18) satisfies Eqs. (7) and (8) and the boundary conditions

$$\mathbf{H}_\mu(0;x)|_{x=0}=\mathbf{H}_\mu(0;x)|_{x=d}=\mathbf{H}_e; \tag{19}$$

moreover, $\mathbf{H}_\mu(0;x)|_{x=d/2}=\boldsymbol{\mu}h(0)$ if $d\gg\lambda$.

The total magnetic flux $\boldsymbol{\Phi}(x)$ in the system is the sum of two terms:

$$\boldsymbol{\Phi}(x)=\int_V \mathbf{B}(\mathbf{r};x)dv=\boldsymbol{\Phi}_0+\delta\boldsymbol{\Phi}(x),\quad \boldsymbol{\Phi}_0=\int_V \mathbf{B}(\mathbf{r})dv, \tag{20}$$

where $\boldsymbol{\Phi}_0$ is the flux without a vortex and $\delta\boldsymbol{\Phi}(x)$ is the flux associated with the vortex. The latter decreases as the vortex approaches an interface:

$$\delta\boldsymbol{\Phi}(x)=\boldsymbol{\mu}\phi_0[1-e^{-x/\lambda}-e^{-(d-x)/\lambda}], \tag{21}$$

so the vortex placed on any interface does not contribute to the total flux in the plate, $\delta\boldsymbol{\Phi}(0)=\delta\boldsymbol{\Phi}(d)=0$ since $d\gg\lambda$.

The Gibbs free energy of a superconducting plate containing a vortex for $\mathbf{H}_i=\mathbf{H}_e$ (or $m=0$) is derived from Eq. (14):

$$\mathfrak{G}_s=\mathscr{F}_{s0}-\frac{1}{8\pi}\int_V \mathbf{H}_e\cdot(\mathbf{B}(\mathbf{r};x)-\mathbf{H}_e)dv+\frac{\phi_0}{8\pi}\,\boldsymbol{\mu}\cdot(\mathbf{H}_\mu(0;x)-\mathbf{H}_e). \tag{22}$$

As a result, we find from Eq. (22), with due account of Eqs. (18) and (21),

$$\mathfrak{G}_s=\mathfrak{G}|_{\mu=0}+\mathscr{G}(x),\quad \mathfrak{G}|_{\mu=0}=\mathscr{F}_{s0}-\frac{1}{8\pi}\times\int_V \mathbf{H}_e\cdot(\mathbf{B}(\mathbf{r})-\mathbf{H}_e)dv,$$

$$\mathscr{G}(x)\left(\frac{\phi_0}{8\pi}\right)^{-1}=\mu^2[h(0)-h(2x)-h(2(d-x))]-2\boldsymbol{\mu}\cdot\mathbf{H}_e(1-e^{-x/\lambda}-e^{-(d-x)/\lambda}), \tag{23}$$

where $\mathfrak{G}|_{\mu=0}$ is the Gibbs free energy of the plate without vortices,[9] and $\mathscr{G}(x)$ is the contribution due to the vortex. Equation (23) is a generalization of the results reported in Refs. 1 and 2 to the case of a plate with a finite thickness $(d\gg\lambda)$. The problem of a chain of vortices in a plate of a finite thickness was studied using a different technique by Brongersma *et al.*[29]

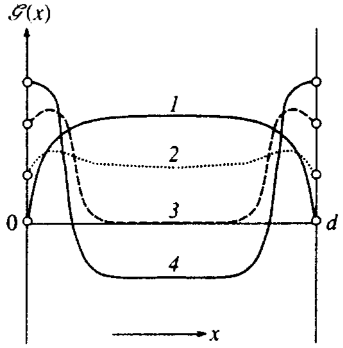


FIG. 2. The function $\mathscr{G}(x)$ [Eq. (23)] in different magnetic fields H_e: (*1*) $H_e=0$; (*2*) $H_e>0$; (*3*) $H_e=H_{c1}$; (*4*) $H_e=H_*$.

Note that for $x\sim d/2$ and $2H_e=h(0)$ it follows from Eq. (23) that $\mathscr{G}=0$, therefore the equilibrium critical field can be defined as $H_{c1}=h(0)/2$. When $H_e=H_{c1}$ holds, a vortex added to the system does not change the total energy: $\mathscr{G}(x)=0$ for $d\gg\lambda$, i.e., the vortex inside the superconductor is in thermodynamic equilibrium.

The behavior of the function $\mathscr{G}(x)$ defined by Eq. (23) is illustrated by Fig. 2. The surface barrier vanishes when the condition

$$d\mathscr{G}(x)/dx|_{x=0}=0$$

is fulfilled, whence the threshold field (i.e., the maximum ''overheat'' field) is derived using Eqs. (15) and (16):

$$H_*=\frac{\phi_0}{2\pi\lambda^2}\quad \kappa=\frac{H_c}{\sqrt{2}},$$

here H_c is the thermodynamic critical field of the superconductor. When $H>H_*$, vortices should move from the interface into the superconductor interior.

4. A VORTEX IN A HOLLOW CYLINDER

The Gibbs free energy of a hollow cylinder containing a vortex is determined by the general formula (14), which requires the total magnetic flux $\boldsymbol{\Phi}(x)$, the field $\mathbf{H}_i(x)$ in the cavity, and the magnetic field $\mathbf{H}_\mu(0;x)$ on the vortex axis. These parameters are usually determined using the mirror-reflection method, but in the case of a hollow cylinder of arbitrary dimensions it is difficult to find an exact solution like (17) because of the surface curvature. Below we consider the case of a circular cylinder characterized by radii r_1 and r_2 with a large cavity radius: $r_1\gg\lambda$. This allows us to neglect the effect of the surface curvature and use the formulas for the field generated by a vortex in the case of a plane interface.

When the total magnetic flux is calculated using Eq. (20) (the integration is performed over the entire cross section of the sample, including the inner cavity), one should take into account that, unlike the case of Eq. (21), the flux associated with a vortex is now

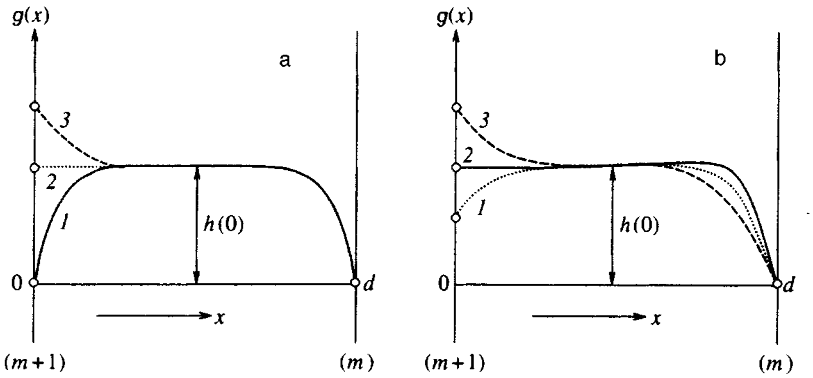


FIG. 3. Function $g(x)=8\pi\mathscr{G}(x)/\phi_0$ [Eq. (29)]: (a) at different m ($H_e=0$, $T=\text{const}$). Curves $g(x)$ connect the states with numbers of quanta $m+1$ and m (the open circles correspond to $x=0$ and $x=d$): (*1*) $m=0$; (*2*) $m=m_*$; (*3*) $m>m_*$; (b) at different temperatures T ($H_e=0$, $m<m_*$): (1) $T<T_*$; (2) $T=T_*$; (3) $T>T_*$. Here m_* is the maximum number of flux quanta which can be contained in the cavity [Eq. (31)].

$$\delta\mathbf{\Phi}(x)=\boldsymbol{\mu}\phi_0[1-e^{-(d-x)/\lambda}], \tag{24}$$

since only the fraction $\phi_0 e^{-(d-x/\lambda)}$ of the flux dissipates into the environment, whereas the other fraction $\phi_0 e^{-x/\lambda}$, which also dissipates into the environment in the case of a planar plate, now remains in the inner cavity and contributes to the trapped field.

The field in the cavity $\mathbf{H}_i(x)$ is a sum of the field $\mathbf{H}_{i0}$ in the absence of vortices and $\delta\mathbf{H}_i(x)$ associated with the vortex:

$$\mathbf{H}_i(x)=\mathbf{H}_{i0}+\delta\mathbf{H}_i(x),\quad \delta\mathbf{H}_i(x)=\boldsymbol{\mu}\,\frac{\phi_0}{\pi r_1^2}\,e^{-x/\lambda}Z_m. \tag{25}$$

The field $\mathbf{H}_{i0}$ corresponds to m flux quanta contained in the cavity and is expressed as follows:[11]

$$\mathbf{H}_{i0}=\mathbf{m}\,\frac{\phi_0}{\pi r_1^2}\,Z_m+\mathbf{H}_e Z_e,\quad Z_m=\frac{D_0}{D_1},\quad Z_e=\frac{2}{\rho_1^2 D_1},$$
$$D_0=K_0(\rho_1)I_0(\rho_2)-I_0(\rho_1)K_0(\rho_2),$$
$$D_1=K_2(\rho_1)I_0(\rho_2)-I_2(\rho_1)K_0(\rho_2), \tag{26}$$

where $K_n(x)$ and $I_n(x)$ are modifed Bessel functions, $\rho_1=r_1/\lambda$, and $\rho_2=r_2/\lambda$. The factors Z_m and Z_e in Eq. (26) are functions of temperature and system dimensions, and describe the shielding properties of a superconducting cylinder of finite thickness. In the case at hand [$d\gg\lambda(T)$, $r_1\gg\lambda(T)$], the shielding factor satifies $Z_m\approx 1$ at all realistic temperatures, and the factor Z_e is exponentially small. In the limiting case of a temperature very close to T_c, in the limit $\lambda\to\infty$ and $d/\lambda\to 0$, we have $Z_m\to 0$ and $Z_e\to 1$. Thus, the shielding factor Z_m in Eq. (26) accounts for the fact that the trapped flux satisfies $\mathbf{\Phi}_{i0}=\mathbf{H}_{i0}\pi r_1^2\to 0$ as $\lambda\to\infty$, i.e., as the cylinder becomes transparent for magnetic field, although the number m of flux quanta remains constant. In this case $\mathbf{H}_i\to\mathbf{H}_e$. We assume that the condition $d\gg\lambda$ holds, therefore $Z_m=1$ and $Z_e=0$.

The flux in the system $\delta\mathbf{\Phi}(x)$ associated with the vortex and the additional field in the cavity $\delta\mathbf{H}_i(x)$ due to the vortex are

$$\delta\mathbf{\Phi}(x)=\boldsymbol{\mu}\phi_0(1-e^{-(d-x)/\lambda}),\quad \delta\mathbf{H}_i(x)=\boldsymbol{\mu}\,\frac{\phi_0}{\pi r_1^2}\,e^{-x/\lambda}, \tag{27}$$

respectively.

When using Eq. (14), one must know the total flux in the system given by Eq. (20), the field $\mathbf{H}_i(x)$ in the cavity given by Eq. (25), and the field $\mathbf{H}_\mu(0;x)$ on the vortex axis. Instead of Eq. (18), the field on the vortex axis is determined now by the equation [here, as in previous calculations we have, $\mathbf{H}_\mu(0;x)\equiv\mathbf{B}_\mu(0;x)$]

$$\mathbf{H}_\mu(0;x)=\boldsymbol{\mu}[h(0)-h(2x)-h(2(d-x))]+\mathbf{H}_i(x)e^{-x/\lambda}$$
$$+\mathbf{H}_e e^{-(d-x)/\lambda}. \tag{28}$$

Using Eqs. (24)–(28), let us express the Gibbs free energy (14) as $\mathfrak{G}_s=\mathfrak{G}(m)+\mathscr{G}(x)$, where $\mathfrak{G}(m)$ is the system energy in the absence of vortices:[11]

$$\mathfrak{G}(m)=F_{s0}-\frac{1}{8\pi}\int_V\mathbf{H}_e\cdot(\mathbf{B}(\mathbf{r})-\mathbf{H}_e)dv$$
$$+\frac{\phi_0}{8\pi}\,\mathbf{m}\cdot(\mathbf{H}_{i0}-\mathbf{H}_e),$$

and $\mathscr{G}(x)$ is the energy due to the vortex:

$$\mathscr{G}(x)\left(\frac{\phi_0}{8\pi}\right)^{-1}=\mu^2[h(0)-h(2x)-h(2(d-x))]$$
$$-2\boldsymbol{\mu}\cdot\mathbf{H}_e(1-e^{-(d-x)/\lambda})+[2\boldsymbol{\mu}\cdot\mathbf{m}e^{-x/\lambda}$$
$$+\mu^2 e^{-2x/\lambda}]\,\frac{\phi_0}{\pi r_1^2}. \tag{29}$$

For $x=d$ we have $\mathscr{G}(d)=0$, i.e., a vortex placed on the outside surface does not affect the system energy. If the vortex is placed on the inside surface ($x=0$), this means, in fact, that the system contains $m+\mu$ quanta in the cavity. Therefore the following condition should hold:

$$\mathfrak{G}(m)+\mathscr{G}(x)|_{x=0}=\mathfrak{G}(m+\mu),$$

where $\mathfrak{G}(m)$ is the Gibbs free energy of the system without vortices. One can easily prove that this condition is satisfied exactly:

$$\mathscr{G}(0)=\frac{\phi_0}{8\pi}\left[-2\boldsymbol{\mu}\cdot\mathbf{H}_e+(2\boldsymbol{\mu}\cdot\mathbf{m}+\mu^2)\,\frac{\phi_0}{\pi r_1^2}\right]$$
$$=\mathfrak{G}(m+\mu)-\mathfrak{G}(m).$$

Note that a vortex placed on the inside surface $(x=0)$ is just a current encircling the cavity and maintaining an additional flux quantum. It does not lead to any singularity in the amplitude of the order parameter $|\Psi(x=0)|$.

The behavior of the function $\mathcal{G}(x)$ given by Eq. (29) in the case of a hollow cylinder with a wall thickness d is illustrated by Fig. 3a (we consider the case of $\mathbf{H}_e=0$). The barrier preventing penetration of a vortex from the cavity into the superconductor vanishes when the condition $\mathcal{G}'(0)=\partial\mathcal{G}(x)\partial x|_{x=0}=0$ is satisfied, i.e., with due account of Eq. (16),

$$\mathcal{G}'(0)=\frac{\phi_0}{8\pi^2\lambda^3}\left[\kappa-2(\boldsymbol{\mu}\cdot\mathbf{m}+\mu^2)\,\frac{\lambda^2}{r_1^2}\right]=0, \tag{30}$$

whence we determine either the maximum number m_* of flux quanta which can be trapped in the cavity at a given temperature:

$$m_*=\frac{\kappa}{2}\,\frac{r_1^2}{\lambda^2(T)}-1, \tag{31}$$

or the maximum ''overheat'' temperature T_* above which a field corresponding to m trapped flux quanta cannot be confined in the cavity:

$$\frac{T_*}{T_c}=1-2\kappa\,\frac{\xi_0^2}{r_1^2}\,(\boldsymbol{\mu}\cdot\mathbf{m}+\mu^2). \tag{32}$$

For $T>T_*$ a flux quantum should be ejected from the cavity (Fig. 3b, curve *3*).

5. VORTEX IN THE PRESENCE OF THERMOELECTRIC CURRENT

The giant thermoelectric effect observed in hollow bimetallic cylinders[14] was investigated previously[13] by solving the model problem of a homogeneous cylinder carrying normal current $\mathbf{j}_{\text{th}}$ circulating around a cavity to simulate the real thermoelectric current. In the following sections, we generalize the above formulas to include the thermoelectric current $\mathbf{j}_{\text{th}}$. The energy conservation is expressed in this case (we assume that the external field $\mathbf{H}_e$ is zero) by

$$\Delta\mathcal{E}=\Delta Q+\frac{c\Delta t}{4\pi}\int_{\sigma_i}\mathbf{E}\times\mathbf{H}_i\cdot d\boldsymbol{\sigma}+\frac{c\Delta t}{4\pi}\int_{\sigma_\mu}\mathbf{E}\times\mathbf{H}_\mu\cdot d\boldsymbol{\sigma}$$
$$-\Delta t\int_{V_s}\mathbf{E}\cdot\mathbf{j}_{\text{th}}dv, \tag{33}$$

where the last term on the right is the work done by the electric field $\mathbf{E}=-c^{-1}\partial\mathbf{A}/\partial t$ on the predetermined current $\mathbf{j}_{\text{th}}=b\nabla T$. We assume that this current flows in the plane perpendicular to the cylinder axis, $\partial\mathbf{j}_{\text{th}}/\partial t=0$, curl $\mathbf{j}_{\text{th}}=0$, $j_{\text{th}}=Q_0/r$, where $Q_0=b\Delta T/\pi$. The total current around the cavity in the normal state is $I_{\text{th}}=Q_0\,\ln(r_2/r_1)$, and the related field in the plane is $H_{\text{th}}=4\pi c^{-1}I_{\text{th}}$.

The Gibbs free energy of the system (provided that $T=$const and $\mathbf{j}_{\text{th}}=$const) is expressed similarly to Eq. (5):

$$\mathfrak{G}=\mathcal{F}_s+\frac{H_i^2}{8\pi}\,V_i-\frac{1}{c}\int_{V_s}\mathbf{A}\cdot\mathbf{j}_{\text{th}}dv, \tag{34}$$

where $\mathcal{F}_s$ is given by Eq. (6), and the variation $\delta\mathfrak{G}/\delta\mathbf{A}$ yields the equation

$$\text{curl }\mathbf{B}=\frac{4\pi}{c}\,(\mathbf{j}_s+\mathbf{j}_{\text{th}}), \tag{35}$$

where $\mathbf{j}_s$ is expressed by Eq. (8). Equation (34) is easily converted to a form similar to Eq. (10) (recall that $\mathfrak{G}=\mathfrak{G}_s+\mathfrak{G}_n$):

$$\mathfrak{G}_s=\mathcal{F}_{s0}+\frac{\hbar}{2e_*}\int_{V_s}\mathbf{j}_s\cdot\nabla\Theta dv-\frac{1}{2c}\int_{V_s}(\mathbf{A}-\mathbf{A}_n)\cdot\mathbf{j}_{\text{th}}dv,$$

where $\mathbf{A}_n$ is the vector potential in the normal state, i.e., $\mathbf{A}_n=\mathbf{A}(|\Psi|=0)$, and, using Eqs. (12), (13), and (35), to a form similar to Eq. (14):

$$\mathfrak{G}_s=\mathcal{F}_{s0}+\frac{\phi_0}{8\pi}\,\{\mathbf{m}\cdot\mathbf{H}_i(x)+\boldsymbol{\mu}\cdot\mathbf{H}_\mu(0;x)\}+\frac{\phi_0}{4\pi}\,\mathbf{m}\cdot\mathbf{H}_{\text{th}}$$
$$-\frac{\phi_0}{4\pi}\,\boldsymbol{\mu}\cdot\mathbf{H}_{\text{th}}\,\frac{\mathcal{L}(x)}{\mathcal{L}_0}+\frac{\lambda^2}{4\mathcal{L}_0}\,\mathbf{H}_{\text{th}}\cdot(\mathbf{H}_i(x)-\mathbf{H}_{\text{th}}), \tag{36}$$

where $\mathcal{L}(x)=\ln(r_2/(r_1+x))$, $\mathcal{L}_0=\mathcal{L}(0)=\ln(r_2/r_1)$. Equation (36) is exact, but after substituting $\mathbf{H}_i(x)$ and $\mathbf{H}_\mu(0;x)$ one has to use approximate expressions similar to those of the mirror-reflection method.

As a result, we obtain $\mathfrak{G}_s=\mathfrak{G}(m)+\mathcal{G}(x)$, where $\mathfrak{G}(m)\equiv\mathfrak{G}_s|_{\mu=0}$ is identical to the expression in Ref. 13:

$$\mathfrak{G}(m)=\mathcal{F}_{s0}+\frac{\phi_0}{8\pi}\,\mathbf{m}\cdot\mathbf{H}_{i0}-\frac{\phi_0}{4\pi}\,\mathbf{m}\cdot\mathbf{H}_{\text{th}}$$
$$+\frac{\lambda^2}{4\mathcal{L}_0}\,\mathbf{H}_{\text{th}}\cdot(\mathbf{H}_{i0}-\mathbf{H}_{\text{th}}), \tag{37}$$

where the field $\mathbf{H}_{i0}$ in the cavity containing m flux quanta is given by[13]

$$\mathbf{H}_{i0}=\mathbf{m}\,\frac{\phi_0}{\pi r_1^2}\,Z_m+\mathbf{H}_{\text{th}}Z_{\text{th}},\quad Z_m=\frac{D_0}{D_1},\quad Z_{\text{th}}=\frac{2\lambda^2}{r_1^2}\,Z_m. \tag{38}$$

As $\lambda\to\infty$, $Z_{\text{th}}\to 1$; for $d\gg\lambda_1$ and $r_1\gg\lambda$ the shielding factor is $Z_{\text{th}}=(2\lambda^2/r_1^2)Z_m\ll Z_m$. By setting $Z_m=1$ and $Z_{\text{th}}=0$, we obtain

$$\mathbf{H}_i(x)=\mathbf{H}_{i0}+\delta\mathbf{H}_i(x),$$

$$\delta\mathbf{H}_i(x)=\boldsymbol{\mu}\,\frac{\phi_0}{\pi r_1^2}\,e^{-x/\lambda},$$

$$\mathbf{H}_\mu(0,x)=\boldsymbol{\mu}[h(0)-h(2x)-h(2(d-x))]+\mathbf{H}_i(x)e^{-x/\lambda}.$$

The contribution to the Gibbs free energy $\mathcal{G}(x)$ due to the vortex is

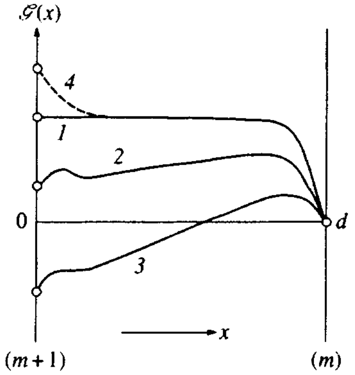


FIG. 4. Function $\mathscr{G}(x)$ [Eq. (39)] at $m=m_*$ [see Eq. (31)] and different thermocurrents j_{th}: (*1*) $j_{\text{th}}=0$; (*2* and *3*) $j_{\text{th}}>0$; (*4*) $j_{\text{th}}<0$. Curve *2* has a minimum, which disappears at higher j_{th} (curve *3*).

$$\mathscr{G}(x)\left(\frac{\phi_0}{8\pi}\right)^{-1}=\mu^2[h(0)-h(2x)-h(2(d-x))]+[2\boldsymbol{\mu}\cdot\mathbf{m}+\mu^2e^{-x/\lambda}]\,\frac{\phi_0}{\pi r_1^2}\,e^{-x/\lambda}-2\boldsymbol{\mu}\cdot\mathbf{H}_{\text{th}}\,\frac{\mathscr{L}(x)}{\mathscr{L}_0}. \tag{39}$$

For $x=d$ we have $\mathscr{G}(d)=0$, as expected. At $x=0$ the condition $\mathscr{G}(0)=\mathfrak{G}(m+\mu)-\mathfrak{G}(m)$ is satisfied exactly.

The expression for $\mathscr{G}'(0)=\partial\mathscr{G}/(x)\partial x|_{x=0}$ has the form

$$\mathscr{G}'(0)=\frac{\phi_0^2}{8\pi^2\lambda^3}\left[\kappa-2(\boldsymbol{\mu}\cdot\mathbf{m}+\mu^2)\,\frac{\lambda^2}{r_1^2}+2\boldsymbol{\mu}\cdot\mathbf{H}_{\text{th}}\,\frac{\pi\lambda^3}{\mathscr{L}_0\phi_0r_1}\right]. \tag{40}$$

Comparison with Eq. (30) demonstrates that the right-hand side of Eq. (40) contains the additional term $\sim\boldsymbol{\mu}\cdot\mathbf{H}_{\text{th}}$, which can be either positive or negative, depending on the direction of current $\mathbf{j}_{\text{th}}$ and field $\mathbf{H}_{\text{th}}$ generated by this current.

The behavior of the Gibbs free energy $\mathscr{G}(x)$ as a function of the thermoelectric current when the cavity contains the largest possible magnetic flux $m_*\phi_0$ [Eq. (31)] is illustrated by Fig. 4. Curve *1* corresponds to the case $j_{\text{th}}=0$, when $\boldsymbol{\mu}\cdot\mathbf{H}_{\text{th}}=0$ holds and the derivative satisfies $\mathscr{G}'(0)=0$ [Eq. (30)], curves *2* and *3* correspond to $j_{\text{th}}>0$, for $\boldsymbol{\mu}\cdot\mathbf{H}_{\text{th}}>0$ and $\mathscr{G}'(0)>0$, and curve *4* to $j_{\text{th}}<0$, for $\boldsymbol{\mu}\cdot\mathbf{H}_{\text{th}}<0$ and $\mathscr{G}'(0)<0$.

This means that, in the presence of a thermoelectric current generating a magnetic field $\mathbf{H}_{\text{th}}$ in the same direction as the captured magnetic flux $\mathbf{m}_*\phi_0$ in the cavity [Eq. (31)], the derivative satisfies $\mathscr{G}'(0)>0$ because $\boldsymbol{\mu}\cdot\mathbf{H}_{\text{th}}>0$. Since the force acting on the vortex $F(x)=-\partial\mathscr{G}(x)/\partial x$ is directed towards the cavity at $x=0$, the thermoelectric current effectively blocks the trapped magnetic flux in the cavity and prevents ejection of magnetic vortices from the superconductor (curve 3 in Fig. 4), which would take place at higher temperatures in the absence of the thermoelectric current j_{th} (Fig. 3b).

If the current is directed oppositely ($\boldsymbol{\mu}\cdot\mathbf{H}_{\text{th}}<0$, curve *4* in Fig. 4), then $\mathscr{G}'(0)<0$, the trapped magnetic flux $m_*\phi_0$ cannot be contained in the cavity, and flux lines should be formed and ejected from the superconductor. Whether this predicted effect really takes place can be checked experimentally.

Note also that the function $\mathscr{G}(x)$ has a minimum at $x\sim(2-3)\lambda$ (curve *2* in Fig. 4). This means that a vortex can occupy a metastable position at some distance from the cavity. Here we do not discuss this effect in detail.

Now compare the behavior of the vortex and antivortex (i.e., a vortex containing the magnetic flux of the opposite sign) inside a superconductor with thermoelectric current $\mathbf{j}_{\text{th}}$. It follows from Eqs. (39) and (40) that $\mathscr{G}(0)<0$ and $\mathscr{G}'(0)>0$ in the case of a vortex ($\mu=1$, $\boldsymbol{\mu}\cdot\mathbf{H}_{\text{th}}>0$) and in a sufficiently strong thermoelectric field $\mathbf{H}_{\text{th}}$ (curves *3* and *4* in Fig. 5a). This means that a force directed towards the cavity acts on a vortex near the inside surface. In the case of an antivortex ($\mu=-1$, $\boldsymbol{\mu}\cdot\mathbf{H}_{\text{th}}<0$), we have $\mathscr{G}(0)>0$ and $\mathscr{G}'(0)<0$, i.e., the driving force is directed from the cavity (curves *3* and *4* in Fig. 5b). Thus, if there is a vortex–antivortex couple, its components can be driven apart by thermocurrent $\mathbf{j}_{\text{th}}$ under certain conditions. This effect can lead to important consequences, which will be discussed in the next section.

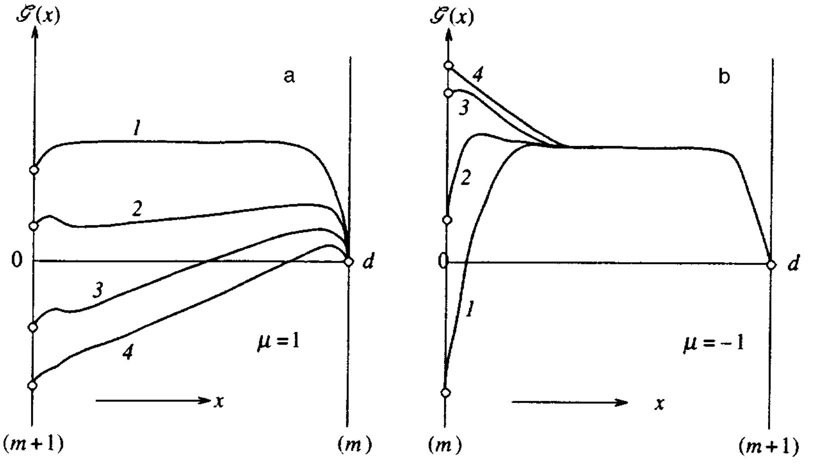


FIG. 5. Function $\mathscr{G}(x)$ [Eq. (39)] at different currents j_{th} in the presence of (a) vortex $\mu=1$ and (b) antivortex $\mu=-1$. Here $m<m_*$ and (*1*) $j_{\text{th}}=0$; (*2–4*) j_{th} increases.

6. GENERATION OF A VORTEX–ANTIVORTEX PAIR BY THERMOCURRENT

In this section, we calculate the Gibbs free energy when the system shown in Fig. 1 contains two vortices, $\boldsymbol{\mu}_1\phi_0$ and $\boldsymbol{\mu}_2\phi_0$, located on a radius vector at distances x_1 and x_2 from the cavity surface (r_1).

Using the same technique as in previous sections, we easily obtain an exact formula which is a generalization of Eq. (36):

$$\mathfrak{G}_s=\mathscr{F}_{s0}+\frac{\phi_0}{8\pi}\,[\mathbf{m}\cdot\mathbf{H}_i(x_1,x_2)+\boldsymbol{\mu}_1\cdot\mathbf{H}_{\mu 1}(0;x_1,x_2)+\boldsymbol{\mu}_2 \times\mathbf{H}_{\mu 2}(0;x_1,x_2)]+\frac{\phi_0}{4\pi}\left[\mathbf{m}\cdot\mathbf{H}_{\rm th}+\boldsymbol{\mu}_1\cdot\mathbf{H}_{\rm th}\frac{\mathscr{L}(x_1)}{\mathscr{L}_0} +\boldsymbol{\mu}_2\cdot\mathbf{H}_{\rm th}\,\frac{\mathscr{L}(x_2)}{\mathscr{L}_0}\right]+\frac{\lambda^2}{4\mathscr{L}_0}\,\mathbf{H}_{\rm th}\cdot(\mathbf{H}_i(x_1,x_2)-\mathbf{H}_{\rm th}). \tag{41}$$

Magnetic fields on the vortex axes $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are determined using the mirror-reflection technique:

$$\begin{aligned}\mathbf{H}_{\mu 1}(0;x_1,x_2)&=\boldsymbol{\mu}_1 H_{c0}f(x_1)+\boldsymbol{\mu}_2 H_{c0}f(x_1,x_2)\\&\quad+\mathbf{H}_i(x_1,x_2)e^{-x_1/\lambda},\\ \mathbf{H}_{\mu 2}(0;x_1,x_2)&=\boldsymbol{\mu}_2 H_{c0}f(x_2)+\boldsymbol{\mu}_1 H_{c0}f(x_1,x_2)\\&\quad+\mathbf{H}_i(x_1,x_2)e^{-x_2/\lambda},\end{aligned} \tag{42}$$

where

$$\begin{aligned}f(x)&=K_0(0)-K_0\left(\frac{2x}{\lambda}\right)-K_0\left(\frac{2d-2x}{\lambda}\right),\\ f(x_1,x_2)&=K_0\left(\frac{|x_1-x_2|}{\lambda}\right)-K_0\left(\frac{x_1+x_2}{\lambda}\right)\\&\quad-K_0\left(\frac{2d-x_1-x_2}{\lambda}\right).\end{aligned} \tag{43}$$

The functions $f(x_1)$, $f(x_2)$, and $f(x_1,x_2)$ are introduced to take into account both the intrinsic fields of each vortex and the fields generated by the counterpart vortex, as well as the contributions of nearest mirror reflections from both interfaces ($x=0$ and $x=d$). The field around each vortex is described by the modified Bessel functions $K_0(\rho)$ and $K_1(\rho)$ for $\rho=x/\xi\geqslant 1$, but on the vortex axis $K_0=\ln\kappa$ and $K_1(0)=\kappa$ in accordance with Eq. (16).

For convenience let us write the expressions for all typical fields in the problem:

$$H_{c0}=\frac{\phi_0}{2\pi\lambda^2},\quad H_{c1}=\frac{\phi_0}{4\pi\lambda^2}\,\ln\,\kappa,\quad H_c=\frac{\phi_0\kappa}{2\sqrt{2}\pi\lambda^2}.$$

The field inside the cavity is

$$\mathbf{H}_i(x_1,x_2)=\mathbf{H}_{i0}+\delta\mathbf{H}_i(x_1,x_2),$$

where $\mathbf{H}_{i0}$ is defined by Eq. (38), and the increase in the cavity field due to the vortices $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ is

$$\delta\mathbf{H}_i(x_1,x_2)=\boldsymbol{\mu}_1\,\frac{\phi_0}{\pi r_1^2}\,e^{-x_1/\lambda}+\boldsymbol{\mu}_2\,\frac{\phi_0}{\pi r_2^2}\,e^{-x_2/\lambda} \tag{44}$$

(here, as in the previous sections, $Z_m=1$ and $Z_{\rm th}=0$).

By expressing $\mathfrak{G}_s$ as $\mathfrak{G}_s=\mathfrak{G}(m)+\mathscr{G}(x_1,x_2)$, where $\mathfrak{G}(m)$ is defined by Eq. (37), we obtain an equation for the contribution of the vortices:

$$\begin{aligned}g(x_1,x_2)&=\mathscr{G}(x_1,x_2)\left(\frac{\phi_0 H_{c0}}{8\pi}\right)^{-1}\\&=\mu_1^2 f(x_1)+\mu_2^2 f(x_2)+2\boldsymbol{\mu}_1\cdot\boldsymbol{\mu}_2 f(x_1,x_2)\\&\quad+2\,\frac{\lambda^2}{r_1^2}\,[\mu_1^2 e^{-2x_1/\lambda}+\mu_2^2 e^{-2x_2/\lambda}\\&\quad+2\boldsymbol{\mu}_1\cdot\boldsymbol{\mu}_2 e^{-(x_1+x_2)/\lambda}+2\boldsymbol{\mu}_1\cdot\mathbf{m}e^{-x_1/\lambda}\\&\quad+2\boldsymbol{\mu}_2\cdot\mathbf{m}e^{-x_2/\lambda}]+a[\boldsymbol{\mu}_1\cdot\mathbf{e}_{\rm th}\mathscr{L}(x_1)\\&\quad+\boldsymbol{\mu}_2\cdot\mathbf{e}_{\rm th}\mathscr{L}(x_2)].\end{aligned} \tag{45}$$

Here $\mathbf{e}_{\rm th}=\mathbf{H}_{\rm th}/H_{\rm th}$ and $a=2H_{\rm th}/\mathscr{L}_0 H_{c0}$.

The free energy (45) satisfies the necessary conditions, namely, at $\mu_2=0$ the expression for $\mathscr{G}(x_1,x_2)$ coincides with Eq. (39), if $\boldsymbol{\mu}_2=-\boldsymbol{\mu}_1$ (i.e., a vortex and antivortex are contained in the superconductor at the same time), $g(x_1,x_2)|_{x_1=x_2=x_0}=0$ for all $x_0=x_1=x_2$. The condition $g(x_0,x_0)=0$ means that nothing happens at this point, that is, the presence of a vortex and antivortex at one point x_0 does not change the system energy because the fields generated by them cancel out $[(\boldsymbol{\mu}_1+\boldsymbol{\mu}_2)\phi_0=0]$ and, therefore, do not affect the order parameter of the superconductor. This means that a vortex and antivortex pair can be generated through fluctuations, and this process does not require energy supply. But when a vortex and antivortex are being separated, the forces acting on them are directed oppositely. The vortex and antivortex are actually attracted to each other,[2,3,6] whereas the thermocurrent tends to separate them by driving the vortex to the cavity and the antivortex outside because the two last terms on the right-hand side of Eq. (45) have opposite signs for $\boldsymbol{\mu}_2=-\boldsymbol{\mu}_1$. The function $g(x_1,x_2)$ in Eq. (45) reflect the presence of countervailing factors, such as the interaction between the vortices and between the vortices and the cavity surface.

In analyzing the function (45), we should determine the point (x_1,x_2) at which the variation $\delta g(x_1,x_2)=0$, i.e., when the barrier preventing vortex separation vanishes. One can check that less energy is needed when a vortex–anitvortex pair is generated on the inside surface r_1 (i.e., $x_1=x_2=0$). It is rather obvious that the inside surface is the best place for the beginning of vortex–antivortex separation since the field $\mathbf{H}_{\rm th}(x)$ generated by the thermocurrent $\mathbf{j}_{\rm th}$ is largest there and acts on the vortex and antivortex in opposite directions. The resulting antivortex is driven outside and carries away magnetic flux $-\phi_0$, whereas the vortex remains on the cavity surface $(x_1=0)$. Thus, if the system previously contained magnetic flux $\Phi_2=m\phi_0$, where Φ_2 is the total magnetic flux inside the circle with radius r_2, the system contains magnetic flux $\Phi_2=(m+1)\phi_0$ after separation of the vortex and antivortex. Note that, if the vortex axis coincides with the cavity surface $x_1=0$, the field on its axis equals the field inside the cavity, and its currents coincide with the currents encircling the cavity, i.e., the vortex has

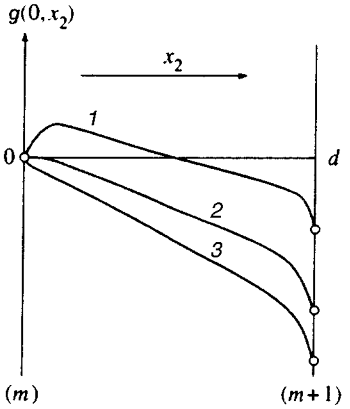


FIG. 6. The function $g(0,x_2)$ [Eq. (45)] versus the distance x_2 between the cavity surface and the antivortex at different temperatures: (*1*) $T<T_*$; (*2*) $T=T_*$; (3) $T>T_*$. Here T_* is the threshold temperature at which the antivortex is driven away from the cavity.

turned into a current encircling the cavity and does not generate a singularity in the order parameter $\Psi(x_1)$.

The behavior of the function $g(0,x_2)$ at different temperatures is illustrated by Fig. 6.

By substituting $\boldsymbol{\mu}_2=-\boldsymbol{\mu}_1$ and $x_1=0$ into Eq. (45), we determine the derivative

$$g_0' = \left.\frac{\partial g(0,x_2)}{\partial x_2}\right|_{x_2=0} = \frac{2}{\lambda}\left[K_1(0)+\frac{2\lambda^2}{r_1^2}\, m-\frac{a}{2}\,\frac{\lambda}{r_1}\right]. \qquad (46)$$

The parameter a in Eq. (46) is a function of temperature and can be expressed in a compact form as

$$a=a_0\,\frac{t}{1-t},\quad a_0=\frac{16\pi}{c}\,\frac{bT_c\lambda^2(0)}{\phi_0},\quad t=\frac{T-T_1}{T_c-T_1}. \qquad (47)$$

Here we have used the formulas

$$a=\frac{2H_{\mathrm{th}}}{\mathscr{L}_0 H_{c0}},\quad H_{\mathrm{th}}=\frac{4\pi}{c}\,\frac{b\Delta T}{\pi}\,\mathscr{L}_0,\quad H_{c0}=\frac{\phi_0}{2\pi\lambda^2},$$

$$\Delta T=T-T_1,\quad \lambda^2(T)=\frac{\lambda^2(0)}{1-T/T_c},$$

$$\frac{T-T_1}{T_c}=\left(1-\frac{T_1}{T_c}\right)t,$$

$$1-\frac{T}{T_c}=\left(1-\frac{T_1}{T_c}\right)(1-t).$$

The reduced temperature t ranges between 0 and 1.

It clearly follows from Eq. (47) that, as $T\to T_c$ $(t\to 1)$, the parameter a increases and the condition $g_0'=0$ is always satisfied. This condition determines t_* at which the barrier to separation between the vortex and antivortex vanishes, because at $t=t_*$ the attractive force between the vortex and antivortex equals the force due to the thermocurrent driving them apart.

Taking into account Eq. (47), we can write the equation for the threshold temperature t_* as

$$\kappa+2m\,\frac{\lambda_1^2}{r_1^2}\,\frac{1}{1-t}-\frac{a_0}{2}\,\frac{\lambda_1}{r_1}\,\frac{t}{(1-t)^{3/2}}=0, \qquad (48)$$

where $\lambda_1=\lambda(T=T_1)$. This cubic equation is solved using the Cardano formulas. In particular, for $m=0$ we have the temperature of transition from level $m=0$ to $m=1$, i.e., the temperature of the first jump in m:

$$t_*=1-(A+B),\quad A=\sqrt[3]{\frac{q}{2}+\sqrt{Q}},$$

$$B=\sqrt[3]{\frac{q}{2}-\sqrt{Q}},$$

$$q=\alpha\left(1-\frac{2}{3}\,\alpha+\frac{2}{3^3}\,\alpha^2\right),\quad Q=\frac{\alpha^2}{4}\left(1-\frac{4}{27}\,\alpha\right),$$

$$\alpha=\left(\frac{a_0}{2\kappa}\,\frac{\lambda_1}{r_1}\right)^2.$$

For $\alpha\ll 1$ we have $t_*=1-\alpha^{1/3}$ or

$$\frac{T_*}{T_c}=1-\Delta_1\left(\frac{a_0\lambda_1}{2\kappa r_1}\right)^{2/3},\quad \Delta_1=1-\frac{T_1}{T_c}.$$

The function $m(t)$ can be easily derived from Eq. (48) by taking the integral part of m:

$$[m]=\frac{r_1^2}{2\lambda_1^2}\left(\frac{a_0}{2}\,\frac{\lambda_1}{r_1}\,\frac{t}{\sqrt{1-t}}-\kappa(1-t)\right), \qquad (49)$$

i.e., the total magnetic flux in the system $\Phi\equiv\Phi_2(t)=[m]\phi_0$ at points of transitions $m\to m+1$ has been found as a function of temperature. For the derivative $d\Phi/dt$ we have

$$\frac{d\Phi}{dt}=\phi_0\,\frac{r_1^2}{2\lambda_1^2}\left[\frac{a_0}{2}\,\frac{\lambda_1}{r_1}\left(\frac{1}{\sqrt{1-t}}+\frac{t}{2(1-t)^{3/2}}\right)+\kappa\right]. \qquad (50)$$

Note that, as $t\to 1$, i.e., $T\to T_c$, the asymptotic limit is described by the formula $d\Phi/dt\sim(1-t)^{-3/2}\sim(T_c-T)^{-3/2}$.

7. COMPARISON TO EXPERIMENT AND DISCUSSION

In this section we discuss mostly the results of the previous section.

Van Harlingen *et al.*[14,15] discovered that in hollow bimetallic superconducting samples with a temperature gradient, the total magnetic flux was anomalously large, several orders of magnitude higher than theoretical predictions. In fact, Eq. (38) indicates that the field in the cavity $\mathbf{H}_i$ is a sum of the field due to the originally trapped magnetic flux $\mathbf{m}\phi_0$ and the field $\mathbf{H}_{\mathrm{th}}$ generated by the thermocurrent $\mathbf{j}_{\mathrm{th}}$. In superconductors, the field due to thermocurrent is suppressed by the factor $Z_{\mathrm{th}}\sim\lambda^2/r_1^2\ll 1$ in comparison with normal metal, because in a bulk superconductor the normal current and superconducting current cancel out $(\mathbf{j}_s+\mathbf{j}_{\mathrm{th}}=0)$,[16] i.e., the Meissner effect occurs (see for details Ref. 17). The resulting magnetic flux should be about $10^{-2}\phi_0$, whereas the measured fluxes[14] were of the order of tens and hundreds of flux quanta ϕ_0. There is no generally accepted interpretation of this giant thermoelectric effect. The hypothesis proposed in

Ref. 13 and later developed in Refs. 18–21, and suggesting that the giant effect can be interpreted in terms of jumps in the trapped magnetic flux due to transitions between quantum states with the fluxes $m\phi_0$ and $(m+1)\phi_0$ induced by thermocurrent raised serious objections. Namely, the quantum number m in the hollow superconductor is a topological invariant[22] and can be changed only by introducing an additional vortex $\boldsymbol{\mu}\phi_0$ through the outside sample surface. But if the external field is zero, there is a barrier to penetration of vortices on the outside interface (Fig. 4), and a vortex cannot be generated inside the sample for topological reasons. It was suggested in Refs. 13, 18–21 on an intuitive basis that there should be a mechanism for increasing the quantum number m without a real transfer of a flux quantum by a vortex, but with a direct generation of additional current around the inner cavity. This hypothesis, however, was not based on solid ground because the mechanism had not been substantiated. We believe that the generation of a vortex–antivortex pair described in Section 6 justifies our hypothesis suggested in Ref. 13.

In fact, if a vortex–antivortex pair is generated at any point in the superconductor, the quantum number m (i.e., the total flux $\Phi_2=m\phi_0$) of the system does not change, so the topological laws are not violated. If the vortex axis remains on the cavity surface $(x_1=0)$, the currents associated with the vortex encircle the cavity and contribute to the internal field $\mathbf{H}_i$, i.e., the vortex is converted to a current encircling the cavity. The field on the vortex axis equals the weak field in the cavity, and the order parameter Ψ has no singularity at point $x_1=0$. As the antivortex moves away from the cavity surface $(x_2>0)$, a region with an oppositely directed field is formed near its axis, and $\Psi=0$ holds on its axis x_2. Note that a detailed description of the field configuration and order parameter near the cavity surface for $x_2<\xi$ would require detailed calculations of the vortex structure like those discussed in Refs. 5 and 6. As the antivortex moves from the interface, the field in the cavity gradually increases, which means generation of additional current encircling it. But the total magnetic flux in the system remains $\Phi_2=m\phi_0$, and only when the antivortex approaches the outside surface within a distance approximately equal to λ and its flux gradually passes to the external space, the total flux becomes $\Phi_2=(m+1)\phi_0$. The quantum number m jumps from m to $m+1$ at the moment when the axis of the antivortex crosses the outside surface (in accordance with topological considerations) and the system is in the state with $(m+1)\phi_0$. Thus, the proposed mechanism allows the system to transfer to a higher magnetic level through generation of a vortex–antivortex pair and its separation by thermoelectric current. As a result, we have a clear physical description of a phenomenon, which probably describes the giant thermoeffect observed in experiments.

Proceeding to a more substantial discussion of experimental results,[14] note that Eq. (49) directly indicates the presence of a giant effect since the additional flux in the system due to each new quantum is two orders of magnitude higher than the value $\sim 10^{-2}\phi_0$ predicted by the simple theory.[17] The total flux measured as a function of temperature[14] is described near T_c by the formula $d\Phi/dt \sim (T_c-T)^{-3/2}$, which is in agreement with Eq. (50) for $t\rightarrow 1$. For lower t the right-hand side of Eq. (50) is a flatter function of temperature because of the large constant κ. This constant also determines the large height of the barrier for a single vortex introduced to a superconductor in the Bean-Livingston theory.[1] Note, however, that this theory applies only to the case of a superconductor with a mirror-smooth surface (when the mirror-reflection technique can be used). If the surface is rough, the measured threshold field[24] is considerably smaller than the theoretical value,[1] which indicates a smaller contribution of the last term on the right-hand side of Eq. (50) and a wider range in which the law $(T_c-T)^{-3/2}$ holds. Moreover, Fig. 5 indicates that the barrier for vortices on the outside surface, where the presence of residual magnetic fields can be important, decreases with increasing j_{th} (i.e., with the temperature of the hotter junction, $T\rightarrow T_c$). Such effects should be also taken into account in comparing between experimental and theoretical results.

Note that the simplified uniform model used in our calculations does not allow us to compare quantitatively the calculations of Sec. 6 and experimental results,[14] therefore, only qualitative comparison is possible. First of all, let us estimate the parameter a_0 in Eq. (47), which determines the magnitude of the effect. By expressing the coefficient b as α/ρ, where α is the thermoelectric coefficient and ρ is the resistivity, and using published data[23] for the constants α and ρ, we find $a_0 \sim 1-50$ for pure superconductors. Van Harlingen *et al.*[14] used bimetallic samples from pure In and Pb, but the parameters of the junction (alloy) were not known. This may be an important point, since a vortex–antivortex pair is generated, most probably, in this junction (because it is a weak link of the system) with large κ and λ, whereas the thermoelectric current j_{th}, hence the parameter a_0, are determined by the characteristics of pure bulk superconductors, where κ is usually small. As a result, the choice of the system parameters is somewhat arbitrary. Taking $T_c=5$ K, $1-T_1/T_c=10^{-2}$, $a_0=10$, $\kappa=10$, $r_1=0.1$ cm, and $\xi_0=10^{-5}$ cm, we derive from Eq. (48) t_* at which jumps in the total flux occur, namely $t_* \sim 0.99$. An anomalously large flux was detected in experiments[14] at smaller t, but this could have various causes. For example, the samples used in experiments[14] were toroids with rectangular cross sections, so the geometrical factors affecting vortex generation were notably different from the case of an infinite cylinder. The effects of surface roughness and junctions have been mentioned above. As follows from the data reported by Mkrtchan *et al.*,[25] the height of the barrier to vortex penetration on an interface between two superconductors with very different λ significantly diminishes. Note also that a vortex–antivortex pair can be generated not in the form of two antiparallel flux lines, but as a closed ring of finite dimension, like a vortex ring in a superfluid helium,[26,27] which takes less energy (see also Ref. 30). All these factors can have considerable effect on the height of the barrier to the pair formation.

Thus, the theory developed in this paper allows us, in principle, to interpret the giant thermoelectric effect,[14] although additional investigation is needed taking into account real experimental conditions.

In conclusion, note that the proposed mechanism of the flux quanta generation can be also applied to the problem of superhigh magnetic fields around rotating neutron stars, whose high-density matter may be in a superconducting or superfliud state.[28]

The work was supported by the Russian Fund for Fundamental Research (Grant No. 94-02-05306).

[1] S. R. Bean and J. D. Livingston, Phys. Rev. Lett. **12**, 14 (1964).

[2] P. G. De Gennes, *Superconductivity of Metals and Alloys*, W. A. Benjamin, Inc., New York–Amsterdam (1968).

[3] M. Tinkham, *Introduction to Superconductivity*, McGraw Hill, New York (1975).

[4] D. Saint-James, G. Sarma, and E. Thomas, *Type II Superconductivity*, Oxford University Press (1969).

[5] A. A. Abrikosov, *Fundamentals of the Theory of Metals*, North-Holland, Amsterdam, 1988.

[6] V. V. Shmidt, *Introduction to the Physics of Superconductors* [in Russian], Nauka, Moscow (1982).

[7] E. Altshuler and R. Mulet, J. Supercond. **8**, 779 (1995).

[8] Yu. Genenko, Phys. Rev. B **49**, 6950 (1994).

[9] V. L. Ginzburg, Usp. Fiz. Nauk **48**, 25 (1952).

[10] V. L. Ginzburg, Zh. Éksp. Teor. Fiz. **34**, 113 (1958) [Sov. Phys. JETP **7**, 78 (1958)].

[11] R. M. Arutyunyan and G. F. Zharkov, Zh. Éksp. Teor. Fiz. **78**, 1530 (1980) [Sov. Phys. JETP **51**, 768 (1980)].

[12] E. A. Demler and G. F. Zharkov, Sverkhprovodimost': Fizika, Khimiya, Tekhnologiya **8**, 276 (1995).

[13] R. M. Arutyunyan and G. F. Zharkov, Zh. Éksp. Teor. Fiz. **83**, 1115 (1982) [Sov. Phys. JETP **56**, 632 (1982)].

[14] D. J. Van Harlingen, D. F. Heidel, and J. C. Garland, Phys. Rev. B **21**, 1842 (1980).

[15] D. J. Van Harlingen, Physica B, **109**–**110**, 1710 (1982).

[16] V. L. Ginzburg, J. Phys. USSR **8**, 148 (1944).

[17] V. L. Ginzburg and G. F. Zharkov, Usp. Fiz. Nauk **125**, 19 (1978) [Sov. Phys. Usp. **21**, 381 (1978)].

[18] R. M. Arutunian and G. F. Zharkov, Phys. Lett. A **96**, 480 (1983).

[19] V. L. Ginzburg and G. F. Zharkov, J. Low Temp. Phys. **92**, 25 (1993).

[20] V. L. Ginzburg and G. F. Zharkov, Physica C **235–240**, 3129 (1994).

[21] E. A. Demler and G. F. Zharkov, Kratkie Soobshcheniya po Fiz. FIAN, No. 3–4, 44 (1995).

[22] R. Rajamaran, *Solitons and Instantons. An Introduction to Solitons and Instantons in Quantum Field Theory*, North Holland, Amsterdam (1982).

[23] *Tables of Physical Constants* [in Russian], ed. by I. K. Kikoin, Atomizdat, Moscow (1976).

[24] W. Buckel, *Supraleitung Grundlagen und Anwendungen* [in German], Bergser, Weinheim (1972).

[25] G. S. Mkrtchan, F. R. Shakirzyanova, E. A. Shapoval, and V. V. Shmidt, Zh. Éksp. Teor. Fiz. **63**, 667 (1972) [Sov. Phys. JETP **36**, 352 (1972)].

[26] C. Bauerle *et al.*, Nature (London) **382**, 332 (1996).

[27] V. M. Ruutu *et al.*, Nature (London) **382**, 334 (1996).

[28] A. D. Sedrakian and D. M. Sedrakian, Astron. J. **477**, 305 (1995).

[29] S. H. Brongersma, E. Verwey, N. J. Koeman, D. G. de Grant, and B. I. Ivlev, Phys. Rev. Lett. **71**, 2319 (1993).

[30] V. P. Galaiko, Zh. Éksp. Teor. Fiz. **50**, 1322 (1966) [Sov. Phys. JETP **23**, 878 (1966)].

Translation provided by the Russian Editorial office.

Spatial correlation of an electron and hole in quasi-two-dimensional electronic system in a strong magnetic field and its relationship to the light-scattering tensor

L. I. Korovin and I. G. Lang

A. F. Ioffe Physicotechnical Institute, Russian Academy of Sciences, 194021 St. Petersburg, Russia

S. T. Pavlov

P. N. Lebedev Physics Institute, Russian Academy of Sciences, 117924 Moscow, Russia

(Submitted 17 October 1996; resubmitted 11 December 1996)
Zh. Éksp. Teor. Fiz. **111**, 2194–2214 (June 1997)

The spatial correlation of light-generated electrons and holes in a quasi-two-dimensional electron gas in a strong magnetic field is investigated in an approximation linear in the intensity of the exciting light. The correlation is due to the interaction of the electrons and holes with longitudinal optical (LO) phonons. The theory permits calculating, on the basis of a special diagrammatic technique, two distribution functions of an electron–hole pair with respect to the distance between the electron and the hole after the emission of N phonons: first, the function determining the total number of pairs which have emitted N phonons and, second, the function related to the rank-4 light-scattering tensor in interband resonance Raman scattering of light. A special feature of the system is that the electron and hole energy levels are discrete. The calculation is performed for a square quantum well with infinitely high barriers. The distribution function and the total number of electron–hole pairs before the emission of phonons as well as the distribution function corresponding to two-phonon resonance Raman scattering are calculated. The theory predicts the appearance of several close-lying peaks in the excitation spectrum under resonance conditions. The number of peaks is related to the number of the Landau level participating in the optical transition. The distance between peaks is determined by the electron–phonon coupling constant. Far from resonance there is one peak, which is much weaker than the peaks obtained under resonance conditions. *© 1997 American Institute of Physics.* [S1063-7761(97)01906-9]

1. INTRODUCTION

As is well known, the energy levels of an electron in a quantum well in a strong magnetic field perpendicular to the plane of the well are discrete. For a certain well width there is only one size-quantization level in the well. For a wide-gap semiconductor and a nondegenerate conduction band the parabolic section of the electron spectrum can be comparable to the depth of the well. In such a situation the spectrum of an electron in a quantum well in a strong magnetic field perpendicular to the plane of the well consists of a system of discrete equidistant levels. For example, for a 0.3 eV square well and electron effective mass $m_e = 0.06 m_0$ (m_0 is the free-electron mass), the second size-quantization level leaves the well when the well width $d \simeq 40$ Å. For cyclotron energy $\hbar\Omega_e = \hbar|e|H/m_e c = \hbar\omega_{LO} = 0.03$ eV (e is the electron charge, H is the intensity of the magnetic field, c is the speed of light in vacuum, and ω_{LO} is the LO phonon frequency), six Landau levels fit in the well. Similar estimates hold for holes.

If the resonance condition

$$\Omega_e = \omega_{LO}, \tag{1}$$

holds, then real emission of a LO phonon and the transition of an electron to the last Landau level are possible (we have in mind low temperatures, at which LO phonons are not excited). The Landau levels are thereby coupled by the electron–phonon interaction.[1,2] The level $n=1$ is coupled to the level $n=0$; by virtue of the equidistant spacing the level $n=2$ is coupled to the levels $n=1$ and $n=0$ (n is the Landau quantum number). In the absence of interaction, for a general system consisting of an electron and LO phonons, the energy levels, regarded as functions of the magnetic field, cross at the point $\Omega_e = \omega_{LO}$. As n increases, the number of crossing levels increases. If $n=1$, then two levels cross: an electron with $n=1$ and an electron with $n=0$ + one phonon. For $n=2$ three levels cross: an electron with $n=1$, an electron with $n=1$ + one phonon and an electron with $n=0$ + two phonons, and so on.

We emphasize that level crossing occurs at a single point (if $n \geqslant 2$) only if the energy levels of the electron spectrum are equally spaced. Interaction with LO phonons removes the degeneracy, as a result of which there appear $n+1$ levels of the electron–phonon system which do not cross one another (Fig. 1). The distance between neighboring levels is determined by the Fröhlich coupling constant α_0 and is proportional to $\sqrt{\alpha_0}$.[3–7]

The lifting of the degeneracy of the levels of the electron–phonon system should be manifested in multiphonon resonance Raman scattering of light, which is an efficient method for investigating both bulk semiconductors and quasi-two-dimensional semiconductor structures.[8–10] The effect entails of the appearance of a series of peaks (phonon repetitions) at the frequencies $\omega_s = \omega_l - N\omega_{LO}$ in the

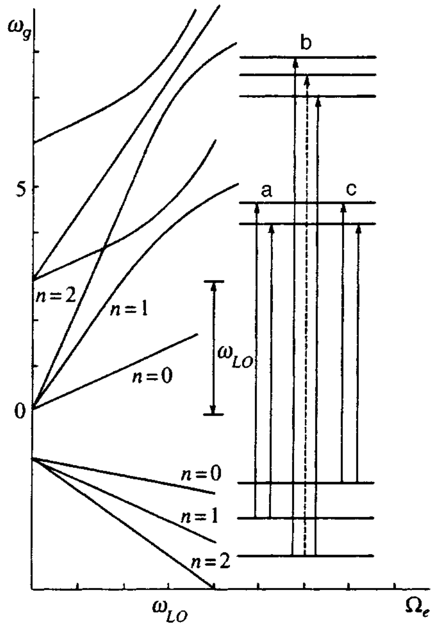


FIG. 1. Lifting of the degeneracy of the levels of the electron–phonon system at the crossing point of the levels for Landau levels $n=0, 1, 2$. The cyclotron frequency Ω_e is plotted along the abscissa and the energy of the levels in arbitrary units is plotted along the ordinate. Electron transitions from the valence band to the conduction band in the region of term splitting is shown on the right-hand side. Input resonance: a) $n=1$, b) $n=2$; output resonance: c) $n=2$. The dashed line represents a transition that is not manifested in two-phonon scattering, if the condition (1) holds. The splitting of the terms at the point r is not shown.

scattered light, where ω_s and ω_l are the frequencies of the scattered and exciting light and N is the number of the phonon repetition, which is equal to the number of phonons emitted in the scattering process.[11]

If the electronic spectrum is continuous, which happens in a three-dimensional $(3D)$ case and in a two-dimensional $(2D)$ system in the absence of a magnetic field, the phonon-repetition peaks are observed over a wide range of frequencies ω_l corresponding to fundamental absorption: the frequency must satisfy the inequality $\omega_l \geqslant \omega_g + (N-1)\omega_{LO}$ $(\hbar\omega_g$ is the band gap). Real creation of electron–hole pairs and real generation of $N-1$ phonons occur at this frequency.

The weak N dependence of the intensity of the phonon-repetition peaks made it possible to observe phonon repetitions with large N.[8,9,12–14] In a system of equidistant levels, which describes a quantum well in a magnetic field, the conditions of observation of the phonon-repetition peaks differ substantially from the conditions presented above. First, the frequencies ω_l lie in narrow intervals corresponding to the direct production of electron–hole pairs (input resonance) or direct annihilation of electron–hole pairs (output resonance). Second, the magnetic field must satisfy the condition (1) in order for real phonon generation to be possible. If $m_e \neq m_h$ $(m_h$ is the hole effective mass), then the condition (1) holds for electrons only or holes only.

Thus, a quantum well in a strong magnetic field exhibits a high degree of selectivity with respect to multiphonon resonance Raman scattering—strong scattering occurs in narrow intervals of exciting light frequency and magnetic fields. If the condition of real electron–hole pair production (or annihilation) does not hold, then the effect is sharply curtailed. As a result, the multiphonon resonance Raman scattering profiles, i.e., curves of the scattering cross sections versus the frequencies ω_l and magnetic field, contain a series of sharp peaks (see, for example, Ref. 15).

The interaction of a light-produced electron and hole with LO phonons leads to spatial correlation between them. The correlation arises because in the process of phonon generation the electron and hole in a $3D$ system wander over a volume determined by their mean-free paths.[16] In a quasi-two-dimensional system the wandering occurs in a plane, which sharply increases the intensity of the phonon-repetition peaks compared with the $3D$ case.[17]

In a $3D$ system in a strong magnetic field, the spatial correlation in a plane perpendicular to the magnetic field of an electron and hole produced by light which then emit some number of LO phonons is qualitatively different from correlation in the case $H=0$.[18] In the classical limit, the electron and hole move in circular orbits in this plane. Light produces an electron and hole at one point and then the electron and hole follow the same orbit in opposite directions. As they emit phonons, the electron and hole cross over to other orbits with other centers. The distance between the old and new centers depends only on the magnitude of the magnetic field and the wave vector of the emitted phonon.[18,19] Thus, the electron (hole) mean-free path, which is proportional to the electron–phonon coupling constant, does not appear in the description of the motion of electrons and holes in a plane perpendicular to the magnetic field. All this is also true of motion in a quantum well, if the magnetic field is perpendicular to the plane of the well.

The distribution function over the relative distance between the electron and hole reflects the spatial correlation. This function, as in other cases, is related to a light-scattering tensor of rank 4, which determines the interband resonance Raman scattering of light.

The distribution function in the absence of a magnetic field has been investigated in Ref. 16 for the $3D$ case and in Refs. 20 and 21 for the quasi-two-dimensional system. In Ref. 18 the distribution function is calculated for the $3D$ case with a strong magnetic field. The quasi-two-dimensional system in a strong magnetic field studied below differs substantially from both the $3D$ system and the quasi-two-dimensional system in the absence of a magnetic field because of the characteristic features of the energy spectrum of the electron–phonon system. Therefore it is helpful to develop a systematic theory that makes it possible to calculate the distribution function of the electron and hole for an arbitrary number of emitted phonons, taking account of the splitting of the terms of the electron–phonon system.

In Sec. 2 a quantum-well model, which is considered as an example of a quasi-two-dimensional system, is formulated, a definition is given for the wave function of electron–hole pairs after the emission of N phonons, and the diagrammatic rules for calculating the wave function are stated. In Secs. 3 and 4 the distribution function of electron–hole pairs before the emission of phonons is calculated and a procedure is presented for renormalizing the electron lines, taking account of the lifting of the degeneracy of the levels of the electron–phonon system. In Sec. 5 the distribution function corresponding to the two-phonon resonance Raman scattering of light is investigated, its relation to the light scattering tensor is established, and the frequency dependence is analyzed. Section 6 is devoted to a discussion of the results.

2. WAVE FUNCTION OF AN ELECTRON–HOLE PAIR AFTER THE EMISSION OF *N LO* PHONONS

The theory is developed for a single square quantum well with infinitely high walls in the case when $d \ll \lambda_e$, λ_h (d is the width of the well, $\lambda_{e(h)}$ is the electron (hole) mean-free path), which gives quantization of electron and hole motion perpendicular to the plane of the well. A magnetic field with intensity H is applied in the same direction. If the well is bounded by the planes $z=0$ and $z=d$, then the wave function of the electron and hole, which do not interact with phonons, has the form, in the effective mass approximation,

$$\Psi_\nu(\mathbf{r},z)=\sqrt{\frac{2}{dL_x}}\,e^{ik_x x}\varphi_n(y)\sin\frac{\pi m z}{d},\quad 0\leqslant z\leqslant d,$$

$$\varphi_n(y)=\frac{1}{\sqrt{2^n n!\sqrt{\pi}R_0}}\exp\left[-\frac{(y-y_0)^2}{2R_0^2}\right]H_n\left[\frac{y-y_0}{R_0}\right],$$

$$\Psi_\nu(\mathbf{r},z)=0,\quad z<0,\quad z>d, \tag{2}$$

where $H_n(x)$ is a Hermite polynomial. The combined index $\nu=m,\ n,\ k_x$; the indices m (size-quantization quantum number) and n refer either to the electron (m_c, n_c) or the hole (m_v, n_v); $\mathbf{r}$ is a two-dimensional vector of the spatial coordinates in the plane of the well (xy plane), where $\mathbf{r}=\mathbf{r}_e$ for the electron and $\mathbf{r}=\mathbf{r}_h$ for the hole; and, L_x is a normalization length. The calculation is performed in the Landau gauge of the vector potential $\mathbf{A}=\mathbf{A}(-yH,0,0)$, $R_0^2=c\hbar/|e|H$, $y_0=-k_x c\hbar/eH$, k_x is the projection of the wave vector on the x axis, and $e<0$ for an electron. The electron and hole energy levels do not depend on k_x and have the form

$$E_e=m_c^2\hbar\omega_e+\left(n_c+\frac{1}{2}\right)\hbar\Omega_e,$$

$$E_h=\hbar\omega_g+m_v^2\hbar\omega_h+\left(n_v+\frac{1}{2}\right)\hbar\Omega_h,$$

$$\omega_{e(h)}=\frac{\hbar\pi^2}{2m_{e(h)}d^2}. \tag{3}$$

At low temperatures the interaction with LO phonons is due to the emission of the phonons by an electron or hole. If the condition (1) holds, then a real phonon is emitted by an electron and a hole emits a phonon in a virtual transition.

A system consisting of an electron–hole pair, arising as a result of the absorption of one photon with frequency ω_l (we have in mind a direct allowed transition), and LO phonons emitted by the electron–hole pair is studied in the stationary state. The wave function of the system in a state when the electron–hole pair has emitted N phonons can be represented in the form

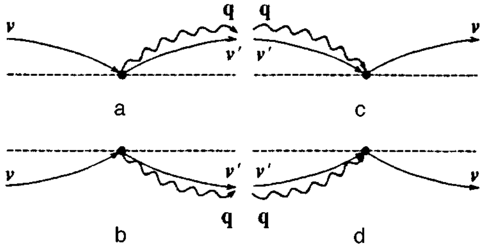


FIG. 2. Various arrangements of the electron, hole, and phonon lines at an electron–phonon interaction vertex.

$$\psi_N(\mathbf{r}_e,z_e,\mathbf{r}_h,z_h,Y_0)=\sum_{\{n_{\mathrm{ph}}\},g}\psi^g_{n_{\mathrm{ph}}}(\mathbf{r}_e,z_e,\mathbf{r}_h,z_h)\psi_{n_{\mathrm{ph}}}(Y_0), \tag{4}$$

where $\psi_{n_{ph}}(Y_0)$ is the wave function of a system of N phonons, corresponding to a state in which the occupation numbers of phonons with wave vectors $\mathbf{q}_1,\ldots,\mathbf{q}_N$ equal 1 and all other occupation numbers equal zero. Here Y_0 is the set of coordinates of the phonon subsystem. The index n_{ph} denotes the collection of vectors $\mathbf{q}_1,\ldots,\mathbf{q}_N$ and

$$\sum_{\{n_{\mathrm{ph}}\}}\ldots=\frac{1}{N!}\sum_{\mathbf{q}_1,\ldots\mathbf{q}_N}\ldots\ . \tag{5}$$

In Eq. (4) $\Sigma_g\psi^g_{n_{\mathrm{ph}}}$ is the wave function of an electron–hole pair after the emission of N phonons with fixed wave vectors and the quantities $\psi^g_{n_{\mathrm{ph}}}$ are calculated by a diagrammatic technique. The rules of this technique are as follows.

1. A wave vector κ_l corresponds to a photon line (dashed line). Nothing is associated with this line; it determines the ratio of the x projections of the wave vectors of the electron and hole lines touching it. It is assumed that $\kappa_l\simeq 0$.

2. Electron lines (solid line) lie above the contour line and hole lines lie below it. The indices $\nu_e=m_c$, n_c, k_{xc} and $\nu_h=m_v$, n_v, and k_{xv}, respectively, are attached to them. These lines are directed from left to right.

3. Phonon lines are wavy. Wave vectors $\mathbf{q}$ are attached to them. The phonon lines are also directed from left to right.

4. The x projections of the wave vectors are conserved at all vertices (light and dark points).

5. The factor

$$M_l=-\left(\frac{2\pi\hbar}{V_0}\right)^{1/2}\frac{|e|}{m_0}(\mathbf{e}_l\cdot\mathbf{p}_{cv})\left(\frac{u_l}{cn_l\omega_l}\right)^{1/2},$$

where V_0 is the normalization volume, $\mathbf{p}_{cv}$ is the interband matrix element of the momentum operator, and $\mathbf{e}_l$, u_l, and n_l are, respectively, the polarization vector of the exciting light, the group velocity, and the refractive index at the frequency ω_l, is associated with a light colored point. The indices m and n of the electron and hole lines, respectively, emanating from a light colored point are equal. The projections k_x have opposite signs.

6. The dark points are electron–phonon and hole–phonon interaction vertices. The arrangement of the lines in the vertices is illustrated in Fig. 2. The factors

$$\pm\mathscr{C}^*e^{\mp i\mathscr{U}}K_{n,n'}(\mp q_x,-q_y);$$

are associated with the vertices a and b and the factors

$$\pm \mathscr{C} e^{\pm i \mathscr{U}} K_{n,n'}(\mp q_x, q_y);$$

are associated with the vertices c and d (the upper signs are taken for a and c). Here

$$\mathscr{C} = \frac{i}{\hbar} C_q M_{m,m'}(q_z), \quad \mathscr{U} = \frac{R_0^2 q_y (2k_x - q_x)}{2}, \tag{6}$$

$$K_{n,n'}(v_x, v_y) = \left[\frac{\min(n!, n'!)}{\max(n!, n'!)}\right]^{1/2} \times e^{-v^2/4} i^{|n-n'|} \left(\frac{v}{\sqrt{2}}\right)^{|n-n'|} \times \exp\left[i\left(\varphi - \frac{\pi}{2}\right)(n-n')\right] L_{\min(n,n')}^{|n-n'|}\left(\frac{v^2}{2}\right),$$

$$\varphi = \arctan\frac{v_y}{v_x}, \quad v^2 = v_x^2 + v_y^2 = R_0^2 (q_x^2 + q_y^2), \tag{7}$$

$L_\beta^\alpha(x)$ are associated Laguerre polynomials;

$$M_{m,m'}(q_z) = \frac{2}{d} \int_0^d dz e^{i q_z z} \sin\frac{\pi m z}{d} \sin\frac{\pi m' z}{d}, \tag{8}$$

$$C_q = -i\hbar\omega_{LO}\left[\frac{4\pi\alpha_0 l_{e(h)}^3}{V_0}\right]^{1/2} \frac{1}{q l_{e(h)}},$$

$$l_{e(h)} = \left(\frac{\hbar}{2 m_{e(h)} \omega_{LO}}\right)^{1/2}, \quad \alpha_0 = \alpha_{e(h)} = \frac{e^2(\kappa_\infty^{-1} - \kappa_0^{-1})}{2\hbar\omega_{LO} l_{e(h)}}, \tag{9}$$

$\varkappa_{0(\infty)}$ is the static (high-frequency) permittivity of the crystal; and, $\mathbf{q} = (q_x, q_y, q_z)$ is the wave vector of the phonon.

7. A factor $(\omega_g - E_j/\hbar + i\delta)^{-1}$, where E_j is the sum of the energies of the line crossing the section, is assigned to each vertical section with number j to the right of a light colored point.

8. The wave function (2) is attached to an electron and hole line with a free end and index ν.

9. Summation is performed over all indices of the electron and hole lines.

In the calculation of the light-scattering tensor and the distribution function in a quasi-two-dimensional system in the absence of a magnetic field, the reciprocal γ of the lifetime, which governs the time dependence of the electronic and hole states, enters in the theory. The quantity γ is determined by the interaction of the electrons and holes with LO phonons, and is proportional to α_0. This system consists of discrete levels and phonons with fixed frequency ω_{LO}, since spatial dispersion and damping of the LO phonons are neglected. In this case the interaction with phonons does not make the states time dependent, but merely lifts the degeneracy of the levels of the electron–phonon system at the point $\Omega_e = \omega_{LO}$. The parameter δ introduced in rule 7 characterizes the time dependence of the states, which is associated with other processes.

Examples of diagrams for calculating the wave function after the emission of two phonons are presented in Fig. 3. The index g denotes the type of diagram. We emphasize that a diagram of type g is characterized by a definite sequence of wave vectors $\mathbf{q}_i$. For example, if $\mathbf{q}_1$ and $\mathbf{q}_2$ are interchanged in diagram a, then the diagram obtained will correspond to a different index g. Therefore each diagram with a fixed arrangement of N phonon wave vectors generates $N!$ diagrams which differ from the initial diagram only by a permutation of the vectors. The sequence in which phonons are emitted by an electron and a hole is also important in the diagrams, and the diagrams b and c therefore correspond to different g.

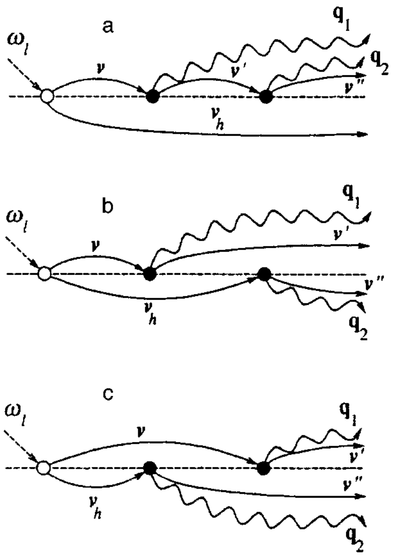


FIG. 3. Diagrams for calculating the wave function after the emission of two LO phonons. $\nu = k_x, m, n$; $\nu' = k_x - q_{1x}, m', n'$; $\nu_h = -k_x, m, n$; $\nu'' = k_x - q_{1x} - q_{2x}, m'', n''$ (a); $\nu'' = -k_x - q_{2x}, m'', n''$ (b,c).

3. WAVE FUNCTION AND DISTRIBUTION FUNCTION OF AN ELECTRON–HOLE PAIR BEFORE PHONON EMISSION

Let us consider a very simple and illustrative example in which the light-produced electron–hole pair has not emitted any phonons. This process corresponds to the diagram shown in Fig. 4a. In accordance with the rules, the wave function of the electron–hole pair is given by

$$\psi_{N=0} = \psi_0 = \frac{M_l}{\hbar} \sum_{k_x, m, n} \frac{1}{\Delta_{m,n}^{m,n} + i\delta} \psi_{\nu_e}(\mathbf{r}_e, z_e) \psi_{\nu_h}(\mathbf{r}_h, z_h). \tag{10}$$

In the quantity

$$\Delta_{m,n}^{m,n} = \omega_l - \omega_g - m^2(\omega_e + \omega_h) - (n + 1/2)(\Omega_e + \Omega_h) \tag{11}$$

the lower and upper indices refer to the hole and the electron, respectively. The selection rules at a light colored point are taken into account in Eq. (10): $k_{xh} = -k_{xe}$, $m_c = m_v = m$, and $n_c = n_v = n$. Using the explicit form (2) of the wave functions ψ_{ν_e} and ψ_{ν_h} and the formula

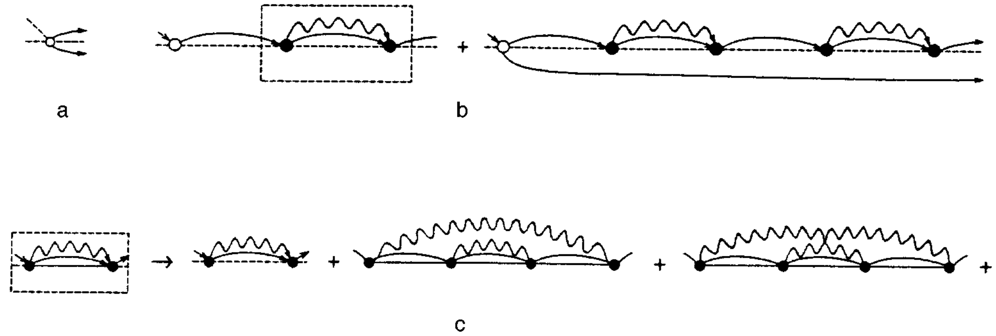


FIG. 4. Diagrams renormalizing an electron line: a—Simplest diagram, b—renormalization for the Landau level $n=1$, c—renormalization for the Landau level $n=2$.

$$\begin{aligned}&\frac{1}{2^n n!\sqrt{\pi}}\sum_{k_x}\exp(ik_x x)\exp\left[-\frac{(y_e-y_0)^2}{2R_0^2}\right]\\&\quad\times\exp\left[-\frac{(y_h-y_0)^2}{2R_0^2}\right]H_n\left(\frac{y_e-y_0}{R_0}\right)H_n\left(\frac{y_h-y_0}{R_0}\right)\\&\quad=\frac{L_x}{2\pi R_0}\exp\left(i\chi_0-\frac{r^2}{4R_0^2}\right)L_n\left(\frac{r^2}{2R_0^2}\right),\end{aligned}\tag{12}$$

we obtain

$$\begin{aligned}\psi_0(\mathbf{r}_e,z_e,\mathbf{r}_h,z_h)=&\frac{M_l}{\pi\hbar d R_0^2}\exp\left(i\chi_0-\frac{r^2}{4R_0^2}\right)\\&\times\sum_{m,n}\sin\left(\frac{\pi m z_e}{d}\right)\sin\left(\frac{\pi m z_h}{d}\right)\\&\times\frac{L_n(r^2/2R_0^2)}{\Delta_{m,n}^{m,n}+i\delta},\end{aligned}\tag{13}$$

where $L_n(x)$ is a Laguerre polynomial and

$$\mathbf{r}=\mathbf{r}_e-\mathbf{r}_h,\quad y_0=R_0^2k_x,\quad \chi_0=\frac{xY}{R_0^2}-\frac{(m_e-m_h)xy}{2R_0^2M},$$

$$\mathbf{R}=\frac{m_e\mathbf{r}_e+m_h\mathbf{r}_h}{M},\quad \mathbf{R}=(X,Y),\quad M=m_e+m_h.$$

The distribution function of the electron and hole over the relative distance between them in the plane of the quantum well after the emission of N phonons is given by[21,18]

$$F_N(\mathbf{r})=\int_0^d dz_e\int_0^d dz_h\int d^2R\int dY_0|\psi_N|^2.\tag{14}$$

Integration over z_e and z_h yields a factor $(d/2)^2\delta_{m,m'}$ ($\delta_{m,m'}$ is the Kronecker delta), and integration over X and Y yields the normalization area S_0. As a result, we obtain in the case $N=0$

$$\begin{aligned}F_0(\mathbf{r})=&\frac{|M_l|^2S_0}{(2\pi\hbar R_0^2)^2}\exp\left(-\frac{r^2}{2R_0^2}\right)\\&\times\sum_m\left|\sum_n\frac{L_n(r^2/2R_0^2)}{\Delta_{m,n}^{m,n}+i\delta}\right|^2.\end{aligned}\tag{15}$$

The physical meaning of the function $F_0(\mathbf{r})$ is as follows: $F_0(\mathbf{r})d^2r$ is the number of pairs, normalized to one photon of the exciting light, which have not emitted a phonon in the interval between $\mathbf{r}$ and $\mathbf{r}+d\mathbf{r}$. One can see from Eqs. (13) and (15) that at the frequency satisfying the condition

$$\Delta_{m,n}^{m,n}=0,\tag{16}$$

the functions ψ_0 and F_0 become infinite if $\delta\to0$. If the frequency ω_l is sufficiently high, the condition (16) can be satisfied for fixed values of m and n. Then one term dominates the multiple sum in Eq. (15), and $F_0(\mathbf{r})\equiv F_{0n}(\mathbf{r})$ assumes the form

$$F_{0n}(\mathbf{r})=\frac{A_n(\mathbf{r})\omega_{LO}^2}{|\Delta_{m,n}^{m,n}+i\delta|^2},\tag{17}$$

where the function

$$A_n(\mathbf{r})=\frac{|M_l|^2S_0}{(2\pi\hbar\omega_{LO}R_0^2)^2}\exp\left(-\frac{r^2}{2R_0^2}\right)\left[L_n\left(\frac{r^2}{2R_0^2}\right)\right]^2\tag{18}$$

determines the coordinate dependence of the distribution function; $A_n(\mathbf{r})$ was introduced in Ref. 18, where the distribution function in a strong magnetic field in the $3D$ case was investigated.

The function $A_n(\mathbf{r})$ does not depend on the quantity number m. Far from the condition (16), many terms make a substantial contribution to the sum—this is the case of the virtual production of an electron–hole pair, when ω_l does not correspond to the frequency of a real transition. One can see from Eq. (17) that F_0 as a function of ω_l consists of one Lorentzian peak, which diverges at the point $\Delta_{m,n}^{m,n}=0$ as δ^{-2}.

To obtain a splitting of the peak, the electronic line in the diagram in Fig. 4a must be renormalized. The hole line is not renormalized, since only the electron can emit phonons in a real transition in the case in which the resonance condition (1) is satisfied. As will be shown below, renormalization depends on the number n of the Landau level that the electron and hole occupy after the production of the electron–hole pair. We therefore examine the renormalization procedure for each n separately, starting with $n=1$. If $n=0$, then renormalization results only in a shift of the peak by a small amount proportional to α_0.

If $n=1$ and the frequency ω_l is chosen so that $\Delta_{m,1}^{m,1}$ is small, then, retaining the resonance term in Eq. (10), we obtain for ψ_0 the expression

$$\psi_0=\frac{M_l}{\hbar}\sum_{k_x}\frac{\psi_{\nu_h}\psi_{\nu_e}}{\Delta_{m,1}^{m,1}+i\delta} \tag{19}$$

($\nu_e=k_x,m,1$ and $\nu_h=-k_x,m,1$), where m assumes a value determined by the condition (16). The electronic line and the corresponding function ψ_{ν_e} are renormalized by summing the series of diagrams shown in Fig. 4b.

This series consists of the diagrams that diverge most strongly in the parameter α_e/δ^2. Taking account of the fact that the energy levels do not depend on k_x and that the x projections of the wave vectors are conserved at all vertices, we obtain the following simple equation for the renormalized functions:

$$\tilde{\psi}_{\nu_e}=\psi_{\nu_e}+\sum_{m',n'}C_{1,n'}^{m,m'}(\Delta_{m,1}^{m',n'}-\omega_{LO}+i\delta)^{-1}\times(\Delta_{m,1}^{m,1}+i\delta)^{-1}\tilde{\psi}_{\nu_e}, \tag{20}$$

where

$$C_{n,n'}^{m,m'}=\hbar^{-2}\sum_{\mathbf{q}_\perp,q_z}|C_q|^2|M_{m,m'}(q_z)|^2|K_{n,n'}(-q_x,-q_y)|^2. \tag{21}$$

The sum in Eq. (20) over m' and n' is dominated by one term corresponding to the small value of the quantity $\Delta_{m,1}^{m',n'}-\omega_{LO}$. The condition (16) in the case $n=1$ determines the frequency of the optical transition for fixed m:

$$\omega_l=\omega_g+m^2(\omega_e+\omega_h)+\frac{3}{2}(\Omega_e+\Omega_h). \tag{22}$$

In order that the quantity $\Delta_{m,1}^{m',n'}-\omega_{LO}$ vanish at this frequency, the condition

$$(m^2-m'^2)\omega_e+(n-n')\Omega_e-\omega_{LO}=0 \tag{23}$$

must be satisfied. Since $\Omega_e=\omega_{LO}$, m and n are integers, and ω_e/Ω_e is arbitrary, the expression (22) holds if $m'=m$, $n'=n-1=0$; in other words, the summation over the quantum numbers m' and n' vanishes in Eq. (20).

Substituting $\tilde{\psi}_{\nu_e}$ into Eq. (19) and summing over k_x using Eq. (12), we obtain for the renormalized wave function of the electron–hole pair the expression

$$\tilde{\psi}_0=\frac{M_l}{\pi\hbar dR_0^2}\exp\left(i\chi_0-\frac{r^2}{4R_0^2}\right)\sin\left(\frac{\pi m z_e}{d}\right)\times\sin\left(\frac{\pi m z_h}{d}\right)L_1\left(\frac{r^2}{2R_0^2}\right)\times\left[\Delta_{m,1}^{m,1}-\frac{\eta\omega_{LO}^2F(1,0)}{\Delta_{m,1}^{m,1}+\Omega_e-\omega_{LO}+i\delta}+i\delta\right]^{-1}. \tag{24}$$

The coefficients $F(n,n-1)$, where $n=1,2,\ldots$, are given by

$$F(n,n-1)=\eta^{-1}\omega_{LO}^{-2}C_{n,n-1}^{m,m},\quad \eta=\frac{\alpha_0}{2}\sqrt{\frac{\Omega_e}{\omega_{LO}}},$$

$$F(n,n-1)=n^{-1}\int_0^\infty du\sqrt{u}\mathscr{F}_m(\beta\sqrt{u})e^{-u}[L_{n-1}^1(u)]^2,$$

$$\beta=\frac{d}{l_H}, \tag{25}$$

$$\mathscr{F}_m(x)=\frac{8}{x^2+4\pi^2m^2}\times\left[\frac{3x}{8}+\frac{\pi^2m^2}{x}-\frac{4\pi^4m^4(1-e^{-x})}{x^2(x^2+4\pi^2m^2)}\right],$$

$$l_H=\frac{R_0}{\sqrt{2}}. \tag{26}$$

Comparing Eq. (24) with the unrenormalized function ψ_0, we see that renormalization changes the resonance denominator in Eq. (13). At resonance, when $\Delta_{m,1}^{m,1}=0$ and $\Omega_e=\omega_{LO}$, this correction is $\sim\eta/\delta$, and it is large if $\eta/\delta\gg1$.

The distribution function corresponding to $\tilde{\psi}_0$ and referring to the Landau level $n=1$ has the form, according to Eq. (14) (the index m is dropped below, if no confusion results),

$$\tilde{F}_{01}(\mathbf{r})=A_1(\mathbf{r})[(\Gamma+\lambda)^2+\delta^2]\Phi_1(\Gamma),$$

$$\Phi_1(\Gamma)=\left[\left(\Gamma+\frac{\lambda}{2}-\sqrt{\frac{\lambda^2}{4}+\eta F(1,0)+\delta^2}\right)^2\left(\Gamma+\frac{\lambda}{2}+\sqrt{\frac{\lambda^2}{4}+\eta F(1,0)+\delta^2}\right)^2+(2\Gamma+\lambda)^2\delta^2\right]^{-1}. \tag{27}$$

In Eq. (27) we introduced the notation

$$\Gamma_n=\Delta_{m,n}^{m,n}/\omega_{LO} \tag{28}$$

(the dimensionless frequency ω_l, measured from the quantity $\omega_g+m^2(\omega_e+\omega_h)+(n+1/2)(\Omega_e+\Omega_h)$) and

$$\lambda=\frac{\Omega_e-\omega_{LO}}{\omega_{LO}} \tag{29}$$

(the dimensionless deviation of the electron cyclotron frequency from the resonance value). The function $\tilde{F}_{01}$ corresponds to the index $n=1$. Therefore

$$\Gamma_{n=1}=\Gamma=\frac{1}{\omega_{LO}}\left[\omega_l-\omega_g-m^2(\omega_e+\omega_h)-\frac{3}{2}(\Omega_e+\Omega_h)\right].$$

It follows from Eq. (27) that $\tilde{F}_{01}$ in the case $\delta=0$ contains two singular peaks, which become infinite at the points $\Gamma=(1/2)(\pm\sqrt{\lambda^2+4\eta F(1,0)}-\lambda)$ and are separated by the distance $\sqrt{\lambda^2+4\eta F(1,0)}$. At resonance $\lambda=0$, and this distance (splitting of the peak) equals $2\sqrt{\eta F(1,0)}$. For $\lambda=0$, in the limit $\eta F(1,0)\gg\delta^2$ there are two peaks; their values at the maxima are $\sim(2\delta)^{-2}$. In the other limiting case, $\eta F(1,0)\ll\delta^2$, there is one peak whose maximum lies at the point $\Gamma=0$. The frequency dependence of the function $\tilde{F}_{01}$ is presented in Fig. 5 (curves *1* and *3*).

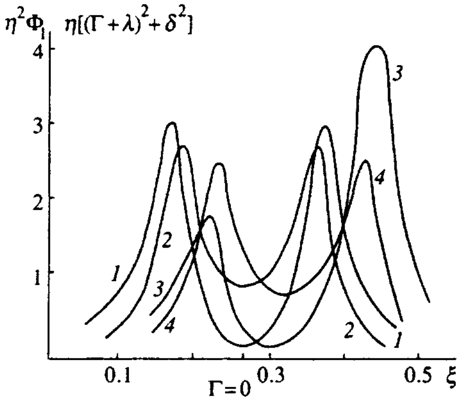


FIG. 5. $\eta[(\Gamma+\lambda)^2+\delta^2]\Phi_1(\Gamma)$ (curves *1,3*) and $\eta^2\Phi_1(\Gamma)$ (Eq. (49), curves *2* and *4*) versus the dimensionless frequency $\xi=1/\omega_{LO}\left[\omega_l-m^2(\omega_e+\omega_h)-3\omega_{LO}/2\right]=\Gamma+\frac{3}{2}\left[\lambda+(1+\lambda)m_e/m_h\right]$, $m_e/m_h=0.176$; $\delta=0.03$; $\eta F(1,0)=0.01$. Curves *1, 2*—$\lambda=0$, curves *3, 4*—$\lambda=0.05$. The curves corresponding to $\lambda=-0.05$ are obtained from the curves *3* and *4* by mirror reflection through the point $\Gamma=0$ ($\xi=0.265$).

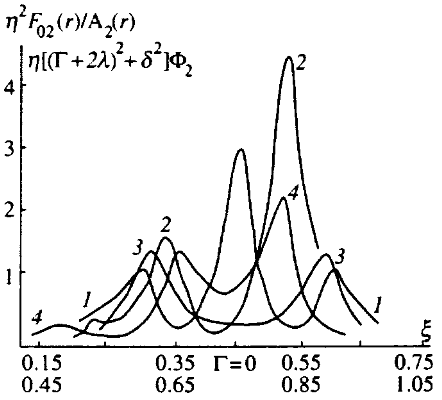


FIG. 6. $\eta\widetilde{F}_{02}(\mathbf{r})/A_2(\mathbf{r})$ (curves *1,2*) and $\eta^2[(\Gamma+2\lambda)^2+\delta^2]\Phi_2$ (curves *3,4*) versus the dimensionless frequency $\xi=\Gamma+5[\lambda+(1+\lambda)m_e/m_h]/2$, $\eta=0.01$, $\delta=0.03$, $m_e/m_h=0.176$; *1, 3*—$\lambda=0$; *2, 4*—$\lambda=0.1$. The top ξ scale is for curves *1* and *3* and the bottom ξ scale is for curves *2* and *4*. $F(1,0)=0.886$; $F(2,1)=0.775$.

4. DISTRIBUTION FUNCTION OF AN ELECTRON–HOLE PAIR FOR LANDAU NUMBERS $n\geqslant 2$

If after absorbing a photon an electron and hole occupy the Landau levels $n_c=n_v=2$, the renormalization of the electronic line performed in Sec. 3 is inadequate, since it neglects all of the most strongly divergent diagrams. Indeed, under resonance conditions, an electron in level $n=2$ can emit two phonons and pass in succession into the levels $n=1$ and $n=0$, while the series of diagrams in Fig. 4a takes account of one resonant transition $n=1\rightarrow n=0$. Renormalization of the line $n=2$ therefore entails replacing each fragment of the diagram in Fig. 4b (one such fragment is enclosed by the dashed rectangle) by the series of diagrams displayed in Fig. 4c. The selection rule for the important diagrams is as follows: if an electron ends up in a Landau level n as a result of pair production, then the diagrams in which any vertical section intersects not more than n phonon lines are important.

As one can see from Fig. 4c, starting with $n=2$, both electron lines and vertex parts (diagrams with intersecting phonon lines) are renormalized. An approximate theory that neglects renormalization of the vertex parts is presented below. It can be shown that neglecting diagrams with intersecting phonon lines does not change the qualitative picture of the splitting of a peak. Only the numerical factors that appear in the expressions that govern the splitting between the peaks change.[22]

Neglecting renormalization of the vertices simplifies the summation of the important diagrams, since the series reduce to a geometric progression. As a result, we obtain for the distribution function near resonance

$$\widetilde{F}_{02}(\mathbf{r})=A_2(\mathbf{r})\left|\Gamma+i\delta-\frac{\eta F(2,1)}{\Gamma+\lambda+i\delta-\eta F(1,0)/(\Gamma+2\lambda+i\delta)}\right|^{-2},$$

$$\Gamma=\Gamma_2, \tag{30}$$

which is a terminating continued fraction with two elements. The frequency dependence of $\widetilde{F}_{02}$ contains three peaks, whose intensities depend on the parameter λ. At resonance $\lambda=0$, and $\widetilde{F}_{02}$ assumes the form

$$\widetilde{F}_{02}(\mathbf{r})=\frac{A_2(\mathbf{r})|(\Gamma+i\delta)^2-\eta F(1,0)|^2}{|\Gamma+i\delta|^2|(\Gamma+i\delta)^2-\eta[F(2,1)+F(1,0)]|^2}, \tag{31}$$

whence it follows that at resonance, peaks occur at the frequencies $\Gamma=0$ and $\Gamma=\pm\sqrt{\eta(F(2,1)+F(1,0))}$. Renormalization of the vertices will change only the number $\sqrt{F(2,1)+F(1,0)}$; the proportionality to the parameter $\sqrt{\eta}$ is preserved, and the position of the central peak does not change.

The frequency dependence of the function (30) is displayed in Fig. 6. Comparing Eqs. (30) and (24) shows that renormalization for the level $n=2$ leads to the appearance of an additional element in the denominator of the continued fraction. The generalization to arbitrary n in the approximate theory proposed is obvious: $\widetilde{F}_{0n}(\mathbf{r})$ is a terminating continued fraction with n elements

$$\widetilde{F}_{0n}(\mathbf{r})=A_n(\mathbf{r})/D_n, \tag{32}$$

$$D_n=\left|\Gamma+i\delta-\frac{\eta F(n,n-1)}{\Gamma+\lambda+i\delta-\eta F(n-1,n-2)/\{\Gamma+2\lambda+i\delta-\cdots-(\eta F(1,0)/(\Gamma+n\lambda+i\delta)\}}\right|^2,\quad \Gamma=\Gamma_n. \tag{33}$$

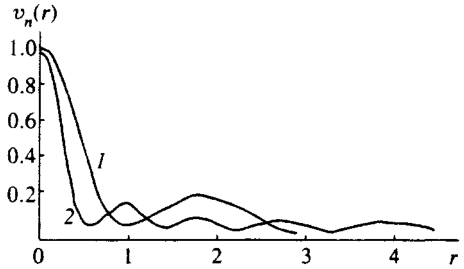


FIG. 7. $v_n(\mathbf{r})=(2\pi R_0^2\hbar\omega_{LO})^2A_n(\mathbf{r})/|M_l|^2S_0$ versus the coordinates $\mathbf{r}$ for different values of the quantum numbers n. Curve *1*—$n=1$, curve *2*—$n=4$.

$\widetilde{F}_{0n}$ as a function of frequency has $n+1$ peaks, which at resonance are split by $\sim\sqrt{\eta}$.

Since the electron–hole pair has not emitted phonons, the wave function (24) can be regarded as a superposition of two two-dimensional clouds—electron and hole—with a common center. As n increases, the distance from the center of the cloud on which an electron and hole can be found with the greatest probability increases. As the clouds increase in size, the probability of electron–hole pairs with large relative distances increases, since the electron and hole in a large cloud can move apart to large distances. Equation (32) shows that the electron–hole distance increases with n—the region next to the central maximum gives an increasingly smaller contribution to the total number of pairs (see below).

The coordinate dependence of the distribution function is shown in Fig. 7. In the limit of large n, circular orbits can be introduced instead of clouds. We note that the function $A_n(\mathbf{R})$ is proportional to the squared radial wave function of a particle in a uniform magnetic field in cylindrical coordinates.[23]

To conclude this section, we calculate the total number $\mathcal{N}_{0n}$ of pairs that have not emitted a phonon,

$$\mathcal{N}_{0n}=\int d\mathbf{r}\widetilde{F}_{0n}(\mathbf{r}), \tag{34}$$

where $\mathcal{N}_{0n}$ is the total number of pairs (per photon of the incident photon flux density) produced in levels with quantum numbers m and n as a result of the absorption of the exciting light, and not emitting any phonons under steady illumination. Substituting the distribution function (32) into the expression (34), we obtain

$$\mathcal{N}_{0n}=\frac{|M_l|^2S_0}{2\pi R_0^2\hbar^2\omega_{LO}^2}D_n^{-1}. \tag{35}$$

The frequency dependence of $\mathcal{N}_{0n}$ is identical to that for $\widetilde{F}_{0n}$. It is also helpful to introduce a distribution function f_{0n}, which is normalized to unity:

$$f_{0n}(\mathbf{r})=\frac{1}{2\pi R_0^2}\exp\left(-\frac{r^2}{2R_0^2}\right)\left[L_n\left(\frac{r^2}{2R_0^2}\right)\right]^2. \tag{36}$$

5. DISTRIBUTION FUNCTION CORRESPONDING TO TWO-PHONON SCATTERING

The distribution function of electron–hole pairs after the emission of two phonons is determined by the wave function (4) with $N=2$. The diagrams for the functions $\psi^g_{n_{\text{ph}}}$ are displayed in Fig. 3. Here the index $g=1$, 2, and 3, and the set $n_{\text{ph}}=\mathbf{q}_1,\ \mathbf{q}_2$. In multiphonon resonant Raman scattering, one transition involving the emission of a phonon is virtual, so diagrams in which a hole emits a virtual phonon are taken into account. It must be borne in mind that two expressions differing by an interchange of the phonon vectors $\mathbf{q}_1$ and $\mathbf{q}_2$ are associated with each diagram, i.e., the index g actually runs through six values.

In accordance with the rules, $\psi^g_{n_{\text{ph}}}$ can be represented in the form

$$\psi^g_{n_{\text{ph}}}=e^{i\chi}\sum_{m,m''}\sum_{n,n''}\sigma^{(g)}K_{\alpha,\beta}(\pm v_x,-v_y)\times\sin\left(\frac{\pi m''z_e}{d}\right)\sin\left(\frac{\pi m z_h}{d}\right). \tag{37}$$

The function $K_{\alpha,\beta}(\pm v_x,v_y)$ is determined by the equations (7)

$$v_x=\frac{y}{R_0}-R_0Q_x,\quad v_y=-\frac{x}{R_0}+R_0Q_y. \tag{38}$$

In $K_{\alpha,\beta}(\pm v_x,-v_y)$ for $g=1$ the plus sign is chosen and $\alpha=n''$, $\beta=n$; for $g=2$ and 3 the minus sign is chosen and $\alpha=n$, $\beta=n''$. The phase χ is the same for all three diagrams shown in Fig. 2:

$$\chi=-Q_xX-Q_yY+\frac{xY}{R_0^2}+\Phi(\mathbf{r}),$$

$$\Phi(\mathbf{r})=\frac{m_e-m_h}{2M}\left(Q_xx+Q_yy-\frac{xy}{R_0^2}\right),$$

$$\mathbf{Q}=\mathbf{q}_{1\perp}+\mathbf{q}_{2\perp},\quad \mathbf{q}_{i\perp}=(q_{ix},q_{iy}). \tag{39}$$

This phase depends on the number N of emitted phonons only through the projection of the sum of the wave vectors of the emitted phonons on the plane of the quantum well $\mathbf{Q}=\Sigma_i^N\mathbf{q}_{i\perp}$, and is present in each diagram with N external phonon lines. The functions

$$\sigma^{(g)}\equiv\sigma^{(g)}_{m'',m,n'',n}=\sigma^{(g)}_{m'',m,n'',n}(\mathbf{q}_1,\mathbf{q}_2)+\sigma^{(g)}_{m'',m,n'',n}(\mathbf{q}_2,\mathbf{q}_1), \tag{40}$$

where

$$\sigma^{(1)}_{m'',m,n'',n}(\mathbf{q}_1,\mathbf{q}_2)=-\frac{M_l C^*_{q_1}C^*_{q_2}}{\pi\hbar^3 R_0^2 d}\exp\left[\frac{iR_0^2}{2}(q_{1x}q_{2y}-q_{2x}q_{1y})\right]$$
$$\times\sum_{m',n'}\frac{M_{m,m'}(-q_{1z})M_{m',m''}(-q_{2z})K_{n,n'}(-q_{1x},-q_{1y})K_{n',n''}(-q_{2x},-q_{2y})}{(\Delta^{m,n}_{m,n}+i\delta)(\Delta^{m',n'}_{m,n}-\omega_{LO}+i\delta)(\Delta^{m'',n''}_{m,n}-2\omega_{LO}+i\delta)}, \tag{41}$$

$$\sigma^{(2)}_{m'',m,n'',n}(\mathbf{q}_1,\mathbf{q}_2)=\frac{M_l C^*_{q_1}C^*_{q_2}}{\pi\hbar^3 R_0^2 d}\exp\left[-\frac{iR_0^2}{2}(q_{1x}q_{2y}-q_{2x}q_{1y})\right]$$
$$\times\sum_{m',n'}\frac{M_{m'm''}(-q_{1z})M_{m',m}(-q_{2z})K_{n',n}(q_{2x},-q_{2y})K_{n',n''}(-q_{1x},-q_{1y})}{(\Delta^{m',n'}_{m',n'}+i\delta)(\Delta^{m'',n''}_{m',n'}-\omega_{LO}+i\delta)(\Delta^{m'',n''}_{m,n}-2\omega_{LO}+i\delta)}. \tag{42}$$

The function $\sigma^{(3)}_{m'',m,n'',n}(\mathbf{q}_1,\mathbf{q}_2)$ differs from Eq. (42) by the fact that $\Delta^{m'',n''}_{m',n'}-\omega_{LO}$ is replaced by $\Delta^{m',n'}_{m,n}-\omega_{LO}$. The derivation of Eqs. (41) and (42) employed the relation

$$\sum_{k_x}\exp[i(x-R_0^2q_y)k_x]\varphi_n(y_h-R_0^2k_x)\varphi_{n'}[y_e-R_0^2(k_x-q_x)]$$
$$=\frac{L_x}{2\pi R_0^2}\exp\left[i\frac{(x-R_0^2q_y)(y_e+y_h+R_0^2q_x)}{2R_0^2}\right]$$
$$\times K_{n',n}(v_x,-v_y).$$

We now establish the correspondence between the distribution function and the light-scattering tensor. As is well known (see, for example, Ref. 16), the light-scattering tensor $S_{\alpha\gamma\beta\lambda}$ is related to the probability $W_{S,N}$ that a photon of the scattered light is emitted per unit time, normalized to one photon of the excited light:

$$\overline{W_{S,N}}=\frac{(2\pi)^3u_lu_s}{V_0c^2n_ln_s}\,\omega_l\omega_s e^*_{s\alpha}e_{l\gamma}e_{s\beta}e^*_{l\lambda}S_{\alpha\gamma\beta\lambda},$$

where the index l refers to the exciting light, the index s refers to the scattered light, and $e_{s(l)\alpha}$ is the α projection of the polarization vector of the light.

On the other hand,

$$\overline{W_{S,N}}=\frac{2\pi}{\hbar}\sum_f|\langle f|M_s|i\rangle|^2\delta(\hbar\omega_l-\hbar\omega_s-N\hbar\omega_{LO}), \tag{43}$$

where M_s is determined by the formula in Sec. 5 by a graphical technique, if the index l in it is replaced by s. The initial state i is determined by the function (4) and the final state is determined by the function

$$\psi_f=\delta(\mathbf{r}_e-\mathbf{r}_h)\delta(z_e-z_h)\psi_{n_{\mathrm{ph}}}(Y_0).$$

It is characterized by the presence of N phonons with a fixed set n_{ph} and photon energy $\hbar\omega_s$. The summation over f in Eq. (43) is equivalent to $\Sigma_{\{n_{\mathrm{ph}}\}}$. In the case $N=2$, using the expressions (37) and (40), we obtain

$$\overline{W_{S,2}}=\frac{2\pi}{\hbar}|M_s|^2S_0\sum_{m,m_1}\sum_{n,n_1}F^{n,n,n_1,n_1}_{2,m,m,m_1,m_1}$$
$$\times\delta(\hbar\omega_l-\hbar\omega_s-N\hbar\omega_{LO}), \tag{44}$$

$$F^{n,n,n_1,n_1}_{2,m,m,m_1,m_1}=\frac{S_0d^2}{2^3}\sum_{q_{1z},q_{2z}}\sum_{\mathbf{q}_\perp}\sum_{g,g'}\sigma^{(g)}_{m,m,n,n}\sigma^{(g')*}_{m_1,m_1,n_1,n_1}. \tag{45}$$

By analogy with Ref. 16, we introduce the distribution function

$$F_N(\mathbf{r},z)=\int d^2R\int dZ\int dY_0|\psi_N|^2,\quad z=z_e-z_h,$$

$$Z=\frac{m_ez_e+m_hz_h}{M},$$

which when integrated over z becomes the function (14). After integration over $\mathbf{R}$, Z, and Y_0, the function $F_N(\mathbf{r},z)$ assumes the form

$$F_N(\mathbf{r},z)=\sum_{m,m',m_1,m_1'}\sum_{n,n',n_1,n_1'}F^{n,n',n_1,n_1'}_{N,m,m',m_1,m_1'}(\mathbf{r})$$
$$\times\xi_{m,m',m_1,m_1'}(z),$$

where

$$\xi_{m,m',m_1,m_1'}(z)=\frac{4}{d^2}\int_{-\infty}^{\infty}dZ\rho_m\left(Z+\frac{m_h}{M}z\right)$$
$$\times\rho_{m'}\left(Z-\frac{m_e}{M}z\right)\rho_{m_1}\left(Z+\frac{m_h}{M}z\right)$$
$$\times\rho_{m_1'}\left(Z-\frac{m_e}{M}z\right),$$

$$\rho_\alpha(x)=\sin\frac{\pi\alpha x}{d}.$$

In turn, $F^{n,n',n_1,n_1'}_{N,m,m',m_1,m_1'}(\mathbf{r})$ can be represented as the sum

$$F^{n,n',n_1,n_1'}_{N,m,m',m_1m_1'}(\mathbf{r})=\sum_{\mathbf{K}}F^{n,n',n_1,n_1'}_{N,m,m',m_1,m_1'}(\mathbf{r},\mathbf{K}).$$

The vector $\mathbf{K}$ is a sum of the electron and hole wave vectors after N phonons are emitted. The vector $\mathbf{K}$ corresponds to the wave vector operator in a magnetic field $\hat{K}_x=-id/dX$, $\hat{K}_y=-id/dY-R_0^2x$. The eigenvalues of the operator are $-Q_x$, $-Q_y$, i.e., $\mathbf{K}=-\mathbf{Q}$.[20] In the case $N=2$

$$F^{n,n',n_1,n_1'}_{2,m,m',m_1,m_1'}(\mathbf{r},\mathbf{K}) = \frac{S_0 d^2}{2^3} \sum_{q_{1z},q_{2z}} \sum_{\mathbf{q}_\perp} \sum_{g,g'} \sigma^{(g)}_{m',m,n',n} \sigma^{(g')*}_{m_1',m_1,n_1',n_1} \times K_{n',n}(\pm v_x,-v_y) K^*_{n_1',n_1}(\pm v_x,-v_y). \tag{46}$$

The expression (46) transforms into Eq. (45), if, first, $\mathbf{r}\to 0$ and $\mathbf{K}\to 0$, i.e., $v_x\to 0$, $v_y\to 0$. This leads to the condition $n'=n$, $n_1'=n_1$. Second, the indices must be the same: $m'=m$, $m_1'=m_1$. Therefore the functions $F^{n,n,n_1,n_1}_{2,m,m,m_1,m_1}(\mathbf{r},\mathbf{K})$ that appear in the expression for the distribution function $F_2(\mathbf{r},z)$ determine the probability $\overline{W_{S,2}}$ (44) and thereby the scattering tensor.

The formal extension of Eqs. (45) and (46) to arbitrary N is simple. The number and form of the functions $\sigma^{(g)}$ will change, and the factor $S_0/N!$ will replace $S_0/2$. The quantities $\sigma^{(g)}$ will consist of a sum of $N!$ terms, which corresponds to the number of permutations of n_{ph}.

We now examine the most interesting case, in which the frequency ω_l corresponds to either input resonance (i.e., when the condition (16) holds) or output resonance (i.e., when $\Delta^{m,n-2}_{m,n}-2\omega_{LO}=0$) and the condition (1) holds. Let $|\Delta^{m,n}_{m,n}|$ (or $|\Delta^{m,n-2}_{m,n}-2\omega_{LO}|$) and $|\Omega_e-\omega_{LO}|\leqslant\sqrt{\eta}\omega_{LO}$, which corresponds to frequencies and magnetic fields in the region of splitting of the peaks. As follows from the form of the functions (41) and (42), the partial function (46), which is diagonal in the Landau numbers, makes the greatest contribution under these conditions.

We represent this function with pairwise equal indices m in the limit $\mathbf{r}=0$ and $\mathbf{K}=0$ in the form of a sum of a part (F^m_n) that is diagonal in m and a part (F^{m,m_1}_n) that is off-diagonal in m. Using the explicit form of the quantities $\sigma^{(g)}$ (41) and (42), we obtain for F^m_n the expression

$$F^m_n = B\eta^2\left[\frac{Pd(n,n-1)}{|\Gamma+i\delta|^2|\Gamma+\lambda+i\delta|^2} + \frac{P^{-1}d(n-1,n-2)(1-\delta_{n,1})}{|\Gamma'+\lambda+i\delta|^2|\Gamma'+2\lambda+i\delta|^2}\right], \tag{47}$$

$$B=\frac{|M_l|^2}{(2R_0\hbar\omega_{LO})^2},\quad P=\left(\frac{\Omega_e+\Omega_h}{\omega_{LO}}\right)^2,$$

$$d(n,n-1)=\frac{1}{n^2}\int_0^\infty du\,u e^{-2u}[L^1_{n-1}(u)]^4\mathscr{F}^2_m(\beta\sqrt{u}),$$

$$n=1,2\ldots,\Gamma=\frac{\Delta^{m,n}_{m,n}}{\omega_{LO}},\quad \Gamma'=\Gamma+2\frac{\Omega_h}{\omega_{LO}}. \tag{48}$$

It can be shown that F^{m,m_1}_n, in contrast to F^m_n, contains the factor

$$\left[\frac{\omega_{LO}}{(m^2-m_1^2)\omega_e}\right]^2+\left[\frac{\omega_{LO}}{(m^2-m_1^2)\omega_h}\right]^2,$$

which is much less than unity if the depth of the quantum well is large compared to $\hbar\omega_{LO}$, since $\omega_{LO}\ll\omega_{e(h)}$. Then it is sufficient to take account of one size-quantization level.

The first term in Eq. (47) corresponds to the channel for direct production of electron–hole pairs (input resonance). It is meaningful starting with the Landau level $n=1$. The second term (direct annihilation or output resonance channel) first contributes at $n=2$, since for $n=1$ the direct annihilation of electron–hole pairs is forbidden by conservation of the quantum numbers n. We recall that Eq. (47) holds in the region Γ, $\Gamma'\sim\sqrt{\eta}$.

Renormalization of the electron lines is performed just as in Secs. 3 and 4. If $n=1$, then it is sufficient to replace $\Gamma+i\delta$ by $\Gamma+i\delta-\eta F(1,0)(\Gamma+\lambda+i\delta)^{-1}$. As a result, we obtain for $\widetilde{F}^m$, instead of Eq. (47),

$$\widetilde{F}^m_1=\eta^2BPd(1,0)\Phi_1(\Gamma). \tag{49}$$

It is clear from Eq. (49) that $\widetilde{F}^m_1$ for $\lambda=0$ has two maxima, located at $\Gamma\pm\sqrt{\eta F(1,0)}$, just like $\widetilde{F}_{01}$ in Sec. 3. As the magnetic field deviates from the resonance value, the ratio of the peak intensities and the position of the peaks change as a result of the change in the contribution of the split terms to the scattering. The maxima of the peaks correspond to two input resonance frequencies

$$\omega_l=\omega_g+m^2(\omega_e+\omega_h)+\frac{3}{2}(\Omega_e+\Omega_h)\pm\sqrt{\eta F(1,0)}. \tag{50}$$

Plots of the function $\widetilde{F}^m_1$ are presented in Fig. 5 (curves *2* and *4*).

In the case $n=2$ both the input and output resonances contribute to the function $\widetilde{F}^m_2$. In the first term of the equation (47) both $\Gamma+i\delta$ and $\Gamma+\lambda+i\delta$ can be renormalized, two elements of the continued fraction being taken into account in the first case and one in the latter case. In the second term only $\Gamma'+\lambda+i\delta$ can be renormalized; it is replaced by $\Gamma'+\lambda+i\delta-\eta F(1,0)(\Gamma'+2\lambda+i\delta)^{-1}$. Substituting $\widetilde{F}^m_2$ in the form $\widetilde{F}^m_2=F_{\mathrm{in}}+F_{\mathrm{out}}$, we obtain for the contributions of the input and output resonances

$$F_{\mathrm{in}}=B\eta^2Pd(2,1)[(\Gamma+2\lambda)^2+\delta^2]\Phi_2(\Gamma), \tag{51}$$

$$F_{\mathrm{out}}=B\eta^2P^{-1}d(1,0)\Phi_1(\Gamma'+\lambda), \tag{52}$$

$$\Phi_2(\Gamma)=\{[(\Gamma-\Gamma^+)^2+\delta^2][(\Gamma-\Gamma^0)^2+\delta^2] \times[(\Gamma-\Gamma^-)^2+\delta^2]\}^{-1}, \tag{53}$$

where $\Gamma^\pm$ and Γ_0 are the roots of the cubic equation

$$\Gamma^3+3\lambda\Gamma^2+[2\lambda^2-\eta(F(1,0)+F(2,1))]\Gamma-2\eta\lambda F(1,0)=0.$$

As in the case $n=1$, F_{out} has two peaks, split by $(\lambda^2+4\eta F(1,0))^{1/2}$. If $\lambda\neq 0$, then F_{in} consists of three peaks, but at exact resonance the central peak vanishes, and two peaks split by $2\sqrt{\eta(F(1,0)+F(2,1))}$ remain. The peaks at the frequencies

$$\omega_l = \omega_g + m^2(\omega_e + \omega_h) + \frac{5}{2}(\Omega_e + \Omega_h) \pm \sqrt{\eta(F(1,0) + F(2,1))} \quad (54)$$

$(\lambda = 0)$ correspond to the input resonance, and the peaks at

$$\omega_l = \omega_g + m^2(\omega_e + \omega_h) + \frac{5}{2}(\Omega_e + \Omega_h) - 2\Omega_h \pm \sqrt{\eta F(1,0)} \quad (55)$$

correspond to the output resonance, i.e., the centers of the two doublets are shifted by $2\Omega_h$. In the heavy-hole approximation $(\Omega_h \to 0)$, the centers of these doublets coincide and we obtain four peaks centered on the point $\Gamma = 0$.[24] Since the coefficients $F(n,n-1) > 0$, in this approximation the inner peaks correspond to the output resonance and the outer peaks correspond to the input resonance. The frequency dependence of the function F_{in} is displayed in Fig. 6 (curves *3* and *4*).

6. DISCUSSION

The theory developed above makes it possible to calculate the wave function of an electron–hole pair that has emitted N LO phonons in a strong magnetic field, and the distribution function of the electron–hole pair over the relative distance between the electron and hole. In a strong magnetic field, the electron and hole energy levels are discrete. This determines the response of the system to the interaction with light: the excitation spectrum (dependence of the total intensity of N-phonon scattering on the exciting light frequency ω_l) consists of a set of narrow peaks, as in one-phonon scattering.[25] Each peak corresponds to a transition of the system through at least one real intermediate state, i.e., a state whose energy equals $\hbar\omega_l$. This is clear, for example, from Eqs. (41) and (42), where the real parts of the expressions in each set of parentheses in the denominator can vanish for certain values of ω_l. For example, if $n = 1$, then $\Delta_{m,1}^{m,1} = 0$ at the frequency

$$\omega_l = \omega_g + m^2(\omega_e + \omega_h) + \frac{3}{2}(\Omega_e + \Omega_h). \quad (56)$$

The other peak lies at

$$\omega_l = \omega_g + m^2(\omega_e + \omega_h) + \frac{\Omega_e}{2} + \frac{3\Omega_h}{2} - \omega_{LO}. \quad (57)$$

If $\Omega_e \neq \omega_{LO}$, then these two peaks are separated by $\Omega_e - \omega_{LO}$. At resonance $\Omega_e = \omega_{LO}$, and the peaks merge, but as a result of the summation of the important diagrams, the peaks split by $\sim \sqrt{\alpha_0}$, and they sharply intensify, since at the maximum they are $\sim \alpha_0/\delta^2$, and off resonance the maximum of a peak is $\sim \alpha_0^2/\delta^2$.

The scattering intensity increases because at resonance, two (and not one) of the three intermediate states of system are real; only the third state is virtual. As a result, it is found that the scattering probability is $\sim \alpha_0$ and not $\sim \alpha_0^2$, as happens in the nonresonance case.

For the input resonance the first two intermediate states are real and the last state is virtual. In the case of the output resonance (Eq. (55)), the first intermediate state is virtual and the second and third are real.

At resonance, the lowest frequency, near which a sharp intensification of two-photon scattering occurs, is determined by Eq. (56). For three-phonon scattering the analogous frequency is

$$\omega_l = \omega_g + m^2(\omega_e + \omega_h) + \frac{5}{2}(\Omega_e + \Omega_h). \quad (58)$$

In three-phonon scattering three of the four intermediate states are real at resonance, as a result of which the intensity of the peaks in terms of the exponent of α_0 remains the same as in two-phonon scattering. This also holds for N-phonon scattering—under resonance conditions the intensity of the peaks in the excitation spectrum is $\sim \alpha_0$.

In the present paper we examined the excitation spectrum for two-phonon scattering at frequencies ranging from the minimum frequency (56) up to the frequency (58). As ω_l increases further, Landau levels with $n > 2$ participate in two-phonon scattering. As n increases, the number of split components of the initial peak increases. For example, if $n = 3$ (four levels of the electron–phonon system cross), then at the input resonance near the frequency

$$\omega_l^{(3)} = \omega_g + m^2(\omega_e + \omega_h) + \frac{7}{2}(\Omega_e + \Omega_h) \quad (59)$$

there will be four peaks in the excitation spectrum. The output resonance at $\omega_l^{(3)} - 2\Omega_h$ is related to the level $n = 2$, i.e., two peaks appear here, since the central peak is absent at resonance.

The theory takes account of the finite mass of a hole. In consequence, the group of peaks referring to the output resonance is displaced by $2\Omega_h$ in the direction of lower frequencies with respect to the group of peaks of the input resonance.

As one can see from Figs. 5 and 6, the intensity of the peaks depends on the magnitude of the magnetic field. This means that the splitting pattern will be sensitive to departures from a parabolic electron band. For sufficiently large departures, any phonon-repetition peak should split into two peaks and the condition (1) will hold for each Landau level.

In the theory it was assumed that the time dependence of a level (homogeneous broadening) is governed by a constant, which appeared by some unspecified mechanism. In addition to homogeneous broadening, inhomogeneous broadening is also present, which is due to both the roughness of the boundaries of the quantum well and the spatial dispersion of the optical phonons. The roughness can probably be neglected, given modern methods for growing heterostructures.

Taking account of the spatial dispersion of a phonon, as shown in Ref. 3, broadens only some of the split peaks, and for weak dispersion its influence is small. Among the mechanisms of homogeneous broadening, elastic scattering must be discarded, since in this case only the quantum number changes (the energy is independent of the latter). Inelastic scattering, which leads to a change in the quantum numbers

m and n, remains. Phonon anharmonicity (decay of a phonon) can also contribute to broadening of the peak.

In conclusion, we briefly discuss the role of excitonic states in the effects examined above. It is known that when the Coulomb interaction of an electron and hole in a strong magnetic field is taken into account, series of excitonic lines arise that converge toward each Landau band from below. In the three-dimensional case,[26–28] only the series below the band $n=0$ corresponds to a discrete spectrum, and all other series are present with a continuous background spectrum. In a quasi-two-dimensional system[29] there is no continuous spectrum and all levels are discrete. As the magnetic field H increases, the distance between the Landau levels increases linearly with H, while the ionization energy $E^{cul}(H)$ of an exciton grows logarithmically.

Excitonic effects can therefore be neglected in sufficiently strong magnetic fields. In the case considered above, the Coulomb energy can be neglected if

$$\hbar\Omega_e=\hbar\omega_{LO}\gg E^{cul}(H).\tag{60}$$

The energy $E^{cul}(H)$ can be represented approximately in the form[28]

$$E^{cul}(H)=\frac{\mu e^4}{2\kappa_0^2\hbar^2}f(H),\quad \mu=\frac{m_e m_h}{m_e+m_h},\tag{61}$$

where $f(H)$ is a weak function of H. The inequality (60) holds for both small and large κ_0. It holds fairly well for III–V materials. Taking account of excitonic states destroys the equal spacing of the levels. This shows up in the emergenceof three points of pairwise intersection of the levels of the electron–phonon system, instead of one point of intersection as in the equidistant case. If the region containing the three points of intersection is small compared to $\sqrt{\eta}\hbar\omega_{LO}$ (the splitting of the levels), then the unequal spacing is unimportant. This condition is expressed by the inequality

$$\sqrt{\eta}\hbar\omega_{LO}\gg|\hbar\omega_{LO}-|E_2-E_1||,\tag{62}$$

where E_i is the energy of the lowest level of the excitonic series belonging to the ith Landau level. In the relation (62) it is assumed that $\hbar\omega_{LO}=|E_1-E_0|$.

This work was supported in part by the Russian Fund for Fundamental Research under Projects Nos. 96-02-17115-a and 95-02-04184-a and the MNTK program ''Physics of Solid-State Nanostructures'' (1-009).

[1] E. J. Johnson and D. M. Larsen, Phys. Rev. Lett. **16**, 655 (1966).

[2] L. I. Korovin and S. T. Pavlov, Zh. Eksp. Teor. Fiz. **53**, 1708 (1967) [Sov. Phys. JETP **26**, 979 (1968)]; JETP Lett. **6**, 50 (1967).

[3] L. I. Korovin, S. T. Pavlov, and B. É. Éshtulatov, Fiz. Tverd. Tela (Leningrad) **20**, 3594 (1978) [Sov. Phys. Solid State **20**, 2077 (1978)].

[4] D. Sarma and O. Madhukar, Phys. Rev. B **22**, 2823 (1980).

[5] A. O. Govorov in *Proceedings of the 13th International Conference on Raman Spectroscopy*, W. Kiefer, M. Cardona, G. Schaack, F. W. Schneider, and H. W. Schrotter (eds.), Wiley, New York (1992), p. 834; J. Raman Spectrosc. **24**, 591 (1993).

[6] A. O. Govorov, Solid State Commun. **92**, 977 (1994).

[7] G. O. Hai, F. M. Peeters, and J. T. Devreese in *Phonons in Semiconductor Nanostructures*, I. P. Leburston, I. Pascual, and C. Sotomayor Torres (eds.), NATO ASI Series, Ser. E: Applied Sciences, Kluwer Academic, Boston (1993), Vol. 236, p. 509.

[8] M. Cardona in *Light Scattering in Solids 2*, M. Cardona and G. Guntherod (eds.), Springer-Verlag, Heidelberg (1982), p. 19.

[9] R. M. Martin and L. M. Falicov in *Light Scattering in Solids 1*, M. Cardona and G. Guntherod (eds.), Springer-Verlag, Heidelberg (1983), p. 79.

[10] T. Ruf and M. Cardona, Phys. Rev. Lett. **63**, 2289 (1989).

[11] R. Merlin, G. Guntherod, and R. Humpery in *Physics of Semiconductors, OIOP Conference Proceedings No. 43*, B. L. Wilson, Institute of Physics and Physics Society, London (1979), p. 852.

[12] M. Yoshida, N. Ohno, M. Mitsutake *et al.*, J. Phys. Soc. Jpn. **54**, 569 (1985).

[13] V. T. Hou, Y. Jin, M. I. Shen *et al.*, Superlattices Microstruct. **12**, 69 (1982).

[14] D. I. Mowbray, M. Cardona, and K. Ploog, Phys. Rev. B **43**, 11815 (1991).

[15] A. Cros, A. Cantarero, and V. I. Belitsky, Phys. Status Solidi B **198**, 653 (1996;).

[16] A. V. Goltsev, Lang, S. T. Pavlov *et al.*, J. Phys. C **63**, 4221 (1983).

[17] L. I. Korovin, S. T. Pavlov, and B. É. Éshpulatov, JETP Lett. **51**, 584 (1990); Zh. Éksp. Teor. Fiz. **99**, 1619 (1991) [Sov. Phys. JETP **72**, 904 (1991)].

[18] V. I. Belitsky, M. Cardona, I. G. Lang, and S. T. Pavlov, Phys. Rev. B **40**, 15767 (1992).

[19] I. G. Lang, S. T. Pavlov, and A. V. Prokhorov, Zh. Éksp. Teor. Fiz. **106**, No. 1, 244 (1994) [JETP **79**, 133 (1994)].

[20] V. I. Belitsky, A. Canterero, S. T. Pavlov, M. Cardona, L. I. Korovin, and I. G. Lang, Phys. Status Solidi B **188**, 863 (1995).

[21] L. I. Korovin, I. G. Lang, and S. T. Pavlov, Zh. Éksp. Teor. Fiz. **108**, 940 (1995) [JETP **81**, 517 (1995)].

[22] L. I. Korovin, Fiz. Tverd. Tela (Leningrad) **13**, 842 (1971) [Sov. Phys. Solid State **13**, 695 (1971)].

[23] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics*, Pergamon Press, New York (1974).

[24] L. I. Korovin, S. T. Pavlov, and B. É. Éshpulatov, Fiz. Tverd. Tela (St. Petersburg) **35**, 1562 (1993) [Phys. Solid State **35**, 788 (1993)].

[25] A. Cros, A. Cantarero, C. Trallero-Giner, and M. Cardona, Phys. Rev. B **46**, 12627 (1992).

[26] R. J. Elliot and R. J. Loudon, J. Phys. Chem. Solids **15**, 196 (1960); **8**, 385 (1959).

[27] H. Hasegawa and R. E. Howard, J. Phys. Chem. Solids **21**, 179 (1962).

[28] A. G. Zhilich and B. S. Monozon, Fiz. Tverd. Tela (Leningrad) **8**, 3559 (1967) [Sov. Phys. Solid State **8**, 2846 (1967)].

[29] M. Altarelli in *Excitons in Confined Systems,* Springer Proc. in Physics **25**, 170 (1988).

Translated by M. E. Alferieff

Ballistic conductance of a quasi-one-dimensional microstructure in a parallel magnetic field

V. A. Geĭler and V. A. Margulis

N. P. Ogaryov Mordovian State University, 430000 Saransk, Russia

(Submitted 12 November 1996)
Zh. Éksp. Teor. Fiz. **111**, 2215–2225 (June 1997)

We discuss the behavior of the ballistic conductance of a quasi-one-dimensional microstructure in a parallel magnetic field when there is electron scattering by a single point impurity inside a channel. An exact analytic formula for the conductance is derived for a model in which the confinement potential is a parabolic well. We show that the conductance curve consists of quantization steps with sharp resonance peaks near the thresholds. Finally, we find the amplitudes and halfwidths of these peaks. © *1997 American Institute of Physics.* [S1063-7761(97)02006-4]

1. INTRODUCTION

Recent theoretical and experimental studies[1–11] have revealed that even a single impurity can have a strong effect on the conductance of a quantum ballistic microstructure. Among the interesting physical effects observed and studied in such microstructures are the quantization of conductance and the destruction of quantization by an impurity.[12–19] Note that a quantizing magnetic field applied to the sample enhances lateral confinement, which changes the electron energy spectrum and, as a result, greatly affects the physical characteristics of the system.[20,21]

Various models have been used in describing theoretically the electron states in the microstructure: an infinitely long waveguide with a constant cross section,[1,2] a saddle point potential for quantum microconstrictions of the bottleneck type,[6–8,22–24] and a parabolic potential for conducting channels, wells, and quantum dots[25–29] in two-dimensional systems.

In the present paper we use a symmetric parabolic potential, $V(x,y)=m^*\omega_0^2(x^2+y^2)/2$, with m^* the effective electron mass in the channel, and ω_0 the characteristic frequency of the confinement potential, to describe the size-confinement potential of a quasi-one-dimensional narrow channel. Such a potential is widely used for the theoretical description of low-dimensional systems. The advantage of using such a potential is that one can derive analytic formulas for the spectral characteristics and scattering parameters.

We consider the situation in which an electron in a narrow conducting channel is scattered by a single impurity at point $\mathbf{r}_0$ inside the channel when the microstructure is placed in a quantizing magnetic field $\mathbf{B}$ directed along the z axis, i.e., along the channel's symmetry axis.

In the model under discussion, one-electron states that are unperturbed by the impurity are described by the Hamiltonian

$$H_0=\frac{1}{2m^*}\left(\mathbf{p}-\frac{e}{c}\,\mathbf{A}\right)^2+\frac{m^*\omega_0^2}{2}\,(x^2+y^2). \tag{1}$$

It is convenient to select the symmetric gauge representation for the vector potential of the magnetic field, i.e.,

$$\mathbf{A}=\tfrac{1}{2}[\mathbf{B}\cdot\mathbf{r}]=-\tfrac{1}{2}(-yB,xB,0).$$

Then in cylindrical coordinates ρ, φ, and z we have

$$H_0=-\frac{\hbar^2}{2m^*}\left[\frac{1}{\rho}\frac{\partial}{\partial\rho}\left(\rho\,\frac{\partial}{\partial\rho}\right)+\frac{1}{\rho^2}\frac{\partial^2}{\partial\varphi^2}+\frac{\partial^2}{\partial z^2}\right]-\frac{i\hbar\omega_c}{2}\,\frac{\partial}{\partial\varphi}+\frac{m^*\Omega^2}{8}\,\rho^2, \tag{2}$$

where

$$\Omega=\sqrt{\omega_c^2+4\omega_0^2}.$$

The spectrum and eigenfunctions of H_0 are

$$E_{mnp}=\frac{\hbar\omega_c}{2}\,m+\frac{\hbar\Omega}{2}\,(2n+|m|+1)+\frac{p^2}{2m^*}, \tag{3}$$

$$\psi^0_{mnp}(\rho,\varphi,z)=\exp\left(\frac{ipz}{\hbar}\right)\,\frac{\exp(im\varphi)}{\sqrt{2\pi}}\,R_{mn}(\rho), \tag{4}$$

where $m\in \mathrm{Z}$ and $n\in \mathrm{N}$. Here

$$R_{mn}(\rho)=c_{mn}\rho^{|m|}\,\exp\left(-\frac{\rho^2}{4l_0^2}\right)L_n^{|m|}\left(\frac{\rho^2}{2l_0}\right),$$
$$c_{mn}=\frac{1}{l_0^{|m|+1}}\,\sqrt{\frac{n!}{2^{|m|}(n+|m|)!}},\quad l_0=\sqrt{\frac{\hbar}{m^*\Omega}}, \tag{5}$$

with $L_n^k(x)$ the generalized Laguerre polynomials.

The kernel of the propagator $\exp(-itH_0/\hbar)$ in the present case is[30]

$$K^0(\mathbf{r},\mathbf{r}';t)=\left(\frac{m^*}{2\pi i\hbar}\right)^{3/2}\frac{\Omega}{2\sqrt{t}\,\sin(\Omega t/2)}\times\exp\left\{\frac{im^*}{2\hbar}\left[\frac{\Omega}{\sin(\Omega t/2)}\left((x'y-xy')\times\sin\frac{\omega_c t}{2}-(xx'+yy')\cos\frac{\omega_c t}{2}+\frac{\rho^2+\rho'^2}{2}\cos\frac{\Omega t}{2}\right)+\frac{(z-z')^2}{t}\right]\right\}. \tag{6}$$

The Green's function of the operator H_0, i.e., the kernel of the resolvent operator $(H_0-E)^{-1}$, is

$$G^0(\mathbf{r},\mathbf{r}';E)=\frac{i}{\hbar}\int_0^\infty K^0(\mathbf{r},\mathbf{r}';t)\exp\left(\frac{itE}{\hbar}\right)dt. \tag{7}$$

The integral in (7) is sure to be convergent after we perform a Wick rotation (i.e., introduce an imaginary time variable through the transformation $t\rightarrow -it$). We now examine the kernel of the semigroup. In view of (6) we can write

$$G^0(\mathbf{r},\mathbf{r}';E)=\frac{1}{\hbar}\int_0^\infty S^0(\mathbf{r},\mathbf{r}';t)\exp\left(\frac{tE}{\hbar}\right)dt, \tag{8}$$

where

$$\begin{aligned}S^0(\mathbf{r},\mathbf{r}';t)=&\left(\frac{m^*}{2\pi\hbar}\right)^{3/2}\frac{\Omega}{2\sqrt{t}\,\sinh(\Omega t/2)}\exp\Biggl\{-\frac{m^*}{2\hbar}\\&\times\Biggl[\frac{\Omega}{\sinh(\Omega t/2)}\Biggl(-i(x'y-xy')\\&\times\sinh\frac{\omega_c t}{2}-(xx'+yy')\cosh\frac{\omega_c t}{2}\\&+\frac{\rho^2+\rho'^2}{2}\cosh\frac{\Omega t}{2}\Biggr)+\frac{(z-z')^2}{t}\Biggr]\Biggr\}.\end{aligned} \tag{9}$$

Equations (8) and (9) clearly show that the integral in (8) is absolutely convergent.

Employing (8), we find that

$$\begin{aligned}&G^0(\mathbf{r},\mathbf{r}';E)\\&=\frac{m^*}{2\pi i\hbar}\sum_{n=0}^{\infty}\sum_{m=-\infty}^{\infty}\exp[im(\varphi-\varphi')]\\&\times\frac{R_{mn}(\rho)R_{mm}(\rho')}{p_{mn}}\exp\frac{ip_{mn}|z-z'|}{\hbar},\end{aligned} \tag{10}$$

where

$$p_{mn}=\left[2m^*\left(E-\frac{\hbar\omega_c}{2}\,m-\frac{\hbar\Omega}{2}\,(2n+|m|+1)\right)\right]^{1/2}. \tag{11}$$

2. GREEN'S FUNCTION AND THE STATE OF A SCATTERED PARTICLE

Now let us suppose that the operator H_0 is perturbed by a short-range impurity modeled by a point potential concentrated at point $\mathbf{r}_0$. Formally such an operator can be written as

$$H=H_0+\lambda\,\delta(\mathbf{r}-\mathbf{r}_0'). \tag{12}$$

As shown, for instance, in Ref. 24, a convenient method for studying operators of this type is to employ an approach based on the Kreĭn formula for resolvent operators.

According to the Kreĭn formula,[31] the Green's function of the operator H has the form

$$\begin{aligned}G(\mathbf{r},\mathbf{r}';E)=&G^0(\mathbf{r},\mathbf{r}';E)-[Q(E;\mathbf{r}_0)+\alpha]^{-1}\\&\times G^0(\mathbf{r},\mathbf{r}_0;E)G^0(\mathbf{r}_0,\mathbf{r}';E).\end{aligned} \tag{13}$$

Here $Q(E,\mathbf{r}_0)$ is the Kreĭn function, which to within a constant is given by

$$Q(E;\mathbf{r}_0)=\lim_{\mathbf{r}\to\mathbf{r}_0}[G^0(\mathbf{r},\mathbf{r}_0;E)-G^0(\mathbf{r},\mathbf{r}_0;0)]. \tag{14}$$

The quantity α is related to the scattering length a by $\alpha=m^*/2\pi\hbar^2 a$. Combining (9) and (10), we find that

$$\begin{aligned}Q(E;\mathbf{r}_0)=&\left(\frac{m^*}{2\pi\hbar}\right)^{3/2}\frac{\Omega}{2\hbar}\int_0^\infty\frac{\exp(Et/\hbar)-1}{\sqrt{t}\,\sinh(\Omega t/2)}\\&\times\exp\Biggl\{-\frac{m^*\Omega\rho_0^2}{2\hbar\sinh(\Omega t/2)}\Biggl[\cosh\frac{\Omega t}{2}\\&-\cosh\frac{\omega_c t}{2}\Biggr]\Biggr\}dt+C,\end{aligned} \tag{15}$$

where $C=(m^*2\pi\hbar^2)(m^*\Omega/2\hbar)^{1/2}\zeta(1/2,1/2)$. In particular,

$$Q(E)\equiv Q(E;\mathbf{0})=\frac{m^*}{2\pi\hbar^2}\sqrt{\frac{m^*\Omega}{2\hbar}}\,\zeta\left(\frac{1}{2};\,\frac{1}{2}-\frac{E}{\hbar\Omega}\right), \tag{16}$$

where $\zeta(s,v)$ is the generalized Riemann zeta function, and ρ_0 is the distance from the point impurity to the symmetry axis of the system.

We now study the asymptotic behavior of Kreĭn's function Q in two limiting cases: large $(\rho_0\gg l_0)$ and small $(\rho_0\ll l_0)$ distances from the impurity to the z axis. We start with the case where $\rho_0\gg l_0$ and partition the integration range in (15) into two ranges: from zero to t_0, where $t_0\ll l_0^2/\rho_0^2\Omega$, and from t_0 to ∞; the respective integrals are $J_1(\rho_0)$ and $J_2(\rho_0)$.

Since

$$\left(\cosh\frac{\Omega t}{2}-\cosh\frac{\omega_c t}{2}\right)\left(\sinh\frac{\Omega t}{2}\right)^{-1}\sim O(1),$$

we have $J_2(\rho_0)\sim O(\exp(-\rho_0^2/l_0^2))$. Replacing the integrand in $J_1(\rho_0)$ with its asymptotic expression, we get

$$\begin{aligned}J_1(\rho_0)\simeq&\frac{1}{2}\left(\frac{m^*}{2\pi\hbar}\right)^{3/2}\frac{E\sqrt{t_0}}{\hbar}\,\mathrm{B}\left(\frac{1}{2},1\right)\\&\times\Phi\left(\frac{1}{2},\,\frac{3}{2};-\frac{m^*\rho_0^2t_0}{8\hbar}\,(\Omega^2-\omega_c^2)\right),\end{aligned} \tag{17}$$

where $\mathrm{B}(x,y)$ is Euler's beta function, and $\Phi(a,b;x)$ is a confluent hypergeometric function. Estimating $\Phi(a,b;x)$ for small x, we obtain

$$J_1(\rho_0)\simeq\frac{2E}{\rho_0}\sqrt{\frac{2}{\pi m^*\hbar(\Omega^2-\omega_c^2)}}.$$

The final expression for $Q(E,\mathbf{r}_0)$ for $\rho_0\gg l_0$ is

$$Q(E;\mathbf{r}_0)=\frac{E}{\rho_0}\frac{m^*}{\pi^2\hbar^3\sqrt{\Omega^2-\omega_c^2}}\left[1+O\left(\frac{1}{\rho_0^2}\right)\right]. \tag{18}$$

Now we turn to the case where $\rho_0\ll l_0$. We have

$$\exp\left\{-\frac{m^*\Omega\rho_0^2}{2\hbar\sinh(\Omega t/2)}\left(\cosh\frac{\Omega t}{2}-\cosh\frac{\omega_c t}{2}\right)\right\}$$
$$\simeq 1-\frac{m^*\Omega\rho_0^2}{2\hbar\sinh(\Omega t/2)}\left[\cosh\frac{\Omega t}{2}-\cosh\frac{\omega_c t}{2}\right].$$

This yields

$$\begin{aligned}
&Q(E;\mathbf{r}_0)-Q(E)\\
&=-\left(\frac{m^*}{2\pi\hbar}\right)^{3/2}\frac{m^*\Omega^2(\Omega^2-\omega_c^2)}{32\hbar^2}\\
&\times\rho_0^2\left[2\int_0^\infty\frac{t^{3/2}\exp(tE/\hbar)}{\cosh(\Omega t)-1}\,dt-3\sqrt{\pi}\zeta\left(\frac{3}{2}\right)\right)\ \Omega^{-5/2}\\
&=-\left(\frac{m^*}{2\pi\hbar}\right)^{3/2}\frac{m^*\Omega^2(\Omega^2-\omega_c^2)}{32\hbar^2}\\
&\times\rho_0^2\left[2\int_0^\infty\frac{t^{3/2}[\exp(tE/\hbar)-\exp(-\Omega t)]}{\cosh\Omega t-1}dt\right.\\
&\left.-3\sqrt{\pi}\zeta\left(\frac{5}{2}\right)\Omega^{-5/2}\right],
\end{aligned}\tag{19}$$

where $\zeta(x)$ is the Riemann zeta function.

Thus, for $\rho_0\ll l_0$ we have $Q(E,\mathbf{r}_0)-Q(E)\sim O(\rho_0^2)$.

We now assume that $\psi_0(\mathbf{r})$ is a delocalized state of the operator H_0. According to (13), the state $\phi(\mathbf{r})$ of the operator H corresponding to $\psi_0(\mathbf{r})$ is the one determined by

$$\psi(\mathbf{r})=\psi_0(\mathbf{r})-[Q(E,\mathbf{r}_0)+\alpha]^{-1}\psi(0)G^0(\mathbf{r},\mathbf{0};E).\tag{20}$$

We note, in particular, that Eqs. (5) and (20) imply that there is scattering in channels with $m\neq 0$. Equations (5), (20), and (18) also imply that asymptotically ($\rho_0/l_0\gg 1$) the function $\psi(\mathbf{r})$ behaves like $(l_0/\rho_0)\exp(-\rho^2/l_0^2)$, so that it rapidly decreases as the distance to the channel axis grows.

Thus, scattering is significant when the impurity is near the channel's symmetry axis, i.e., when $\rho_0/l_0\ll 1$. In this case the estimate (19) shows that the dependence of ρ_0 is weak and can be ignored. Note that the case where $\rho_0=0$ is a problem in its own right because it corresponds to a pinch of the microstructure.[7] It is precisely this case that we describe below.

3. PARTIAL TRANSMISSION COEFFICIENTS FOR A MICROSTRUCTURE

When the impurity is on the channel axis, there exists a simple analytic expression for Kreĭn's function Q (16), which means that the exact eigenfunctions of the Hamiltonian H can be written as

$$\begin{aligned}
\psi(\rho,\varphi,z)=&\frac{R_{0n_0}(\rho)}{\sqrt{2\pi}}\exp\left(\frac{ipz}{\hbar}\right)-\frac{R_{0n_0}(0)}{Q(E)+\alpha\sqrt{2\pi}}\frac{m^*}{2\pi i\hbar}\\
&\times\sum_{n=0}^{\infty}R_{0n}(0)R_{0n}(\rho)\\
&\times\frac{\exp(ip_{0n}|z|/\hbar)}{\{2m^*[E-\hbar\Omega(n+1/2)]\}^{1/2}}.
\end{aligned}\tag{21}$$

The pole of the Green's function $G(\mathbf{r},\mathbf{r}';E)$ corresponds in this case to the condition that $Q(E)=-\alpha$. Clearly, there is only one bound state, whose energy E_c is below the edge of the continuous spectrum of the Hamiltonian H_0, i.e., below $\hbar\Omega/2$. We introduce the notation $E_c=\hbar\Omega(\delta-1/2)$. Using the Hermite formula for the generalized Riemann zeta function, we have[32]

$$\begin{aligned}
\zeta\left(\frac{1}{2},1-\delta\right)\simeq&\frac{1}{\sqrt{1-\delta}}+\frac{1}{2}\frac{1}{\sqrt{2-\delta}}-2\sqrt{2-\delta}\\
&+\frac{1}{24}\frac{1}{(2-\delta)^{3/2}}.
\end{aligned}\tag{22}$$

Then, to estimate the energy of the bound state, we have

$$\frac{1}{\sqrt{1-\delta}}+\frac{1}{2}\frac{1}{\sqrt{2-\delta}}-2\sqrt{2-\delta}+\frac{1}{24}\frac{1}{(2-\delta)^{3/2}}\simeq\frac{\sqrt{2}l_0}{a}.\tag{23}$$

Under ordinary conditions $|a|\sim 10^{-7}$ cm and $l_0\sim 10^{-6}$ cm, only the first term makes a substantial contribution (23), with the result that for E_c we have the simple estimate

$$E_c\simeq\frac{\hbar\Omega}{2}-\frac{m^*\Omega^2a^2}{2}.\tag{24}$$

The E_c vs. B curve, obtained by numerically solving Eq. (23), is depicted in Fig. 1.

Since only delocalized states participate in scattering, below we examine a scattered particle with an energy $E>\hbar\Omega/2$.

As noted earlier, there is no scattering in which the projection of the particle's orbital angular momentum on the channel's symmetry axis is finite, so that we consider only transitions from the mode $(n_0,0)$ to the mode $(n_1,0)$. We denote the corresponding transmission coefficient by $T_{n_0\to n_1}(E)$.

Suppose that the wave that propagates in the mode $(n_0,0)$ with energy $E=\hbar\Omega(n_0+1/2)+p^2/2m^*$ is

$$\psi_{0n_0p}(\mathbf{r})=\frac{1}{2\pi^2\hbar}\exp\left(\frac{ipz}{\hbar}\right)R_{0n_0}(\rho),\quad \rho>0.$$

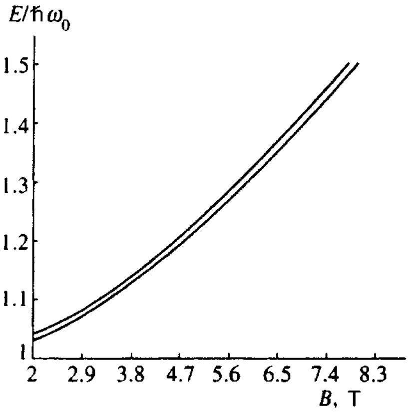


FIG. 1. Bound-state energy vs. magnetic field strength (lower curve). For the sake of comparison we depict the dependence of the edge of the continuous spectrum on the magnetic field strength (upper curve).

The exact eigenfunctions (21), which are solutions of the equation $H\psi=E\psi$, correspond to this wave if the energy E is the same.

In order to obtain the partial transmission coefficient $T_{n_0\to n_1}(E)$, we isolate the coefficient of $R_{n_0 0}(\rho)$ for $z<0$ and the coefficient of $R_{n_1 0}(\rho)$ for $z>0$.

For $z<0$ Eq. (21) yields

$$\psi_{n_0}^{(-)}(\rho,\varphi,z)=\exp\left(\frac{ipz}{\hbar}\right)-\frac{|R_{0n_0}(0)|^2}{Q(E)+\alpha}\,\frac{m^*}{2\pi i\hbar p}\times\exp\left(-\frac{ipz}{\hbar}\right), \tag{25}$$

and for $z>0$ we have

$$\psi_{n_1}^{(+)}(\rho,\varphi,z)=\delta_{n_0n_1}\exp\left(\frac{ipz}{\hbar}\right)-\frac{R_{0n_0}(0)R_{0n_1}(0)}{Q(E)+\alpha}\,\frac{m^*}{2\pi i\hbar}\times\frac{\exp[ip_{0n_1}z/\hbar]}{[p^2-2m^*\hbar\Omega(n_0-n_1+1/2)]^{1/2}}. \tag{26}$$

We allow for the fact that $R_{n0}(0)=1/l_0$ for all values of n. Then, by comparing (25) and (26) we find that

$$T_{n_0\to n_1}(E)=\left|\delta_{n_0n_1}-\frac{m^*}{2\pi i\hbar l_0^2\,[p^2+2m^*\hbar\Omega(n_0-n_1+1/2)]^{1/2}[Q(E)+\alpha]}\right|^2. \tag{27}$$

We now introduce into (27) the notation $E=\hbar\Omega(N+\delta+\frac{1}{2})$, where $N\in\mathrm{N}$ and $0<\delta<1$. It is convenient at this point to transform (27) by separating the real and imaginary parts of the generalized Riemann zeta function. Using the shift formula we obtain[32]

$$\zeta\left(\frac{1}{2};N-\delta\right)=\zeta\left(\frac{1}{2};1-\delta\right)+i\sum_{n=1}^{N}(N+\delta-n)^{-1/2}. \tag{28}$$

As the estimate (22) implies, $\mathrm{Re}\,\zeta(\frac{1}{2};N-\delta)=\zeta(\frac{1}{2};1-\delta)$, while the corresponding finite sum in (28) is $\mathrm{Im}\,\zeta$. Let us introduce the dimensionless pseudopotential coupling constant $\gamma=|a|/\sqrt{2}l_0$. As noted earlier, in real situations for narrow channels $\gamma\ll 1$. Using this notation, we can transform (28) into

$$T_{n_0\to n_1}=\delta_{n_0n_1}\left\{1+\frac{2\gamma^2(N+\delta-n_1)^{-1/2}\mathrm{Im}\,\zeta}{(1+\gamma\,\mathrm{Re}\,\zeta)^2+(\gamma\,\mathrm{Im}\,\zeta)^2}\right\}+\frac{\gamma^2(N+\delta-n_1)^{-1}}{(1+\gamma\,\mathrm{Re}\,\zeta)^2+(\gamma\,\mathrm{Im}\,\zeta)^2}. \tag{29}$$

We see that the partial amplitudes are finite for all values of N and δ, i.e., for all values of E. Moreover, the last two terms on the right-hand side of Eq. (29), containing powers of the small parameter γ, are much smaller than the first term.

4. CONDUCTANCE OF THE MICROSTRUCTURE

Following the Landauer–Buttiker formalism,[6,22,23] we can find the conductance of a one-dimensional channel by the following formula:

$$\frac{G(E)}{2e^2/h}=\sum_{n_0,n_1=0}^{N}T_{n_0\to n_1}(E). \tag{30}$$

Plugging (29) into (30) and calculating the sum over n_0, we get

$$\frac{G}{2e^2/h}=\sum_{n_1=0}^{N}\left\{1+\frac{2\gamma^2(N+\delta-n_1)^{-1/2}\,\mathrm{Im}\,\zeta}{(1+\gamma\,\mathrm{Re}\,\zeta)^2+(\gamma\,\mathrm{Im}\,\zeta)^2}\right\}+\frac{\gamma^2(N+1)}{(1+\gamma\,\mathrm{Re}\,\zeta)^2+(\gamma\,\mathrm{Im}\,\zeta)^2}\sum_{n_1=0}^{N}\frac{1}{N+\delta-n_1}. \tag{31}$$

We use the following formulas to calculate the sums in (31):

$$\sum_{n_1=0}^{N}(N+\delta-n)^{-1/2}=\mathrm{Im}\,\zeta,$$

$$\sum_{n_1=0}^{N}\frac{1}{N+\delta-n_1}=\Psi(N+\delta)-\Psi(\delta)+\frac{1}{N+\delta}.$$

Here Ψ is the logarithmic derivative of the Γ-function. We then find that

$$\frac{G}{2e^2/h}=N+1+\gamma^2\times\frac{2(\mathrm{Im}\,\zeta)^2+(N+1)[\Psi(N+\delta)-\Psi(\delta)+(N+\delta)^{-1}]}{(1+\gamma\,\mathrm{Re}\,\zeta)^2+(\gamma\,\mathrm{Im}\,\zeta)^2}. \tag{32}$$

If in (32) we go back to the variable E, we finally obtain

$$\frac{G(E)}{2e^2/h}=1+[x]+\gamma^2(1+2\gamma\zeta(\tfrac{1}{2};1-\{x\})+\gamma^2|\zeta(\tfrac{1}{2};-x)|^2)^{-1}\Bigg(\left|\zeta(\tfrac{1}{2};-x)-\zeta(\tfrac{1}{2};1-\{x\})\right|^2+(1+[x])\bigg(\psi(x)-\psi\bigg(\{x\}+\frac{1}{x}\bigg)\bigg)\Bigg), \tag{33}$$

where $x=E/\hbar\Omega-\frac{1}{2}$, $[x]$ denotes the integer part of x, and $\{x\}$ the fractional part.

The channel conductance $G(E)$ consists of two terms. The first term, $G_1=[(2E+\hbar\Omega)/2]$ is responsible for the steps in the conductance, which depend on the magnetic field B in a nonlinear manner; the step width is $\hbar\Omega=\hbar\sqrt{\omega_c^2+4\omega_0^2}$. The height of these steps is equal to unity (in units of $2e^2/h$). The second term G_2 contains a small factor $\gamma^2\ll 1$ and determines the deviations of conductance quantization steps from the ideal shape. It is obvious that scattering of electrons by a point impurity produces this term.

The diagrams of the $G(E)$ and $G(B)$ dependences built according to formula (33) are depicted in Fig. 2. Clearly visible are conductance quantization steps between the reso-

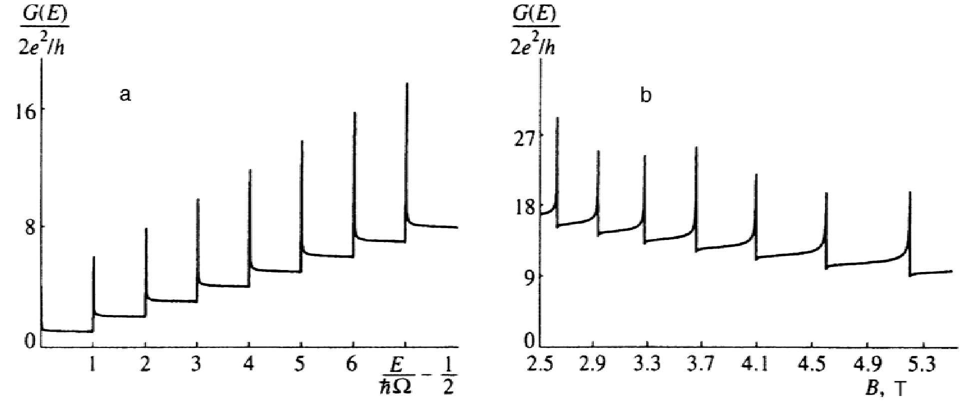


FIG. 2. Conductance vs. scattered-particle energy at $B=5$ T and $\omega_0=2\times10^{13}\ \mathrm{s}^{-1}$ (a) and vs. magnetic field strength at $E=10^{-12}$ erg and $\omega=2\times10^{13}\ \mathrm{s}^{-1}$ (b).

nance peaks at the threshold of each step. The resonance peaks are due to the behavior of the second term in the general formula (33) near a step threshold. The detailed behavior of the resonance curve near a resonance point is depicted in Fig. 3.

Let us analyze formula (33). With the exception of a small neighborhood of the energy values at which $\delta\ll1$, the second term in this formula is of order $\gamma^2E/\hbar\Omega$. Hence for realistic values of E such that $E/\hbar\Omega\ll\gamma^{-2}$, this term is of order $O(\gamma^2)$. Thus, the second term in (33) is very small compared to the first term for all energy values except in the neighborhood of the step threshold. But at values of E in such neighborhoods, the situation changes. Let $\delta_0=\{(2E-\hbar\Omega)/2\hbar\Omega\}$ satisfy the condition

$$-\gamma\ \operatorname{Re}\ \zeta\left(\frac{1}{2};\ \frac{1}{2}-\frac{E}{\hbar\Omega}\right)=1. \tag{34}$$

Using formula (28), we can write this condition as follows:

$$-\gamma\zeta(\tfrac{1}{2};1-\delta_0)=1. \tag{35}$$

At $\delta=\delta_0\ll1$ the small factor γ^2 in $G_2(E)$ disappears, and at such energies the value of G_2 can be estimated. Allowing for the fact that both the numerator and denominator in $G_2(E)$ are proportional to δ_0^{-1}, we can easily arrive at an estimate at the resonance points: $G_2(E)\sim3+2E/\hbar\Omega$.

Hence $G_1(E)$ and $G_2(E)$ at the resonance points are of the same order. Note that the resonance condition (34) corresponds to the existence of a bound state: $Q(E)=-\alpha$. This clarifies the physical nature of a resonance: resonance peaks are generated by bound states.

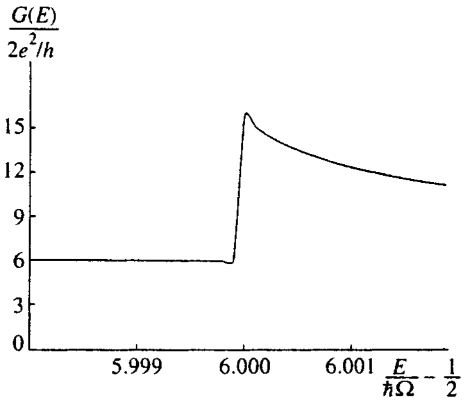


FIG. 3. Structure of the G vs. E resonant curve near the resonance at $B=5$ T and $\omega_0=2\times10^{13}\ \mathrm{s}^{-1}$.

Now let us study the resonance peaks in G vs. E. To the left of the threshold of each step, where δ approaches unity, $G_2(E)$ is of order $G_2\sim(1-\delta)(N+3)$. Hence the value of G_2 to the left of a resonance is small, and the resonance peaks originate at the point where $\delta=0$. Since δ_0 is small, the ascending section of a resonance peak is almost vertical. The descending section of the curve near a resonance to the right of the peak threshold after the point where $\delta=\delta_0$ can be approximately described by a formula that follows from (32) with $\delta\ll1$:

$$\frac{G(E)}{2e^2/h}\simeq N+1+\frac{N+3}{1+\delta\gamma^{-2}}. \tag{36}$$

Here we have allowed for the fact that $\gamma\zeta(\frac{1}{2};1-\delta)\ll1$ for $\delta\ll1$.

Now let us find the halfwidth $\Gamma(E)$ and amplitude of the resonance peaks. The above implies that the resonance maxima are shifted slightly to the right from the threshold of each step. Then the peak amplitudes ΔG can easily be estimated as follows:

$$\frac{\Delta G}{2e^2/h}=\frac{E}{\hbar\Omega}+\frac{3}{2}. \tag{37}$$

We see that the peak amplitudes increase linearly with E. To estimate the halfwidths $\Gamma(E)$ of the peaks we recall that the descending section of the resonance curve is not so steep as the ascending. Let us find a value $\delta=\delta_1$ at which $G(\delta_1)=G(\delta_0)/2$. The condition (35) leads to the following estimate: $\delta_1\simeq\gamma^2(N+2)$. Then for the halfwidth of a peak we have $\Gamma(E)\simeq\gamma^2(E+3\hbar\Omega/2)$. Since for all realistic values $\gamma^2(E+3/2\hbar\Omega)\ll\hbar\Omega$, the halfwidth is small compared to the step width. Note that the function $\Gamma(E)$ increases linearly with E. The behavior of G vs. E discussed above agrees with the corresponding curve in Fig. 2.

The phenomenon of conductance quantization and the resonance peaks near the step thresholds also exist in the limit of zero magnetic field. Here Ω must be replaced by $2\omega_0$ in all formulas. As noted earlier, the conductance step width increases with the magnetic field (Fig. 2). Hence by measuring the step widths, the separation of the resonance peaks, and the peak amplitudes we can determine the frequency ω_0 of the confinement potential and the cyclotron effective mass of the charge carriers.

How do displacements of the impurity from the channel axis affect the conductance? As noted in Sec. 2, small deviations in the position of a point scattering center from the channel axis yield a small correction to Kreĭn's function Q, $\sim(\rho_0/l_0)^2$, and hence to the scattering amplitude. Consequently, this correction can affect only the last term on the right-hand side of Eq. (33) for the conductance, the term due to scattering, and, as noted earlier, it is small far from resonance. In view of this we can state that the step shape remains essentially unchanged when the impurity moves away from the axis of a narrow channel.

The present work was supported by the Russian Fund for Fundamental Research (Grant No. 95-02-04871 and the Universities of Russia Program.

[1] C. S. Chu and R. S. Sorbello, Phys. Rev. B **40**, 5941 (1989).
[2] P. F. Bagwell, Phys. Rev. B **41**, 10354 (1990).
[3] P. F. Bagwell, J. Phys., Condens. Matter. **2**, 6179 (1990).
[4] A. Kumar and P. F. Bagwell, Phys. Rev. B **44**, 1747 (1991).
[5] Y. Takagaki and D. K. Ferry, Phys. Rev. B **45**, 6718 (1992).
[6] Y. B. Levinson, M. I. Lubin, and E. V. Sukhorukov, Phys. Rev. B **45**, 11976 (1992).
[7] Y. B. Levinson, M. I. Lubin, and E. V. Sukhorukov, JETP Lett. **54**, 401 (1991).
[8] M. I. Lubin, JETP Lett. **57**, 361 (1993).
[9] M. Buttiker, in *Semiconductors and Semimetals*, M. Reed (ed.), Academic Press, New York (1992).
[10] J. Faist, P. Guerert, and H. Rothuizen, Phys. Rev. B **42**, 3217 (1986).
[11] D. H. Gobden, N. K. Patel, M. Pepper, D. A. Ritchie, J. E. F. Frost, and G. A. C. Jones, Phys. Rev. B **42**, 1938 (1991).
[12] A. B. Fowler, G. L. Timp, J. J. Wainer, and R. A. Webb, Phys. Rev. Lett. **57**, 138 (1986).
[13] T. E. Kopley, P. L. McEuen, and R. G. Wheller, Phys. Rev. Lett. **61**, 1654 (1988).
[14] P. L. McEuen, B. W. Alphenaur, R. G. Wheller, and R. N. Sacks, Surf. Sci. **312**, 229 (1990).
[15] S. J. Bending and M. R. Beasley, Phys. Rev. Lett. **55**, 324 (1985).
[16] M. Naito and M. R. Beasley, Phys. Rev. B **42**, 1492 (1990).
[17] M. W. Dellow, P. H. Beton, C. J. G. M. Landerak, T. J. Foster, P. C. Main, L. Eaves, M. Henini, S. P. Beaumont, and C. D. W. Wilkinson, Phys. Rev. Lett. **68**, 1754 (1992).
[18] A. K. Geim, P. C. Main, N. La Scala , Jr., L. Eaves, T. J. Foster, P. H. Foster, P. H. Beton, J. W. Sakai, F. W. Sheard, M. Henini, G. Hill, and M. A. Pate, Phys. Rev. Lett. **72**, 2061 (1994).
[19] A. K. Geim, T. J. Foster, A. Nogaret, N. Mori, P. J. McDonnell, N. La Scala , Jr., P. C. Main, and L. Eaves, Phys. Rev. B **50**, 8074 (1994).
[20] J. K. Jain and S. A. Kivelson, Phys. Rev. Lett. **60**, 1542 (1988).
[21] M. Y. Azbel, Phys. Rev. B **43**, 2435 (1991).
[22] H. A. Fertig and B. I. Halperin, Phys. Rev. B **36**, 7969 (1987).
[23] M. Buttiker, Phys. Rev. B **40**, 7906 (1990).
[24] V. A. Geĭler, V. A. Margulis, and I. I. Chuchaev, JETP Lett. **58**, 648 (1993).
[25] R. Merlin, Solid State Commun. **64**, 99 (1987).
[26] V. A. Geĭler, V. A. Margulis, and O. B. Tomilin, JETP Lett. **63**, 578 (1996).
[27] V. A. Geĭler, V. A. Margulis, and I. V. Chudaev, Zh. Éksp. Teor. Fiz. **109**, 762 (1996) [JETP **82**, 409 (1996)].
[28] H. Tamura and T. Ando, Phys. Rev. B **44**, 1792 (1991).
[29] P. Streda, J. Kucera, and A. H. MacDonald, Phys. Rev. Lett. **59**, 1973 (1987).
[30] T. J. Popadopoulos, J. Phys. A **4**, 773 (1971).
[31] B. S. Pavlov, Usp. Mat. Nauk **42**, No. 6, 99 (1987).
[32] V. A. Geĭler, V. A. Margulis, I. V. Chudaev, and I. I. Chuchaev, Zh. Éksp. Teor. Fiz. **107**, 187 (1995) [JETP **80**, 100 (1995)].

Translated by Eugene Yankovsky

Effect of charge exchange on the energy distribution of fast multiply charged ions propagating through matter

V. V. Balashov, A. V. Bibikov, and I. V. Bodrenko

D. V. Skobel'tsyn Institute of Nuclear Physics, M. V. Lomonosov Moscow State University, 119899 Moscow, Russia

(Submitted 30 August 1996; resubmitted 31 December 1996)
Zh. Éksp. Teor. Fiz. **111**, 2226–2236 (June 1997)

A technique for calculating, in the diffusion approximation, energy distributions of multiply-charged ions with an arbitrary number of charge states propagating through matter has been suggested. Examples of numerical solutions of kinetic equations taking into account charge exchange between ions and matter are given. A compact solution for the special case of two charge states has been found. The calculations are compared to experimental data. © *1997 American Institute of Physics.* [S1063-7761(97)02106-9]

1. INTRODUCTION

Propagation of multiply-charged ions through matter during which they capture or lose electrons has been studied extensively. Owing to charge exchange, the regime of ion deceleration in matter is complicated, and no accurate theoretical description of the process has been proposed as yet. The energy distribution of charged particles propagating through matter in the regime of intense charge exchange can be radically different from a Gaussian, but the issue of how multiple transitions between its states affect the shape of the energy distribution in the general case (i.e., before an equilibrium among different charge states of ions is established) has not been settled. On the other hand, in most cases of practical interest, information about characteristics of elementary events of charge exchange is insufficient. In view of this, more attention is focused on the inverse problem of deriving cross sections of electron capture and loss by an ion propagating through matter from measurements of the ion energy distribution.

Recently Sigmund and Närmann[1–3] have made a significant contribution to the theory of ion deceleration in matter. They have derived specific analytic expressions for the set of moments of energy distributions of ions which have passed a certain layer of matter in the presence of charge exchange. Unfortunately, such calculations performed with the aim of deriving the energy distribution function are cumbersome in the general case, when more than two charge states of ions must be taken into account.

In this paper, we describe a different approach to the problem based on fundamental principles of the fluctuation theory of ionization deceleration formulated by Landau.[4] This approach enables us to calculate spectra of ion energy losses taking into account the straggling (energy spread) of the initial distribution with an arbitrary number of charge states without recourse to Monte Carlo techniques. In the important special case of two charge states, we have obtained a compact analytic expression for the energy loss spectrum of such ions.

2. THE CASE OF TWO CHARGE STATES; DISTRIBUTION DENSITIES OF DISTANCES TRAVERSED IN EACH STATE

Suppose that an ion with energy E_0 enters a sample in the charge state $\langle 1\rangle$, then transfers to the state $\langle 2\rangle$ in the course of its motion through the sample, then returns to the original state $\langle 1\rangle$, etc., changing its charge state many times. Let S_1 and S_2 be the effective deceleration rates in states $\langle 1\rangle$ and $\langle 2\rangle$, respectively, and let Ω_1^2 and Ω_2^2 be straggling parameters in these states. Then, in a thin layer, i.e., when changes in the parameters S_1, S_2 and Ω_1^2,Ω_2^2 due to ion deceleration and the energy loss in each recharging event can be neglected, the average energy E of an ion that has passed through a layer with thickness x, so that on a certain fraction of its path x_1 it was in state $\langle 1\rangle$ and on the other fraction $x_2=x-x_1$ in state $\langle 2\rangle$, is

$$E=E_0-S_1x_1-S_2x_2. \tag{1}$$

Suppose that an ion that entered the sample in state $\langle 1\rangle$ has been detected at point x in the same state. Denote by $P_1(x,x_1)$ the probability density that the ion, which passed through distance x was in state $\langle 1\rangle$ for a fraction x_1 of this distance, so that it passed through the rest of this distance, $x_2=x-x_1$, in state $\langle 2\rangle$. Similarly we introduce the probability density $P_2(x,x_1)$ for an ion entering the sample in state $\langle 1\rangle$ to be detected at point x in state $\langle 2\rangle$. We call these parameters the distribution densities for the paths traversed by an ion in a certain charge state.

Let $f_1(x,E)$ and $f_2(x,E)$ be energy distributions of ions detected in the original charge state $\langle 1\rangle$ or in the state $\langle 2\rangle$ at distance x from the entrance to the sample (from now on, we denote by f distribution functions calculated without taking account of straggling due to ionization; in the general case we denote the distribution functions by $F_k(x,E)$). Equation (1) uniquely relates the ion energy loss in deceleration or, more accurately, its average (to within the straggling) energy E at point x, to the distribution of full ion path between the parts x_1 and $x_2=x-x_1$ through which it has passed in states $\langle 1\rangle$ and $\langle 2\rangle$. Using Eq. (1), we determine the relation between the distribution density of distances and the energy spectrum of ions in each charge state:

$$f_k(x,E)dE=P_k(x,x_1)dx_1|_{x_1=(E_0-E-S_2x)/(S_1-S_2)}. \qquad (2)$$

The distributions $P_k(x,x_1)$ and $f_k(x,E)$ are related to probabilities of detecting an ion in states $\langle 1\rangle$ and $\langle 2\rangle$ at distance x from the entrance to the sample by the integral equations

$$p_k(x)=\int_0^x P_k(x,x_1)dx_1=\int f_k(x,E)dE. \qquad (3)$$

We normalize the parameters introduced above in accordance with conservation of the number of particles:

$$p_1(x)+p_2(x)=1. \qquad (4)$$

We now introduce the distance distribution density $P_{1(2)}^{(N)}(x,x_1)$ for ions that have changed charge states N times. An ion entering the sample in state $\langle 1\rangle$ and changing charge an even number of times is again in state $\langle 1\rangle$, whereas after changing charge an odd number of times it is in state $\langle 2\rangle$. Let us start with the distance distribution density for ions that have undergone zero, one, and two transitions:

$$P_1^{(0)}(x,x_1)=e^{-\lambda_{12}x_1}\delta(x_2), \qquad (5)$$

$$P_2^{(1)}(x,x_1)=e^{-\lambda_{12}x_1}\lambda_{12}e^{-\lambda_{21}x_2}, \qquad (6)$$

$$P_1^{(2)}(x,x_1)=e^{-\lambda_{12}x_1-\lambda_{21}x_2}\lambda_{12}\lambda_{21}x_1, \qquad (7)$$

where λ_{12} and λ_{21} are the transition rates from state $\langle 1\rangle$ to $\langle 2\rangle$ and vice versa. The expressions for $P_1^{(N=2n)}(x,x_1)$ and $P_2^{(N=2n+1)}(x,x_1)$ at arbitrary n are derived using the mathematical induction method:

$$P_1^{(2n)}(x,x_1)=e^{-\lambda_{12}x_1-\lambda_{21}x_2}\lambda_{12}^n\lambda_{21}^n\frac{x_1^n}{n!}\frac{x_2^{(n-1)}}{(n-1)!}, \qquad (8)$$

$$P_2^{(2n+1)}(x,x_1)=e^{-\lambda_{12}x_1-\lambda_{21}x_2}\lambda_{12}^{(n+1)}\lambda_{21}^n\frac{x_1^n}{n!}\frac{x_2^n}{n!}. \qquad (9)$$

Summation of all these series yields

$$\begin{aligned}P_1(x,x_1)&=\sum_{n=0}^{\infty}P_1^{(2n)}(x,x_1)\\&=e^{-\lambda_{12}x_1-\lambda_{21}x_2}\Bigg[\delta(x_2)\\&\quad+\sum_{n=1}^{\infty}\lambda_{12}^n\lambda_{21}^n\frac{x_1^n}{n!}\frac{x_2^{(n-1)}}{(n-1)!}\Bigg]\\&=e^{-\lambda_{12}x_1-\lambda_{21}x_2}\Bigg[\delta(x_2)\\&\quad+\sqrt{\frac{\lambda_{12}\lambda_{21}x_1}{x_2}}\,I_1(2\sqrt{\lambda_{12}\lambda_{21}x_1x_2})\Bigg], \qquad (10)\end{aligned}$$

$$\begin{aligned}P_2(x,x_1)&=\sum_{n=0}^{\infty}P_2^{(2n+1)}(x,x_1)\\&=e^{-\lambda_{12}x_1-\lambda_{21}x_2}\lambda_{12}I_0(2\sqrt{\lambda_{12}\lambda_{21}x_1x_2}). \qquad (11)\end{aligned}$$

Here $I_0(x)$ and $I_1(x)$ are the Infeld functions (Bessel functions of imaginary argument). We have used their series expansions[5]

$$I_0(x)=\sum_{k=0}^{\infty}\frac{1}{k!k!}\left(\frac{x}{2}\right)^{2k}, \qquad (12)$$

$$I_1(x)=\sum_{k=0}^{\infty}\frac{1}{k!(k+1)!}\left(\frac{x}{2}\right)^{2k+1}. \qquad (13)$$

Using the known relations for these functions, one can easily prove that substitution of the solutions (10) and (11) into Eq. (3) yields functions that satisfy the normalization condition (4).

Equations (10) and (11) enable one to visualize the energy distribution of ions and its evolution in the course of ion propagation through matter. According to Eq. (10), the distribution $P_1(x,x_1)$ plotted as a function of x_1 has a sharp peak at $x_1=x$, corresponding to the fraction of ions that have not changed charge state (its contribution drops exponentially as $e^{-\lambda_{12}x}$ with increasing penetration depth), and a broad peak due to recharged ions. The shape of the energy distribution $f_1(x,E)$ of ions is similar. If $S_1<S_2$, which corresponds to the charge in state $\langle 2\rangle$ being larger than that in state $\langle 1\rangle$, the spectrum of outgoing ions contains a narrow peak near its high-energy edge due to ions that have passed through matter without changing charge, and a broad peak at lower energies (higher energy losses). If $S_1>S_2$, the relative locations of these peaks are interchanged.

3. GENERAL CASE OF AN ARBITRARY NUMBER OF CHARGE STATES; KINETIC EQUATION FOR DECELERATION WITH ALLOWANCE FOR CHARGE EXCHANGE

Let us start from the principles of the theory of deceleration for particles whose charge remains constant as they propagate through matter.[4] If the mean energy loss $\langle\Delta\rangle=S(E)x$ of a particle on a given section of the path x is much less than its total energy E, while on the other hand $\langle\Delta\rangle$ is much larger than the possible energy transfer $T_{\max}$ to the ambient medium in one collision (i.e., the ''diffusion approximation'' applies), the distribution of energy losses by one particle will be Gaussian:

$$F(\Delta,x|E)=\frac{\exp[-(\Delta-S(E)x)^2/2\Omega^2(E)x]}{\sqrt{2\pi\Omega^2(E)x}}, \qquad (14)$$

where Ω^2 is the straggling parameter for a particle with given energy E:

$$\langle(\Delta-\langle\Delta\rangle)^2\rangle=\Omega^2(E)x. \qquad (15)$$

Under these conditions, the motion of a particle capable of changing charge state due to deceleration is described by the system of kinetic equations

$$\begin{aligned}F_k(x,E)&=e^{-\lambda_k x}\int g_k(x,E'-E)F_k(x=0,E')dE'\\&\quad+\sum_{l\neq k}\lambda_{lk}\int_0^x dx'\int e^{-\lambda_k(x-x')}\end{aligned}$$

$$\times g_k(x-x',E'-\Delta_{lk}-E)F_l(x',E')dE',$$

$$k=1,2,\ldots,N, \tag{16}$$

where, in accordance with Eq. (14),

$$g_k(x-x',E'-E) = \frac{\exp\{-[E'-E-S_k(x-x')]^2/2\Omega_k^2(x-x')\}}{\sqrt{2\pi\Omega_k^2(x-x')}}, \tag{17}$$

S_k and Ω_k^2 are the deceleration ability and straggling parameter, respectively, in state k; $\lambda_k=\Sigma_l\lambda_{kl}$ is the total rate of ion transition from charge state k to other states l; λ_{kl} is the partial rate of ion transition from charge state k to state l; Δ_{lk} is the mean energy loss due to the transition from state l to k with charge exchange; the function $F_k(x=0,E')$ is the initial distribution of ions. If an ion beam enters the sample in the definite charge state $\langle 1\rangle$ and with the definite energy E_0, we have

$$F_k(x=0,E')=\delta_{k1}\delta(E'-E_0). \tag{18}$$

4. ALLOWANCE FOR STRAGGLING IN THE CASE OF TWO CHARGE STATES

The specific shape of the ion energy spectrum defined by Eqs. (10) and (11) can be derived from the general solution of Eqs. (16)–(18) under the conditions summarized in Sec. 2. The analytic expressions (10) and (11) are convenient for calculations and simulations of various situations in the case of two charge states, but, unlike the general solution, they do not take into account the ionization straggling, and the spread of the ions' energy due to their penetration through matter is totally determined by charge exchange. Although quantitative calculations based on Eqs. (16)–(18) indicate that in many cases of practical interest the output ion spectrum is controlled not by straggling but by charge exchange, it is of interest to generalize the results of Sec. 2 with allowance for straggling.

In the diffusion approximation and given a Gaussian energy distribution of ion energies in each charge state, the total energy spread of an ion that has traversed a layer of matter with thickness x, such that it was in state $\langle 1\rangle$ over the fraction x_1 of this distance and in state $\langle 2\rangle$ over the fraction $x_2=x-x_1$, is characterized by the variance

$$D(x,E)=\langle(\Delta E)^2\rangle=\Omega_1^2x_1+\Omega_2^2x_2; \tag{19}$$

and the mean ion energy E can be uniquely determined with Eq. (1).

In this approach, the result of Sec. 2 is generalized to the case of straggling using the simple convolution integral:

$$f_1(x,E)\rightarrow F_1(x,E)=\int g(x,E-E')f_1(x,E')dE', \tag{20}$$

where

$$g(x,E-E')=\frac{\exp[-(E-E')^2/2D(x,E)]}{\sqrt{2\pi D(x,E)}}. \tag{21}$$

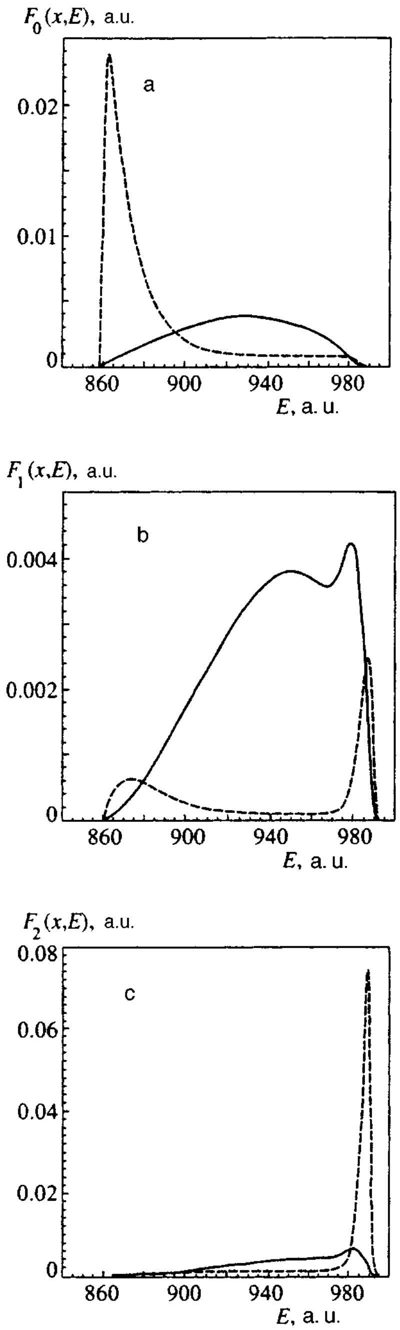


FIG. 1. Energy distributions of ions in states a) $\langle 0\rangle$, b) $\langle 1\rangle$, and c) $\langle 2\rangle$ after passing through a layer $n_cx=100$ (the case of three charge states). The solid curves were calculated with the following parameters: $E_0=1000$, $S_0=1.4$, $S_1=0.35$, $S_2=0.1$, $W_1=W_0=W_2=0.01$, $\sigma_{10}=\sigma_{01}=\sigma_{12}=\sigma_{21}=0.05$; $\sigma_{02}=\sigma_{20}=0$. The dashed curves were calculated at the same parameters, but at a smaller rate of reverse transitions from states $\langle 1\rangle$ and $\langle 2\rangle$ to $\langle 1\rangle$: $\sigma_{01}=\sigma_{21}=0.005$.

5. COMPUTER SIMULATION OF THE DECELERATION PROCESS

Following Närmann,[3] we use the following model parameters: E_0 and E are the initial ion energy and its energy after penetrating a layer of matter of thickness x; n_c is the number density of particles in the layer; σ_{kl} is the effective transition cross section from state k to l, which is related to the partial transition rate λ_{kl} by

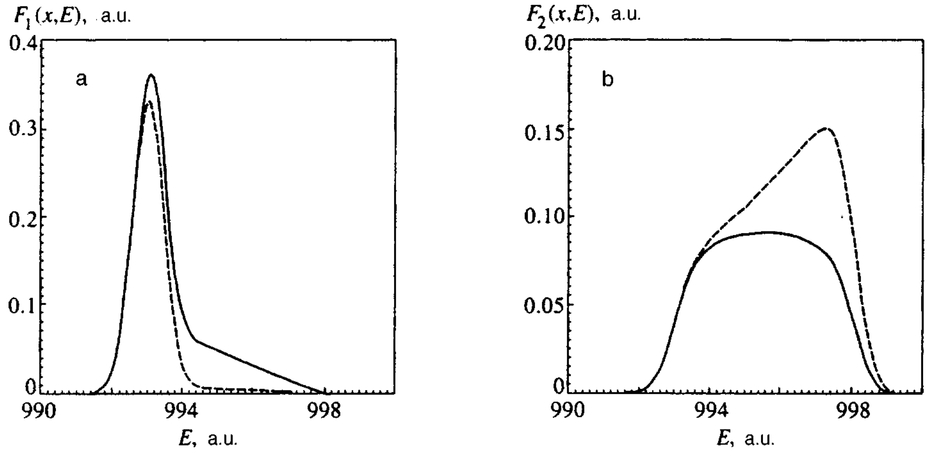


FIG. 2. Energy distributions of ions in states a) $\langle 1\rangle$ and b) $\langle 2\rangle$ after traversing a layer $n_c x = 20$ (the case of two charge states). Solid curves correspond to $E_0 = 1000$, $S_1 = 0.35$, $S_2 = 0.1$, $W_1 = W_2 = 0.01$, $\sigma_{12} = \sigma_{21} = 0.05$. The dashed curves correspond to the same parameters, but $\sigma_{12} = 0.05$ and $\sigma_{21} = 0.005$.

$$\lambda_{kl} = \sigma_{kl} n_c, \tag{22}$$

S_k is the effective deceleration rate of an ion in state k; W_k is the ionization straggling parameter in state k, which is related to the straggling parameter Ω_k^2 by

$$\Omega_k^2 = W_k n_c. \tag{23}$$

All the parameters are expressed in atomic units. In all examples under consideration we assume that an ion beam enters the sample in charge state $\langle 1\rangle$ with definite energy E_0. When three or more states are taken into consideration, the system of kinetic equations (16)–(18) is solved. In cases with two charge states, we use the technique of the distance distribution density with a correction for straggling (Eq. (20)). The results are equivalent to those derived from Eqs. (16)–(18), but obtained without cumbersome calculations demanding a lot of computer time.

It is necessary to extend our analysis beyond the case of two charge states discussed previously by other authors, above all because of the interest in the situation in which charge exchange proceeds in both directions, namely, when the ion charge can be either reduced through the capture of electrons from the environment or increased through the loss of one of its electrons (stripping).

Figure 1 shows the first-ever published calculations of energy distributions of multiply-charged ions with three charge states taken into account. We denote by $\langle 1\rangle$ the initial state, and denote the other two states by $\langle 0\rangle$ and $\langle 2\rangle$. Having assumed that the first of the final states is a result of stripping, while the second results from electron capture, we selected effective deceleration parameters in each state in accordance with the condition

$$S_0 > S_1 > S_2. \tag{24}$$

For comparison, Figs. 2 and 3 also show calculations taking into account only charge exchange during the transitions $\langle 1\rangle \rightarrow \langle 2\rangle$ and $\langle 2\rangle \rightarrow \langle 1\rangle$. Above all, note that since the effective deceleration rate in the initial state $\langle 1\rangle$ is assumed to be higher than in state $\langle 2\rangle$, the spectrum of ions in state $\langle 2\rangle$ is slightly shifted to the high-energy side, as compared to the spectrum in state $\langle 1\rangle$. The comparison between the solid and dashed lines in Figs. 2 and 3 demonstrates how the relation between the transition rates from the initial state $\langle 1\rangle$ to the state $\langle 2\rangle$ and back affects the shapes of ion spectra in each of these states.

The solid curves in Fig. 2 correspond to the case in which the cross sections σ_{12} and σ_{21} are equal. If the transi-

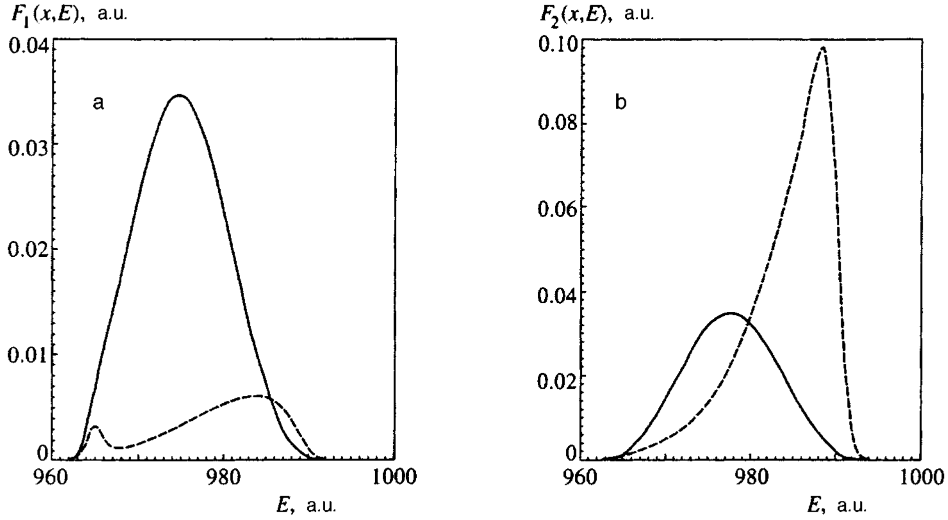


FIG. 3. Energy distributions of ions after traversing a layer $n_c x = 100$ (the case of two charge states). Calculation parameters as the same as in Fig. 2.

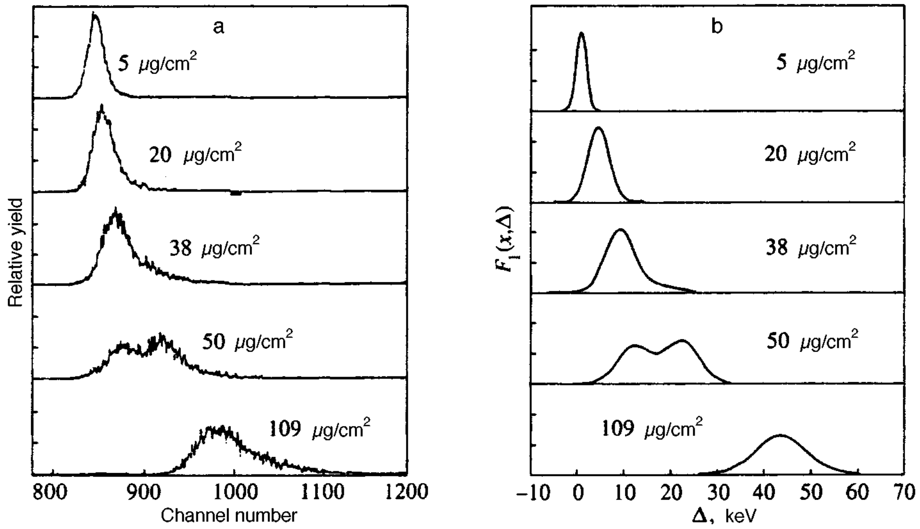


FIG. 4. Energy distribution of Li^{2+} ions with initial energy $E_0 = 63.4$ MeV after traversing thin carbon films of different thicknesses: (a) experimental data by Ogawa *et al.*[6] (the abscissa is the channel number); (b) our calculations (the abscissa is the energy loss in keV).

tion rate from state ⟨1⟩ to ⟨2⟩ (dashed lines) is higher, the reverse transition of particles from state ⟨2⟩ to the initial state will be more difficult. As a result, the amplitude of the peak corresponding to state ⟨1⟩ will be lower, while at the same time the contribution of higher energies to the spectrum of ions in the state ⟨2⟩ will increase (a change in the ratio between the cross sections in favor of the transition from the state ⟨2⟩ to ⟨1⟩ has the opposite effect). Useful information concerning the ratio between the rates of direct and inverse charge exchange due to ion deceleration can probably be derived from the asymmetry of the peak $F_2(E)$ in the energy distribution.

The spread in the output ion energy in either state increases with path length (Fig. 3), mostly owing to charge exchange, whereas the contribution of ionization straggling becomes less and less important. Given a small rate of the ⟨2⟩→⟨1⟩ reverse transition and a large rate of the direct transition ($\sigma_{12} n_c x > 1$, $\sigma_{21} n_c x < 1$), which corresponds to dashed curves in Fig. 3, the flux of ions in state ⟨1⟩ is very weak and consists mostly of ions that have undergone charge exchange twice, ⟨1⟩→⟨2⟩→⟨1⟩. The fraction of ions that have traversed the sample without charge exchange is represented by only the small, narrow peak on the left slope of the main peak. As was first shown by Närmann,[3] such a two-bump curve of energy distribution of multiply-charged ions which have passed through matter is a typical indicator of charge exchange.

The energy distribution of ions calculated for both increases and decreases in the charge state is much more complicated than in the problem of two charge states. The range of energy loss is notably wider, and the spectrum shape is more complex and very sensitive to the relationships among transitions rates between various charge states during deceleration.

Let us compare the calculations shown by the solid curves in Figs. 1 and 3. The difference between these graphs is that the calculations given in Fig. 1 take into account transitions from the initial state ⟨1⟩ to the state ⟨0⟩ and back, the deceleration rate in the state ⟨0⟩ being very high. It is natural that these transitions result in broader peaks corresponding to all states in Fig. 1, as compared with those in Fig. 3. At the same time, a narrow peak similar to the main peak in Fig. 3 can be seen on the right of the ion spectra in states ⟨1⟩ and ⟨2⟩ in Fig. 1. An interesting point is that this peak, which is due to ions decelerated in state ⟨1⟩, is equally prominent in the spectrum of ions in state ⟨2⟩ as that of ions in state ⟨1⟩.

The dashed curves shown in Fig. 1 are interesting from another point of view. If the rates of reverse transitions from states ⟨0⟩ and ⟨2⟩ to initial charge state ⟨1⟩ are low, the spectrum of ions in state ⟨1⟩ will contain two clearly resolved features corresponding to ions decelerated in states ⟨1⟩ and ⟨0⟩. Note also a wide plateau on the right of the peak due to ions in state ⟨0⟩.

6. COMPARISON WITH EXPERIMENTAL DATA

Recently Ogawa *et al.*[6] measured energy distributions of lithium ions with initial energies of about 10 MeV/nucleon downstream of a carbon target. Generally, the changes in the shapes of these distributions with increasing ion path length are in agreement with the calculations described in the previous section. In particular, the experimental curves clearly demonstrate evolution of the spectrum from a shape with one bump at small target thicknesses to a two-bump curve, and then to a spectrum with one much broader bump, which is typical of ion deceleration in the presence of intense charge exchange. By measuring the gradual drop in the number of ions traversing the target without changing their charge state, Ogawa *et al.*[6] determined partial characteristics (attributed to different charge states) of the effective ion deceleration in matter. They did not discuss the issue of a comprehensive theoretical description of all resulting distributions, however; therefore the problem of separating the contribution of ions traversing a layer of matter without charge exchange remained unsolved.

Using the technique described above, we have calculated parameters of Li^{2+} ions shown in Fig. 4 alongside measurements by Ogawa *et al.*[6] The parameters of ion deceleration and charge exchange were taken to be close to estimates given in Ref. 6, namely, $S(\mathrm{Li}^{2+}) = 250\ \mathrm{eV}\cdot(\mu\mathrm{g/cm}^2)^{-1}$,

$S(\mathrm{Li}^{3+}) = 360$ eV $\cdot(\mu g/\mathrm{cm}^2)^{-1}$, $\sigma(\mathrm{Li}^{2+}\rightarrow\mathrm{Li}^{3+}) = 4\times10^{-18}$ cm^2, $\sigma(\mathrm{Li}^{3+}\rightarrow\mathrm{Li}^{2+}) = 2\times10^{-22}$ cm^2. The energy loss due to charge exchange in one cycle $\mathrm{Li}^{2+}\rightarrow\mathrm{Li}^{3+}\rightarrow\mathrm{Li}^{2+}$ has been estimated to be 5.6 keV.[7] A full comparison between the calculations and measurements is difficult because, firstly, the data of Ref. 6 are plotted on a relative energy scale, and secondly, we do not know the initial energy spread of ions in the experiment. Thus, the straggling parameter $W_1 = W_2 = W$ introduced into our calculations as a fitting parameter includes, in fact, both the straggling proper and the initial energy spread of the ion beam. Note that the value $W = 2.5\times10^5$ eV$^2\cdot(\mu$g/cm$^2)^{-1}$ obtained by fitting our calculations to the measurements is below the Bohr limit calculated for asymptotically high ion velocities.

In our opinion, good agreement between the main features of theoretical and experimental results is beyond doubt.

7. CONCLUSIONS

Interaction between fast multiply-charged ions and matter has aroused the growing interest of researchers working in various fields of modern physics,[8–11] such as a solid state physics, atomic and nuclear physics, physics of cosmic rays and astrophysics, and biophysics. Recent experimental studies concerning the role of charge exchange in formation of energy distributions of ions in these processes and new possibilities for accurate theoretical analysis show promise for purposeful studies in this field.

The technique of coupled integral equations, on which our study has been based, is a reliable method of investigating penetration kinetics of fast multiply-charged ions through matter in the presence of charge exchange. It allows us to calculate energy distributions of ions in various charge states as a function of their path and ratios among rates of transitions between various charge states of ions without limitations on the number of charge states. A solution to the problem of energy distribution of ions in a specific case of two charge states has been found in the form of compact analytic expressions.

[1] P. Sigmund, Nucl. Instrum. Methods Phys. Res. B **69**, 113 (1992).

[2] A. Närmann and P. Sigmund, Phys. Rev. A **49**, 4709 (1994).

[3] A. Närmann, Phys. Rev. A **51**, 548 (1995).

[4] L. Landau, J. of Phys. (USSR) **8**, 201 (1944).

[5] I. S. Gradshtein and I. M. Ryzhik, *Tables of Integrals, Series, and Products*, Academic Press, New York (1965).

[6] H. Ogawa, I. Katayama, I. Sugai *et al.*, Nucl. Instrum. Methods Phys. Res. B **115**, 66 (1996).

[7] H. Ogawa, I. Katayama, H. Ikegami *et al.*, Phys. Lett. A **160**, 77 (1991).

[8] F. Närmann, W. Heiland, R. Manreal *et al.*, Phys. Rev. B **44**, 2003 (1991).

[9] H. Bichsel, Rev. Mod. Phys. **60**, 663 (1988).

[10] H. Paul and M. J. Berger, in *Atomic and Molecular Data Needed in Radiotherapy*, IAEA-TECDOC; Vienna (1993), Ch. 7.

[11] P. Stavrev, N. Gavrilova-Stavreva, S. Petrov, J. Phys. G **18**, 1833 (1992).

Translation provided by the Russian Editorial office.

Effect of impurities on thermoelectric power due to phonon drag in metals

K. D. Belashchenkov*⁾

Kurchatov Institute, 123182 Moscow, Russia,

D. V. Livanov

Moscow Institute of Steel and Alloys, 117936 Moscow, Russia

(Submitted 21 November 1996)
Zh. Éksp. Teor. Fiz. **111**, 2237–2242 (June 1997)

The effect of inelastic impurity scattering of electrons on the thermoelectric power due to phonon drag in metals has been studied. It is shown that this is the main cause of the thermoelectric power suppression due to doping at low temperatures. The thermoelectric power in a metal with a quadratic electron spectrum has been calculated as a function of temperature and impurity concentration. In addition to the impurity concentration, the correction to the thermoelectric power due to inelastic scattering contains the large factor Θ_D/T. *© 1997 American Institute of Physics.* [S1063-7761(97)02206-3]

It is well known that the thermoelectric power in normal metals contains two main contributions, namely, the diffusion thermoelectric power and the contribution due to phonon drag. The diffusion thermoelectric power is linear in temperature and can be derived from the expansion in terms of the parameter T/ϵ_F, where T is the temperature and ϵ_F is the Fermi energy. The thermoelectric power induced by the phonon drag results from a nonequilibrium contribution to the phonon distribution function due to external electric field, which does not contain an expansion near the Fermi level and leads to a thermal flow of phonons. There is, usually, a broad temperature range in which the two contributions are comparable in magnitude. In pure metals at low temperatures, the contribution of the phonon drag is usually greater than the diffusion thermopower, but is suppressed considerably by doping.[1] It is clear that the thermoelectric power due to the phonon drag should strongly depend on relative contributions of various phonon scattering processes. If the electron–phonon scattering is dominant, the contribution of the phonon scattering in the thermoelectric power achieves its maximum, and if the contribution of other processes (such as phonon–phonon or phonon–impurity scattering) is larger, the thermoelectric power is lower, because these processes, unlike the scattering on nonequilibrium electrons, tend to mix the phonon distribution over the angles. The probability of impurity scattering of a phonon is proportional to $c_{\mathrm{imp}}q^4$, where c_{imp} is the impurity concentration and q is the phonon wave vector, whereas the probability of mutual scattering of two long-wave phonons is proportional to the product of the wave vectors of all phonons involved in the process: $w \propto q_1q_2q_3$. At low temperatures, typical phonon wave vectors are proportional to T, hence the probabilities of these processes drop rapidly with decreasing temperature. Moreover, since the total wave vector is conserved in mutual scattering of long-wave phonons, this process does not change the total quasimomentum of phonons. It can be changed only by umklapp processes, and for the metals with a closed Fermi surface their probability drops with decreasing temperature exponentially as $\exp(-\Delta_0/T)$, where Δ_0 is of the same order as Θ_D.

It will be shown below that the thermoelectric power is suppressed by impurities at c_{imp} notably lower than that needed for strong impurity scattering of phonons. The more important factor in the suppression of the thermoelectric power is nonelastic impurity scattering of electrons, i.e., the electron scattering with the emission or absorption of a phonon. Along with other processes of the electron–phonon–impurity interference, this scattering determines a conductivity of impure metals at low temperatures.[2,3] Compared to the vertex g of electron–phonon scattering, the vertex γ referring to inelastic impurity scattering of an electron contains an additional factor of the order of $1/ql$, where $l=v_F\tau$ is the electron mean free path and τ is the electron collision time.[3] The additional term in the drag thermopower S_g due to the inelastic impurity scattering of electrons can be estimated assuming that this correction is of the same order as the ratio of the effective rate of absorption (emission) of phonons in the events of the inelastic scattering of electrons to the effective rate of the ''pure'' phonon–electron scattering, which equals $1/q_Tl$, where $q_T\sim T/u_l$ is the typical wave vector of a thermal longitudinal phonon and u_l is the velocity of longitudinal sound.[3] It turns out that the drag thermopower as a function of impurity concentration is estimated by the following formula:

$$\frac{S_g(c_{\mathrm{imp}})-S_g(0)}{S_g(0)}\sim-\frac{1}{q_Tl}\sim-c_{\mathrm{imp}}\,\frac{\Theta_D}{T}. \tag{1}$$

In order to describe the thermoelectric power in metals at low temperatures, when interference of various scattering mechanisms is significant, we have used the kinetic equation based on the diagrammatic technique developed by Keldysh,[4,2] in which Green's functions and polarization operator are expressed in the form of matrices:

$$\hat{G}=\begin{pmatrix}0 & G^A\\ G^R & G^C\end{pmatrix},\quad \hat{D}=\begin{pmatrix}0 & D^A\\ D^R & D^C\end{pmatrix},\quad \hat{\Pi}=\begin{pmatrix}\Pi^C & \Pi^R\\ \Pi^A & 0\end{pmatrix}. \tag{2}$$

Here the superscripts A and R denote the advanced and retarded functions, respectively, and the diagonal components

G^C and D^C contain information about the nonequilibrium state of the system. The phonon Green's function is expressed as

$$D^R(\omega,\mathbf{q}\lambda)=[D^A(\omega,\mathbf{q}\lambda)]^*=\frac{\omega_{\mathbf{q}}\lambda}{2}\,[(\omega-\omega_{\mathbf{q}\lambda}+i0)^{-1}-(\omega+\omega_{\mathbf{q}\lambda}+i0)^{-1}], \tag{3}$$

where $\omega_{\mathbf{q}\lambda}=u_\lambda q$ is the frequency of the phonon with wave vector $\mathbf{q}$, and u_λ is the velocity of phonons with polarization λ.

In calculating the conductivity, one can substitute into the kinetic equation the equilibrium distribution function of phonons, but in analysis of phonon drag, one must calculate this distribution using an independent kinetic equation. In the semiclassical approximation, the desired system of two kinetic equations is reduced to two classical Boltzmann equations for electrons and phonons.[5] In order to take into account the inelastic impurity scattering of electrons, we derive a kinetic equation for phonons similar to the corresponding equation for electrons.[4] Let us express the Dyson's equation for phonons in two versions:

$$\begin{aligned}
D_0^{-1}(x_2)\hat{D}(x_2,x_1)&=-u^2\Delta_2\hat{\sigma}_x\left[\delta(x_1-x_2)+\int d^4y\,\hat{\Pi}(x_2,y)\hat{D}(y,x_1)\right],\\
D_0^{-1}(x_1)\hat{D}(x_2,x_1)&=-u^2\Delta_1\left[\delta(x_1-x_2)+\int d^4y\,\hat{D}(x_2,y)\hat{\Pi}(y,x_1)\right]\hat{\sigma}_x,
\end{aligned} \tag{4}$$

where

$$x=(t,\mathbf{r}),\quad D_0^{-1}(x)=-\partial^2/\partial t^2+u^2\Delta. \tag{5}$$

Here Δ is the Laplacian. The equations for different branches of the phonon spectrum are written separately, and the subscripts on the sound velocity u for different branches are omitted. Since phonons do not carry an electric charge, the potentials of electromagnetic field do not enter the Dyson's equation for phonons explicitly.

Let us calculate the thermal flux of phonons as a response to an external electric field. Subtract the second of Eqs. (4) from the first, change the variables x_1 and x_2 for $X=(x_1+x_2)/2=(t,\mathbf{R})$ and $x=x_2-x_1=(\theta,\mathbf{r})$, and calculate the Fourier transform with respect to the variables of relative motion θ and $\mathbf{r}$. In a steady state, when the temperature is uniform, all the components of functions $\hat{\Pi}$ and $\hat{D}$ are independent of $\mathbf{R}$ and t. (Since phonons are not electrically charged, no Poisson brackets[2] emerge in the equations.) Therefore the left-hand side of the resulting equation is identically equal to zero. The right-hand side is the collision integral for phonons and is expressed as follows:

$$I_{\text{ph}}=\omega_q^2[\Pi^C(D^R-D^A)-(\Pi^R-\Pi^A)D^C]. \tag{6}$$

The nonequilibrium component of the phonon Green's function D^C is expressed in terms of the phonon distribution function:

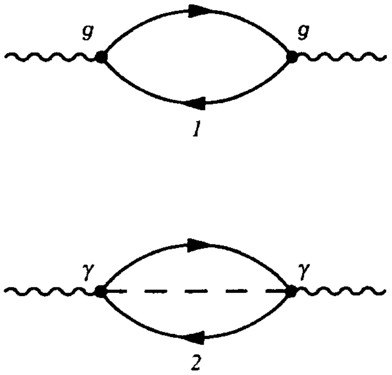


FIG. 1. Feynman graphs responsible for the main contribution to the phonon polarization operator in calculations of the drag thermopower. The solid lines correspond to the electron Green's functions, wavy lines stand for the phonon Green's functions, and the dashed lines denote impurity scattering.

$$D^C=h(\omega,\mathbf{q})(D^R-D^A). \tag{7}$$

In equilibrium $h(\omega,\mathbf{q})=h_0(\omega)=\coth(\omega/2T)$. Let us express the nonequilibrium component of the polarization operator Π^C similarly:

$$\Pi^C=g(\omega,\mathbf{q})(\Pi^R-\Pi^A). \tag{8}$$

In equilibrium $g_0(\omega)=h_0(\omega)$. By substituting Eqs. (7) and (8) into the expression for the phonon collision integral (6) and equating it to zero, we obtain a solution of the phonon kinetic equation:

$$h(\omega,\mathbf{q})=g(\omega,\mathbf{q}). \tag{9}$$

In order to calculate drag the thermopower, we have to calculate the function $h(\omega,\mathbf{q})$ using Eqs. (8) and (9) and substitute it into the expression for the phonon thermal flux:

$$\mathbf{Q}_{\text{ph}}=\frac{i}{2}\int\omega\,\frac{d\omega_{\mathbf{q}}}{d\mathbf{q}}\,\frac{D^C(\omega,\mathbf{q})}{\omega_{\mathbf{q}}}\,\frac{d^3q\,d\omega}{(2\pi)^4}. \tag{10}$$

We assume that the electron momentum is dissipated mostly due to the elastic impurity scattering, when the so-called ''weak Mattissen rule'' applies. This regime occurs at low temperatures, when the effective rate of the electron–phonon scattering $\nu_{\text{e-ph}}\sim T^3/\Theta_D^2$ (see Ref. 5) is smaller than the rate τ^{-1} of the electron–impurity scattering, i.e., $T\tau\ll(\Theta_D/T)^2$.

As was noted above, at low temperatures the pure electron–phonon scattering and inelastic electron–impurity scattering should be taken into account. Figure 1 shows the Feynman graphs which make the main contribution to the drag thermopower. We must calculate the function $I=-i[\Pi^C-h(\omega,\mathbf{q})(\Pi^R-\Pi^A)]$, which is proportional to the phonon collision integral (6), and equate I to zero in order to determine the function $h(\omega,\mathbf{q})$. Let us first express this function disregarding the inelastic impurity scattering (diagram 1):

$$I_0=\frac{1}{2}\,g_{\mathbf{q}\lambda}^2\int\frac{d^3p\,d\epsilon}{(2\pi)^4}\,[1-S_{\epsilon,\mathbf{p}}S_{\epsilon+\omega,\mathbf{p}+\mathbf{q}}-h(\omega,\mathbf{q})\times(S_{\epsilon,\mathbf{p}}-S_{\epsilon+\omega,\mathbf{p}+\mathbf{q}})]\,\text{Im}\,G^A_{\epsilon,\mathbf{p}}\,\text{Im}\,G^A_{\epsilon+\omega,\mathbf{p}+\mathbf{q}}. \tag{11}$$

Next we introduce the standard linearization of the electron and phonon distribution functions in the usual form:

$$S(\epsilon,\mathbf{p})=S_0(\epsilon)-\frac{\partial S_0(\epsilon)}{\partial\epsilon}\,\phi(\epsilon,\mathbf{p}),$$

$$h(\omega,\mathbf{q})=h_0(\omega)-\frac{\partial h_0(\omega)}{\partial\omega}\,\chi(\omega,\mathbf{q}). \tag{12}$$

Here $S_0(\epsilon)=\tanh[(\mu-\epsilon)/2T]$ is the equilibrium electron distribution function. Since we assume that the electron momentum is dissipated due to the elastic impurity scattering, the correction to the electron distribution function is expressed in the conventional form:

$$\phi(\epsilon,\mathbf{p})=e\tau\mathbf{v}\mathbf{E}, \tag{13}$$

where e is the electron charge, $\mathbf{v}=d\epsilon_{\mathbf{p}}/d\mathbf{p}$, and $\mathbf{E}$ is the applied electric field. After linearization, the expression in the square brackets in Eq. (11) is reduced to

$$R^{\omega,\mathbf{q}}_{\epsilon,\mathbf{p}}=-[S_0(\epsilon)-S_0(\epsilon+\omega)]\,\frac{dh_0(\omega)}{d\omega}$$
$$\times(\phi_{\epsilon+\omega,\mathbf{p}+\mathbf{q}}-\phi_{\epsilon,\mathbf{p}}-\chi_{\omega,\mathbf{q}}). \tag{14}$$

It is obvious that, if no umklapp processes take place,1) the right-hand side of Eq. (11) vanishes if $\chi_{\omega,\mathbf{q}}=e\tau\mathbf{q}\cdot\mathbf{E}/m$. If both umklapp processes and inelastic impurity scattering are neglected, the thermoelectric coefficient β (by definition, $\mathbf{Q}=-\beta T\mathbf{E}$) due to the phonon drag is given by

$$\beta^0_q=\frac{2\pi^2}{45}\,\frac{e\tau T^3}{mu_l^3}. \tag{15}$$

This is the well-known result deriving from the classical Boltzmann equation for electrons and phonons.[5]

Now let us consider the graphs taking into account the impurity scattering. Our analysis indicates that the main contribution to the collision integral is due to diagram 2 in Fig. 1. It is clear that, in the combination of distribution functions $R^{\omega,\mathbf{q}}_{\epsilon,\mathbf{p}}$ for these diagrams, $\mathbf{q}$ is replaced with the momentum $\mathbf{p}'$ imparted to an impurity. In this case, integration should be also performed over $\mathbf{p}'$. By substituting the expressions for the vertices $g_{\mathbf{q}\lambda}$ and $\gamma_{\mathbf{p}-\mathbf{p}',\lambda}$[3] into the collision integral and equating it to zero, we obtain an equation from which $\chi_{\omega,\mathbf{q}}$ can be derived:

$$0=\int\frac{d^3p\,d\epsilon}{(2\pi)^4}\,[S_0(\epsilon)-S_0(\epsilon-\omega)]\,\frac{dh_0(\omega)}{d\omega}$$
$$\times\mathrm{Im}\,G^A_{\mathbf{p}\epsilon}\Bigg[\epsilon_F^2(\mathbf{q}\mathbf{e}_\lambda)^2(\phi_{\mathbf{q}}-\chi_{\omega,\mathbf{q}})\mathrm{Im}\,G^A_{\epsilon+\omega,\mathbf{p}+\mathbf{q}}+\frac{9}{\pi\nu\tau}$$
$$\times\int\frac{d^3p'}{(2\pi)^3}\,(\mathbf{p}'\mathbf{e}_\lambda)^2(\phi_{\mathbf{p}'}-\chi_{\omega,\mathbf{q}})\mathrm{Im}\,G^A_{\epsilon+\omega,\mathbf{p}+\mathbf{p}'}\Bigg]. \tag{16}$$

The integrals over $\mathbf{p}$ for $(q,p')l\gg 1$ have the form

$$\frac{1}{\pi\nu\tau}\int\frac{d^3p}{(2\pi)^3}\,\mathrm{Im}\,G^A_{\epsilon,\mathbf{p}}\,\mathrm{Im}\,G^A_{\epsilon+\omega,\mathbf{p}+\mathbf{p}'}$$
$$=\begin{cases}\pi/4p'l, & p'<2p_F\\ \sim(2p_F/p')^2/(p'l)^2, & p'>2p_F\end{cases} \tag{17}$$

(compare to Eq. (44) in Ref. 2).

In crystals with cubic lattices, electrons are decoupled from transverse phonons (the vertex $g_{\mathbf{q}\lambda}$ for transverse phonons vanishes). Thus, it follows from Eq. (16) that transverse phonons are not ‘‘dragged.’’ For longitudinal phonons $\mathbf{e}_\lambda=\mathbf{q}/q$. It is clear that $\chi_{\omega,\mathbf{q}}$ depends only on $\mathbf{q}$, hence follows that the phonon collision integral is zero when the following condition holds:

$$\epsilon_F^2 q(\phi_{\mathbf{q}}-\chi_{\mathbf{q}})=\frac{9}{\pi\nu\tau}\,\chi_{\mathbf{q}}\int_0^{2p_F}\frac{d^3p}{(2\pi)^3}\,\frac{(\mathbf{p}\cdot\mathbf{n})^2}{p}. \tag{18}$$

Here $\mathbf{n}=\mathbf{q}/q$. The upper limit of integration over p is selected in accordance with Eq. (17). The divergence of the integral in the region $p>2p_F$ is purely formal because Eq. (13) for the distribution function $\phi_{\mathbf{p}}$ was derived in the linear approximation and is valid only near the Fermi surface. Therefore, it clearly follows from Eq. (17) that the contribution of the region $p>2p_F$ to the collision integral is negligible. It is quite natural, because $2p_F$ is the maximum momentum which can be transferred to an impurity from an electron without its leaving the Fermi surface.

Finally, we obtain the expression

$$\chi_{\mathbf{q}}=\frac{ql}{ql+24/\pi}\,\phi_{\mathbf{q}}. \tag{19}$$

Hence follows that the smaller the phonon wave vector, the weaker the phonon drag. Let us calculate the relative contribution of the inelastic impurity scattering to the drag thermopower. To this end, the correction to the phonon distribution function due to the inelastic impurity scattering must be substituted into the expression (10) for the thermal flux. After integrating over $\mathbf{q}$ and ω, we obtain the relative correction to the thermoelectric coefficient:

$$\frac{\beta_g-\beta_g^0}{\beta_g^0}=-1.41\,\frac{1}{\epsilon_F\tau}\,\frac{\Theta_D}{T}. \tag{20}$$

In order to express our result in the form convenient for comparison with experimental data, let us present it in the form of the differential drag thermopower $S_g=-\beta_g/\sigma$. In the considered temperature range, the relative correction to the residual conductivity due to the inelastic impurity scattering and the electron–phonon–impurity interference is of the order of $T^2/\epsilon_F\Theta_D$ and is much smaller than the correction (20).[3] Finally, we have

$$S_g=\frac{2\pi^2}{45}\,\frac{T^3}{n_0 e u_l^3}\left(1-1.41\,\frac{1}{\epsilon_F\tau}\,\frac{\Theta_D}{T}\right). \tag{21}$$

Here n_0 is the number of electrons per unit volume. The dependence of the thermoelectric power on the impurity concentration is fully described by Eq. (20).

To sum up, the relative correction to the drag thermoelectric power due to inelastic impurity scattering can be, according to the simple estimate (1), considerably larger than the impurity concentration, which equals by the order of magnitude $1/\epsilon_F\tau$. The accurate calculation of the integral in Eq. (17) shows that it deviates notably from the proposed asymptotics only in the region $ql\sim 1$. In this region the integral rapidly approaches the constant value 1/2, while $S_g\rightarrow 0$. Therefore, Eq. (20) is valid almost everywhere in the region $q_T l > 1$, and the drag thermoelectric power totally vanishes when $c_{\text{imp}}\sim T/\Theta_D$. Thus, the effect of phonon drag is suppressed at low temperatures by relatively low concentrations of impurities.

The authors acknowledge helpful discussions with A. V. Sergeev. The work was supported by the Russian Fund for Fundamental Research (Grant No. 97-02-16877).

*⁾E-mail: kdbel@cerber.net.kiae.su

[1] In order to incorporate umklapp processes, one should add to the collision integral the sum over the vectors **G** of the reciprocal lattice and replace the subscript $\mathbf{p}+\mathbf{q}$ with $\mathbf{p}+\mathbf{q}+\mathbf{G}$.

---

[1] F. Blatt *et al.*, *Thermoelectric Power of Metals*, Plenum Press, New York (1976).
[2] B. L. Al'tshuler, Zh. Éksp. Teor. Fiz. **75**, 1330 (1978) [Sov. Phys. JETP **48**, 670 (1978)].
[3] M. Yu. Reĭzer and A. V. Sergeev, Zh. Éksp. Teor. Fiz. **92**, 2291 (1987) [Sov. Phys. JETP **65**, 1291 (1987)].
[4] L. V. Keldysh, Zh. Éksp. Teor. Fiz. **47**, 1515 (1964) [Sov. Phys. JETP **20**, 1018 (1964)].
[5] E. M. Lifshitz and L. P. Pitaevskiĭ, *Physical Kinetics*, Pergamon Press, New York (1982).

Translated by Russian Editorial office

Nonlinear evolutionary processes in a free-electron laser generator of diffracted radiation

V. P. Shestopalov

A. Ya. Usikov Institute of Radiophysics and Electronics, Ukrainian National Academy of Sciences, 310085 Khar'kov, Ukraine

(Submitted 5 June 1996)
Zh. Éksp. Teor. Fiz. **111**, 2243–2262 (June 1997)

Nonlinear evolutionary processes with two control parameters, one of which is related to the electrodynamic structure (positive feedback) and the other is related to the constant electric field applied to the electron flux, are studied in a free-electron laser, which is a diffracted-radiation oscillator. To this end, first, the linear spectral problem for an open-cavity resonator is investigated and the dispersion relation near the Morse critical point is established. Then the nonlinear evolutionary equation with two control parameters is constructed. Analysis of the latter makes it possible to determine the properties of the parametric dependence of the bifurcation and structural stability, which are determined by small variations of the control parameter (tuning of the cavity). This explains the operating efficiency of a diffracted-radiation oscillator in the millimeter and submillimeter wavelength ranges. © *1997 American Institute of Physics.* [S1063-7761(97)02306-8]

1. INTRODUCTION

In a diffracted-radiation oscillator—a free-electron laser,[1] just as in any other O-type microwave oscillator, the electron noise cloud produced by the hot cathode is converted by specially produced static magnetic and electric fields in the region near the cathode into a flux of chaotic particles. As a result of properly designed positive feedback, implemented with the aid of an appropriate electrodynamic structure and an applied constant electric field, these particles give rise in the interaction space to uniform and rectilinear motion of the electrons, which for certain parameters of the structure and the electron flux are then capable of generating coherent radiation. How can completely disordered free electrons be converted into a highly organized assemblage of particles capable of producing high-frequency coherent electromagnetic fields? This question can just as well be asked about physical, chemical, biological, computational, and other problems which require solving for the dynamics of complex processes, bearing in mind above all their nonlinearity and the possibility of self-organization, an understanding and description of which can be obtained by introducing control parameters (or order parameters). In this respect a diffracted-radiation oscillator is an ideal model for studying nonlinear evolutionary processes with two control parameters, one of which is associated with the electrodynamic structure, i.e., positive feedback, while the other — a force parameter — is associated with the constant electric field applied to the electron flux. In Ref. 2, the experimental data were analyzed for the purpose of studying time-independent bush resonances of a diffracted-radiation oscillator with an open structure. Here a nonlinear stationary theory of dynamical processes in the oscillator will be constructed.

The open electrodynamic structure of the oscillator is a complicated open cavity consisting of two reflectors with a special geometric shape: the top reflector is usually spherical or spheroidal (sometimes spherocylindrical) and has a coupling aperture; the bottom reflector is flat or cylindrical and a diffraction grating is positioned on it (the grating only partly covers the surface of the mirror). An electron flux, modulated at the frequency of the radiation and moving uniformly and rectilinearly near the diffraction grating, produces a characteristic surface electromagnetic field that is scattered by the grating and engenders, besides surface waves, at least one volume wave—diffracted radiation, which feeds the open cavity. On account of the multiple reflection of the diffracted radiation from the mirrors in the cavity, the radiation transferred to the load through the coupling slot builds up in the cavity. This is the standard description of the regime of steady-state oscillations in a diffracted-radiation oscillator.[3]

In studying nonlinear evolutionary processes in an oscillator, we first investigate the linear spectral problem for an open cavity, and we determine both the regular and critical points of the spectrum. The theory of Morse critical points of the dispersion relations for a cavity forms the basis for the approach developed here for the construction of nonlinear evolutionary equations for a diffracted-radiation oscillator. Since near the Morse points the dispersion relations can be represented in analytic form, the corresponding evolutionary equations (first linear and then nonlinear) can be written out in the usual manner.

The nonlinear evolutionary equations so constructed describe the characteristic oscillatory process, which depends on one control parameter—the geometry of the open cavity.

The other control parameter is phenomenological; it is related to the constant electric field applied to the electron flux. The resulting nonlinear equation differs from the existing equations in that it reflects, above all, the nonlinear process near the Morse points, where the dispersion laws change drastically (normal dispersion turns into anomalous dispersion and vice versa), while taking account of the effects of the applied electrostatic field on the electron flux introduces asymmetry, which alters the stability of the nonlinear system.

Analysis of the nonlinear evolutionary equations makes it possible to establish the properties of the parametric dependence of the bifurcation on the quality of the open cavity, the behavior of the electron flux, and the corresponding idea of structural stability. These are affected by small changes in the control parameters near the critical points, thereby explaining the efficiency of a diffracted-radiation oscillator at millimeter and submillimeter wavelengths.

2. LINEAR SPECTRAL PROBLEM AND EXCITATION OF AN OPEN CAVITY

The construction of an evolutionary equation for a diffracted-radiation oscillator is closely related to the deeper investigation of linear stationary processes, i.e., the solution of the linear spectral problem, and to the development of methods combining homogeneous (spectral) and inhomogeneous (excitation) problems for open electrodynamic structures. The latter structures differ fundamentally from closed structures by virtue of radiation losses, multiconnectedness in the transverse cross section, and the behavior of the electromagnetic field at infinity. The spectrum of the characteristic oscillations (for open cavities) is therefore no longer real, additional requirements arise for the energy relations in different regions of space, and the character of the linear spectral problem itself changes. The emerging problems concerning the characteristic oscillations of an open cavity form a new class of non-self-adjoint linear problems in which the spectral parameter ordinarily enters nonlinearly. Such nonlinear spectral problems of mathematical physics have led to the development of the method of operator functions of one or more complex variables.[4]

1. The details of the operator function method in the construction of linear spectral problems for an open cavity can be properly formulated by a study of the linear nonstationary homogeneous Maxwell's equations with certain initial and boundary conditions:

$$\frac{\partial \mathbf{u}}{\partial t}=A\mathbf{u}, \quad t>0, \quad \mathbf{u}|_{t=0}=\mathbf{u}_0. \tag{1}$$

Here A is the Maxwell operator

$$A=\begin{bmatrix} 0 & c\boldsymbol{\nabla}\times \\ -c\boldsymbol{\nabla}\times & 0 \end{bmatrix}, \quad \mathbf{u}=\begin{bmatrix}\mathbf{E}\\ \mathbf{H}\end{bmatrix},$$

E and **H** are the electric and magnetic field intensities and c is the speed of light in vacuum. We note that in the absence of external sources the free oscillations occurring in a cavity on account of the initial energy introduced must decay with time.

The Laplace transform yields

$$u(t)=\frac{1}{2\pi i}\int_{a-i\infty}^{a+i\infty} e^{pt}(pI-A)^{-1}u_0\,dp, \tag{2}$$

where the integral extends over any straight line Re $p=a$ that does not contain the spectrum of the Maxwell operator A, and I is the identity operator.

Thus, the linear temporal evolution of free cavity oscillations is based on the integral (2); to calculate it, is necessary to know the spectrum of the Maxwell operator A, i.e., the operator function $(pI-A)^{-1}$ as a function of the parameter p. An efficient method for solving this problem is obtained by transforming from Eq. (1) to reduced equations of the type

$$-i\lambda v=Av, \tag{3}$$

i.e., to a study of solutions of Eqs. (1) of the form $u(t)=v\,\exp(-i\lambda t)$, where v satisfies Eq. (3) and certain radiation conditions.

The role of the linear spectral problem in the solution of the problem of excitation of an open cavity now becomes clear. It reduces to a system of operator equations of the second kind with a Hilbert–Schmidt type operator.

2. We now examine the linear spectral problem of the characteristic **H** oscillations of a two-dimensional open cavity formed by two perfectly conducting open circular cylindrical surfaces, and we construct the matrix operator function that depends on the desired spectral parameter.

The problem is to determine the values of the complex spectral parameter k (characteristic frequency $\omega\simeq ck$ of the cavity) for which there exists a nontrivial solution of the homogeneous Helmholtz equation, satisfying boundary conditions on the cavity mirrors, the Meixner type condition, and the Sveshnikov–Reinhart condition for outgoing radiation. This linear problem differs fundamentally from the analogous problem for closed cavities, due to the radiation condition at infinity. The solution of the linear problem by the method of separation of variables in local coordinates has the form of a linear superposition of Fourier series, and by virtue of the properties of Bessel functions and the addition theorem for them, the desired eigenfunction $u(x,y)$ is infinitely differentiable with respect to x and y and is a solution of the Helmholtz equation everywhere in the region $\mathbb{R}^2$ with the exception of the cavity mirrors; in addition, the function $u(x,y)$ satisfies the radiation condition for any $k \in \mathbb{C}_k=\{k:|\arg k|<\pi\}$.

In constructing the matrix operator functions, $u(x,y)$ must satisfy boundary conditions. This leads to paired-off functional equations with a kernel in the form of trigonometric functions for the Fourier coefficients of $u(x,y)$. Therefore the initial linear problem in terms of differential equations is equivalent to paired functional equations, which can be reduced by the method of the Riemann–Hilbert problem[5] to an investigation of an infinite linear homogeneous system of algebraic equations of the second kind

$$x_p^n=\sum_{s=1}^{2}\sum_{m=-\infty}^{\infty} a_{pm}^{ns}(k)x_m^s, \quad n=1,2, \tag{4}$$

where x_p^n are the Fourier components of the function $u(x,y)$ and the matrix elements a_{pm}^{ns} are functions of the spectral parameter k, and are defined in terms of Bessel functions and Legendre polynomials (they are not presented here because of their complexity[5]).

We now introduce the following notation: $\sigma(\Delta)$ is the spectrum of the initial problem; $\sigma_H(A)$ is a set of complex numbers $k \in \mathbb{C}_k$ for which a nontrivial solution of the system

(4) exists. Then it can be proved that the matrices $\|a_{pm}^{ns}(k)\|_{p,m=-\infty}^{+\infty}$ specify the operator kernels in the space l_2 for a definite set k. The matrices $\|a_{pm}^{ns}(k)\|$ define bounded operators $A^{ns}(k):l_2\rightarrow l_2$ in l_2, which we shall regard as the operator functions of the spectral parameter k with domain $\mathbb{C}_k$ and range given by the set of kernels. The system of kernel operator functions $A^{ns}(k)$ is a faithful representation of the initial linear spectral problem and, as follows from Eq. (4), admits of a compact description.

We now write the system (4) in the operator form

$$x^n=\sum_{s=1}^{2} A^{ns}(k)x^s, \quad n=1, 2. \tag{5}$$

The system of operator equations (5) is an eigenvalue problem for a matrix operator function, and it—and hence the system (4)—can also be written as an operator equation in l_2^2:

$$\{I-A(k)\}x=\theta, \tag{6}$$

where I is the identity operator in l_2^2 and θ is the zero element.

Now the properties of the set $\sigma(\Delta)$ can be studied by the spectral-theory methods as applied to the operator functions $I-A(k)$. This is related to the study of the analytic properties of the matrix operator function $A(k)=\|A^{ns}(k)\|_{n,s=1}^{2}$ as a function of the parameter k. In so doing, it is proved that the spectrum of two-dimensional electromagnetic oscillations of an open cavity is of finite multiplicity, discrete, and lies in the region $\{k:\mathrm{Im}\, k<0\}$. Thus, for any $R_1>0$ and $R_2>0$, the region $\{k:\pi<\arg k<2\pi;\ R_1<|k|<R_2\}$ contains only a finite number of eigenfrequencies of the initial linear spectral problem.

The operator function method described above for solving the linear spectral problem makes it possible to construct and verify a universal algorithm for calculating the eigenfrequencies. The algorithm is based on the fact that the operator equation (6) has a unique solution for all k, and that there exist regular characteristic numbers and Morse critical points of the operator function $I-A(k)$.

If k_m are approximate eigenvalues of the operator function $I-A(k)$, then they converge to the eigenvalues of the operator function $I-A(k)$ as $m\rightarrow\infty$. Since the $A_m(k)$ are finite-dimensional operator functions ($A_m(k)$ is determined from $A_m(k)x=\theta$), the k_m satisfy the equation

$$\det\{A_m(k)\}=0, \tag{7}$$

where $A_m(k)$ is the sequence of operator functions that approximates the operator function $I-A(k)$. The function $\det\{A_m(k)\}$ is analytic in the region $\mathbb{C}_k$. Therefore Newton's method can be used to solve Eq. (7) numerically. Thus, the approximate characteristic numbers are the roots of Eq. (7).

3. We now consider the diffraction of a monochromatic electromagnetic field $\mathbf{E}^0(\omega)\exp(-i\omega t)$, $\mathbf{H}^0(\omega)\exp(-i\omega t)$ by the open cavity whose spectral characteristics were studied in Sec. 2. The requirement is to determine the solution $\mathbf{E}^g(\omega,x,y)$, $\mathbf{H}^g(\omega,x,y)$ of the homogeneous system of Maxwell's equations that satisfy the boundary conditions at the perfectly conducting mirrors of the cavity, the Meixner condition, and the Malyuzhints extinction condition. A uniqueness theorem holds for this vector boundary-value problem. It follows from this theorem and the properties of $\mathbf{E}^0$ and $\mathbf{H}^0$ that the problem reduces to two scalar problems with $u^1=H_z^g$ and $u^2=E_z^g$, and u^1 and u^2 must satisfy the aforementioned conditions for $\mathbf{E}^g$, $\mathbf{H}^g$. Functions of the type

$$G^H=H_z^0+u^1, \quad G^E=E_z^0+u^2$$

are ordinarily called Green's functions of the Helmholtz equations of the problem under study.

Using the method of separation of variables, we seek $u^{1,2}$ in the form of a Fourier series, and after applying the boundary conditions we find paired functional equations with a kernel in the form of trigonometric functions, which can be reduced by the Riemann–Hilbert method[5] to an inhomogeneous system of algebraic equations of the second kind of the form (4). This system can be regarded as a system of operator equations in l_2,

$$x^q=\sum_{i=1}^{2} A^{qi}x^i+a^q, \quad q=1, 2, \tag{8}$$

where the a^q are determined by the form of the exciting field. If an l_2 solution of the system (8) exists, then we obtain a solution of the initial diffraction problem. Note that the system (8) can be represented as a canonical Fredholm equation in some Hilbert space

$$x=Ax+a, \tag{9}$$

with a completely continuous operator $A:l_2^2\rightarrow l_2^2$. The Fredholm alternative holds for Eq. (9).

The initial problem for the Green's functions G^{HE} describing forced monochromatic oscillations of an open cavity is closely related to the linear spectral problem of characteristic oscillations. It follows from the validity of the Fredholm alternative for Eq. (9) that there exists a bounded operator $(I-A)^{-1}$ defined on the entire space l_2^2, and it follows from the properties of the operator that the resolvent set $\rho(A)$ is not empty; this proves the discreteness and finite multiplicity of the set of eigenvalues of the linear cavity problem under study. The Green's function as a function of a spectral parameter possesses a pole-type singularity in the spectrum of the linear problem; the residues of the Green's function at these poles equal, to within a constant, the functions describing the characteristic oscillations of the cavity: they do not satisfy the radiation condition for the excitation problem, but they do satisfy the more general condition for outgoing radiation (Sveshnikov–Reinhart condition). The converse is also true: if the Green's function possesses a pole-type singularity at some eigenvalue, then the eigenvalue belongs to the set of eigenfrequencies of the open cavity under study.

The Fourier coefficients determined from Eq. (9) make it possible to find, by approximate summation of the series, the Green's functions G^H and G^E, which are important in practice in the resonant frequency range determined by the condition $ka=O(1)$, where a is the characteristic size of the open cavity, and also near the Morse critical points. The solution of Eq. (9) by the method of reduction or the method of successive approximations makes it possible, together with computer calculations, to establish qualitative laws that govern the spectral characteristics and Green's functions of the cavities under study.

3. THEORY OF MORSE CRITICAL POINTS

As shown in Sec. 2, the spectral theory of an open cavity makes it possible to find from the dispersion relation

$$\mathscr{F}(\kappa)\equiv\det\{I-A(k)\}=0, \tag{10}$$

where $\kappa=ka$ ($k=\omega/c$, ω is an eigenfrequency of the cavity, a is any geometric dimension of the cavity), the spectral parameter κ for given nonspectral parameters. The dispersion relation (10) is a characteristic infinite determinant of the kernel operator $A(\kappa)$, and it can only be solved with the aid of a computer. It is important that $A(\kappa)$ also depends on nonspectral parameters, specifically, geometric parameters (the distance l between the mirrors and the radius a and aperture b of a mirror). We confine our attention to one of them: $\chi=b/a$. Then the operator function will have the form $A(\kappa,\chi)$, and the dispersion relation becomes

$$\mathscr{F}(\kappa,\chi)\equiv\det\{I-A(\kappa,\chi)\}=0, \tag{11}$$

where χ takes on values in the domain $\mathscr{D}_\chi \in \mathbb{C}$ of analyticity of $A(\kappa,\chi)$ ($\mathbb{C}$ is the complex plane).

The roots found for Eq. (11) make it possible to construct the functions $\kappa(\chi)$, which determine the dispersion laws of a given cavity. The regular points of the spectrum (11) yield the classic (normal or anomalous) dispersion laws, where the dispersion curves vary continuously as a function of χ. Ordinarily, the investigation is confined to these laws. However, there exist situations in which small variations of a set of nonspectral parameters lead to sharp (catastrophic) variation of the dispersion, strong growth or decrease of the diffraction losses, and appearance of hybrid oscillations. This was found to be possible because of the presence of singular points in the spectrum in Eq. (11)—Morse critical points—near which the indicated events occur. We now construct an analytical theory of such points.

1. The domain of the analytic kernel operator function $A(\kappa,\chi)$ of two complex variables κ and χ is $(\kappa,\chi)\in\Lambda\times\mathscr{D}_\chi$. From a physical standpoint, the parameter χ determines the tuning of the cavity. We assume that the eigenvalues $\kappa\in\Lambda_0$, where $\Lambda_0=\{\kappa\in\mathbb{C}:-\pi<\arg\kappa<\pi;\kappa\neq 0\}$ is the zeroth sheet of Λ.[6]

Since $A(\kappa,\chi)$ is an analytic kernel operator function, the set of zeros of the dispersion relation $\mathscr{F}(\kappa,\chi)=0$ is identical to the set of characteristic numbers $\sigma(\kappa)$ of the operator functions $I-A(\kappa,\chi)$ (the set $\sigma(\kappa)$ is defined on a subset of the set $\mathscr{D}$ in $\mathbb{C}^2$ of the form $\mathscr{D}=\Lambda_0\times\mathscr{D}_\kappa$). It is necessary to determine what happens to $\sigma(\kappa)$ when the eigenvalues κ vary in $\mathscr{D}_\kappa$. For this we introduce an analytic set $\mathscr{D}$ of the form $\sigma_0=\{(\kappa,\chi)\in\mathscr{D}:\mathscr{F}(\kappa,\chi)=0\}$ and treat $\mathscr{F}(\kappa,\chi)$ as a map $\mathscr{F}:\mathbb{C}^2\to\mathbb{C}$ with domain $\mathscr{D}$ (we assume that χ is a complex quantity). If an isolated singular point (κ_0,χ_0) of the map $\mathscr{F}$ exists nearby, then the local structure of σ_0 near (κ_0,χ_0) is determined by the position and type of the point (κ_0,χ_0) (this is consequence of the theory of functions of many complex variables and the characteristic features of continuous differentiable maps[7,8]). Far from (κ_0,χ_0) the set is locally a hyperplane. Clearly, small variations of χ near (κ_0,χ_0) lead to small variations of $\sigma(\kappa)$.

The phenomenological features of hybrid oscillations, which are well known for different oscillatory structures, are determined by the dispersion law—the special behavior of two eigenfrequencies as functions of χ with $\operatorname{Im}\chi=0$ (Wien plot). We are therefore faced with the problem of describing in terms of regular and singular points of the map $\mathscr{F}$ the dispersion law typical of hybrid oscillations. The local structure of σ_0 as a hypersurface does not correspond to the indicated dispersion law. We therefore examine the set of critical points

$$\sigma_c=\left\{(\kappa,\chi)\in\mathscr{D}:\nabla\mathscr{F}=\left(\frac{\partial\mathscr{F}}{\partial\kappa},\frac{\partial\mathscr{F}}{\partial\chi}\right)=0\right\}$$

and the set of isolated Morse critical points

$$\sigma_{Mc}=\left\{(\kappa,\chi)\in\sigma_c:\frac{\partial^2\mathscr{F}}{\partial\kappa^2}\frac{\partial^2\mathscr{F}}{\partial\chi^2}-\left(\frac{\partial^2\mathscr{F}}{\partial\kappa\partial\chi}\right)^2\neq 0\right\}.$$

Let (κ_0,χ_0) be an isolated Morse point near σ_0. Then Eq. (11) can be expanded in a Taylor series about (κ_0,χ_0) and, to within cubic infinitesimals, we obtain

$$\delta+\frac{1}{2}\frac{\partial^2\mathscr{F}}{\partial\kappa^2}(\kappa-\kappa_0)^2+\frac{\partial^2\mathscr{F}}{\partial\kappa\partial\chi}(\kappa-\kappa_0)(\chi-\chi_0)+\frac{1}{2}\frac{\partial^2\mathscr{F}}{\partial\chi^2}(\chi-\chi_0)^2+O^3=0. \tag{12}$$

According to the Morse theorem,[9] by a change of variables the series (12) can be expressed as a canonical quadratic form

$$\tilde{\kappa}^2+\tilde{\chi}^2+\delta=0, \tag{13}$$

where $\delta=\mathscr{F}(\kappa_0,\chi_0)\neq 0$, $\tilde{\kappa}=\kappa-\kappa_0$, and $\tilde{\chi}=\chi-\chi_0$. From Eq. (13) we have $\tilde{\kappa}=\pm i\sqrt{\tilde{\chi}^2+\delta}$. For $\delta=0$ we obtain $\tilde{\kappa}=\pm i\tilde{\chi}$. Equation (13) contains a dispersion law (the dependence of the spectral parameter $\tilde{\kappa}$ on the nonspectral parameter $\tilde{\chi}$) typical of hybrid coupling of the oscillations, and the quantity δ determines the degree of this coupling.

It is sufficient to investigate the functions $\operatorname{Re}\tilde{\kappa}(\tilde{\chi})$ and $\operatorname{Im}\tilde{\kappa}(\tilde{\chi})$ with $\tilde{\chi}$ ranging along the straight lines $\operatorname{Re}\tilde{\chi}=\xi$ and $\operatorname{Im}\tilde{\chi}=\alpha\xi+\beta$, where ξ is the real parameter of the straight line, α is the slope of the line, and $\beta\neq 0$ for $\operatorname{Im}\chi_0\neq 0$. As a result, we have

$$\tilde{\kappa}_\pm(\xi)=\pm\left(\sqrt{\frac{\sqrt{a^2+b^2}+a}{2}}+i\,\operatorname{sign}b\sqrt{\frac{\sqrt{a^2+b^2}-a}{2}}\right), \tag{14}$$

where

$$a=\alpha\xi+\beta^2-\xi^2-\delta_1,\quad b=-\delta_2-2\xi(\alpha\xi+\beta),$$

$$\delta=\delta_1+i\delta_2,\quad \delta_1\delta_2\in\mathbb{R}^1.$$

2. The two-parameter (α,β) families of curves $\operatorname{Re}\tilde{\kappa}(\xi)$, $\operatorname{Im}\tilde{\kappa}(\xi)$ contain for certain values of α and β dependences that are similar to the well-known Wien plot for the resonant

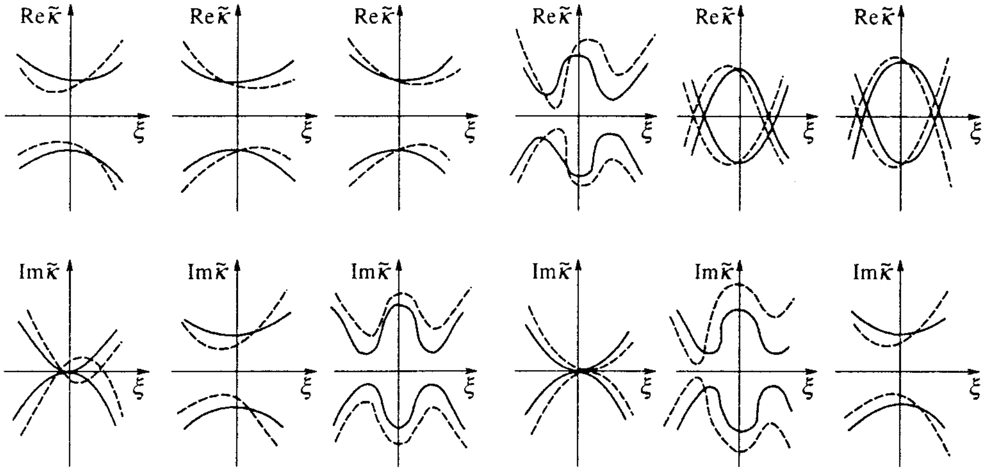


FIG. 1. Dispersion law of an open cavity near the Morse point for different values of the parameters α and β: Solid curves—$\beta=0$, dashed curves—$\beta\neq 0$.

frequencies of coupled oscillatory circuits. Figure 1 displays the dispersion curves obtained from Eqs. (13) and (14) for all possible qualitatively different dispersion laws near the Morse point, for various values of α and β (the functions $\tilde{\kappa}(\xi)$ with $\beta=0$ are plotted as solid lines; their perturbations ($\beta\neq 0$) are plotted as dashed lines).

Thus, the presence of an isolated Morse point (κ_0,χ_0) of the characteristic determinant $\mathscr{F}(\kappa,\chi)$ of an open cavity leads to the existence of two solutions $\kappa_+(\chi)$ and $\kappa_-(\chi)$ of the dispersion relation $\mathscr{F}(\kappa,\chi)=0$ near (κ_0,χ_0). The behavior of these solutions as a function of the detuning χ of the cavity is completely determined by the quadratic form (13) and describes hybrid oscillations of the cavity; specifically, it governs the rapid variation of the quality factor $Q=-\mathrm{Re}\ \kappa/2\ \mathrm{Im}\ \kappa$ (diffraction losses) of the cavity for small variations of its geometric parameters.

Figure 2 displays the real parts of the eigenfrequencies and the logarithm of the Q for H_{03} and H_{41} oscillations as a function of Re $\chi=b/a$. One can see that the Q of the oscillations has a pronounced dependence on b/a: as b/a increases, the Q of the (high-Q) H_{03} oscillation, which has three variations of the field along the cavity axis, grows monotonically to $Q\sim 10^4$ ($b/a=0.650$) and then rapidly decreases (by almost two orders of magnitude), and near $b/a=0.775$ it becomes comparable to the Q of the H_{41} oscillation, which increases rapidly with b/a. The real parts of the eigenfrequencies (Fig. 2a) for $0.73\leqslant b/a\leqslant 0.83$ form dispersion curves similar to a Wien plot. For $\chi_0=0.775$ and $\kappa_0=10.130$ the absolute difference of the eigenfrequencies is minimum (the Qs of the H_{03} and H_{41} oscillations are equal).

The structure of the fields of the H_{03} and H_{41} oscillations near the Morse point ($\kappa_0=10.130$ and $\chi_0=0.775$) is of great interest. To determine it, it is necessary to make use of the excitation of the open cavity by an H-polarized plane wave at frequencies corresponding to H_{03} and H_{41} oscillations, as well as the frequency of the oscillations at the Morse point itself. Figure 3 displays the distribution of the fields (lines of constant H_z components of the field) in an open cavity (the cases 1, 2, and 3 correspond to the frequencies 1, 2, and 3 in Fig. 2a).

Fundamental conclusions can be drawn from Figs. 2 and 3: near a Morse point, substantial changes occur in the behavior of the dispersion laws and intense hybrid oscillations arise. As we show below, this leads to a new situation in the study of nonlinear evolutionary processes near a Morse point.

3. Any dispersion relation obtained in electrodynamics, plasma physics, solid-state physics, and other fields must be investigated for the presence (or absence) of Morse points and the resultant feasibility of hybrid oscillations. Note that

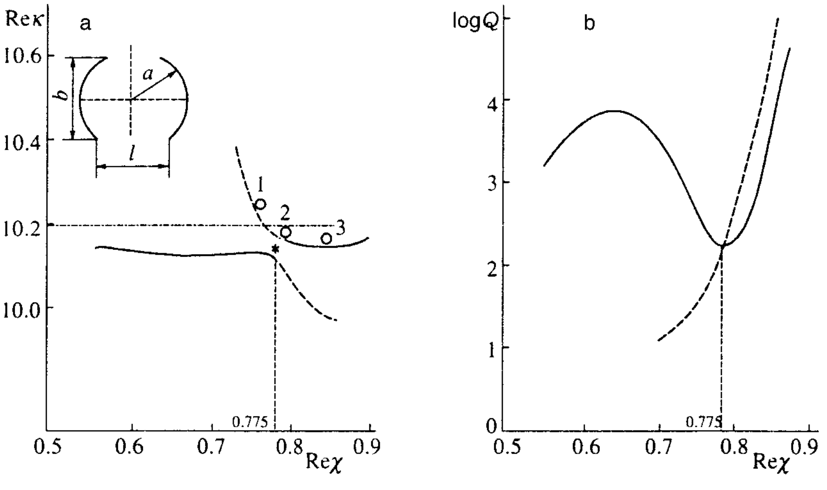


FIG. 2. Dispersion (a) and diffraction losses (b) in an open cavity near the Morse point (asterisk in Fig. 1a) of H_{03} oscillations (solid curves) and H_{41} oscillations (dashed curves).

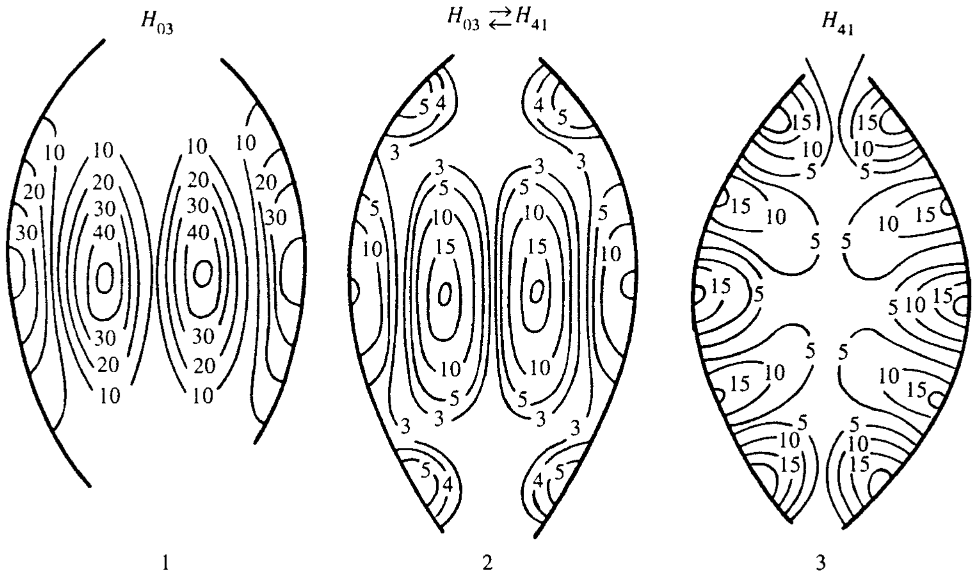


FIG. 3. Distribution of H_{03} and H_{41} oscillations of the field of an open cavity near the Morse point (2—hybrid oscillations $H_{03} \leftrightarrow H_{41}$ at the Morse point).

the method of investigating Morse points in which the function $\mathscr{F}(\kappa,\chi)$ determining the dispersion relation $\mathscr{F}(\kappa,\chi)=0$ depends on two complex variables κ and χ extends to the case in which three and more quantities are independent variables (for example, the frequency, the projection of the wave vector on one of the axes, and one of the control parameters).

4. EXPERIMENTAL INVESTIGATIONS OF A DIFFRACTED-RADIATION OSCILLATOR

A complicated open electrodynamic structure and the presence of a distributed electron flux lead to additional difficulties in the study of nonlinear evolutionary processes in diffracted-radiation oscillators. The experimental data used for constructing the theory of such oscillators must first confirm the fact that a Morse critical point exists in the open cavity of the oscillator and that near the Morse point the same dispersion dependences and hybrid oscillations are observed as in the simplest cavity studied in Secs. 2 and 3, and that new qualities of the process associated with the electron flux likewise appear.

1. We present experimental data on the behavior of an open cavity near a Morse point.[10] The measurements were performed on a standard experimental setup, described in detail in Ref. 3; it is illustrated schematically in Fig. 4. The cavity is formed by flat (with 60 mm aperture) and spherical (radius $a=60$ mm) mirrors. A diffraction grating with period $d=0.4$ and width 10 mm is positioned at the center of a mirror. The width and depth of the channels are $g=0.14$ mm and depth $h=0.86$ mm, respectively. The cavity is excited by a klystron in the 4-mm range through a coupling slot in the spherical mirror. The cavity is tuned by moving the flat mirror along the cavity axis. A signal is extracted from the cavity with the aid of a second slotted coupling element mounted on the spherical mirror. The distances between the mirrors is l.

According to the classification adopted, the fundamental modes of the cavity will be quasi-TEM_{20q}, where q is the number of half-waves between the mirrors. Hybrid quasi-TEM_{20q} and quasi-TEM_{40q} modes, where quasi-TEM_{40q} is the highest mode of the open cavity, were observed experimentally. It was established that for $\chi=l/a=0.324$ the quasi-TEM_{40q} mode (with high Q) dominates for large l; after this point the quasi-TEM_{40q} mode decays rapidly (Q drops), while the quasi-TEM_{20q} modes become high-Q. Clearly, we are dealing with the scheme of transformations of hybrid oscillations discovered theoretically for the simplest open cavity. Therefore, these variations make it possible to assert that a Morse critical point has been found at $\chi_0=0.324$. The curve of Q for quasi-TEM_{20q} (curve *1*) and quasi-TEM_{40q} (curve *2*) modes in Fig. 4 confirms the existence of a Morse point at $\chi_0=0.324$.

2. The dispersion curve of an open cavity of a diffracted-radiation oscillator near the Morse point ($\chi_0=0.324$) is displayed in Fig. 5. It was obtained for quasi-TEM_{20q} and quasi-TEM_{40q} oscillations by determining the resonant distance l for a specific value of q. The ratio l/a is chosen so that the resonant distance for quasi-TEM_{20q} and quasi-TEM_{40q} oscillations at constant q lie on both sides of the

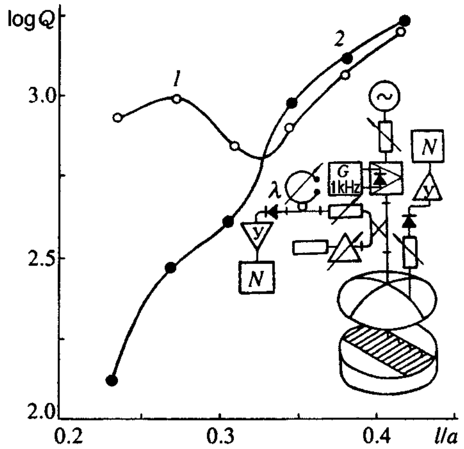


FIG. 4. Experimental diffraction losses in an open cavity of a diffraction-radiation oscillator for quasi-TEM_{20q} oscillations (curve *1*) and quasi-TEM_{40q} oscillations (curve *2*) and a block diagram of the experimental setup.

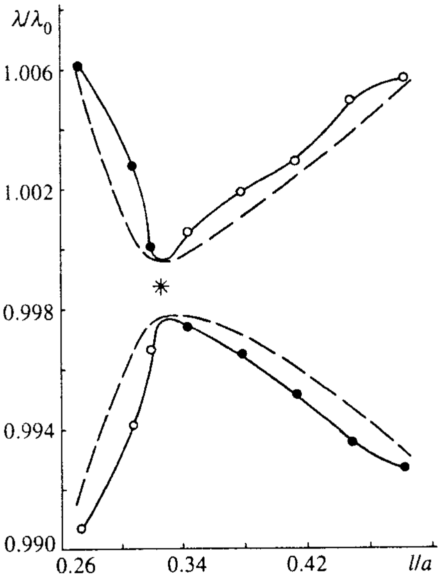


FIG. 5. Experimental dispersion curves (solid lines) of an open cavity of a diffracted-radiation oscillator (dashed curves—theory) near the Morse point (*) for quasi-TEM_{20q} (○) and quasi-TEM_{40q} (●) modes.

chosen value of χ_0. Then the wavelength λ of the oscillations under study is determined with a wavemeter. For $\chi_0=0.324$ the absolute difference of the characteristic wavelengths assumes its minimum value (Fig. 5), and the Qs of the modes (Fig. 4) become the same. After the Morse point the modes transform into one another and exchange energy; as l/a decreases further, the branches of the dispersion curves diverge. Therefore, measurements of the dispersion in an open cavity of the oscillator also confirm the existence of a Morse point with $\chi_0=0.324$. The dispersion curve constructed according to Eqs. (13) and (14) is shown in Fig. 5 (dashed lines).

3. The structure of the fields of hybrid modes was determined by the test-body method. It was found that far from the Morse point the modes are of a pure form: quasi-TEM_{4011} for $l/a=0.413$ and quasi-TEM_{2012} for $l/a=0.410$. As the distance between the cavity mirrors decreases, the quasi-TEM_{4010} mode ($l/a=0.377$) changes first—this is the first higher-order mode of an open cavity: it has a larger ''spot'' size at the mirror than the quasi-TEM_{2011} mode ($l/a=0.375$). As the Morse point is approached, the quasi-TEM_{409} mode ($l/a=0.342$) transforms into a quasi-TEM_{2010} mode ($l/a=0.341$); for this mode the field ''spot'' increases in size. Intense energy exchange between these modes starts. After passage through the Morse point, the mode types interchange. Thus, the quasi-TEM_{409} mode becomes a quasi-TEM_{209} ($l/a=0.307$) mode and the quasi-TEM_{2010} mode becomes a quasi-TEM_{408} ($l/a=0.304$) mode.

4. Our experimental investigations show that in an open cavity with a diffraction grating, an interaction exists between two types of modes as a function of the distance between the mirrors. Significantly, the Q of the working-type oscillations increases. This results in a higher efficiency of interaction of the electron flux with the field of the open cavity.

It can now be claimed that the theory of Morse critical points, which was developed for a very simple open cavity, also extends to the complicated cavity of a diffracted-radiation oscillator. This makes it possible to thoroughly investigate the construction of nonlinear evolutionary equations for such an oscillator.

5. NONLINEAR EVOLUTIONARY EQUATIONS OF A DIFFRACTIED-RADIATION OSCILLATOR

The results obtained in Secs. 2–4 concerning the properties of open electrodynamic structures near Morse critical points are the foundation for the construction of nonlinear evolutionary equations of a diffracted-radiation oscillator. However, they must be understood from the standpoint of nonlinear dynamics. Furthermore, it is also necessary to determine the effect of the electron flux on the excitation of oscillations in the oscillator and to establish the characteristic features of the continuous generation of the oscillations.

1. As is well known,[11] one nonlinear bistable element (a flip-flop) in nonlinear systems is governed by the nonlinear interaction of the elements of the system. The flip-flop possesses two stationary stable states in each of which it can reside forever, and only external actions can transfer it from one state to the other.

Let us examine from this standpoint the behavior of linear stationary oscillations in an open cavity near a Morse point. For this, we make use of Figs. 2, 3, and 5. As one can see from Fig. 2a (and Fig. 5), there exist two branches of the dispersion curves determined by the solution of the dispersion equation (13). Confining attention to the upper branch, it is evident that an anomalous dispersion law holds from point 1 to the Morse point 2 and a normal dispersion law holds from the Morse point 2 to point 3. Therefore at the Morse point the anomalous dispersion law is replaced by a normal dispersion law. As one can see from Fig. 3, this leads to a corresponding behavior of the linear oscillations in distributed systems: as the nonspectral parameter varies, there arises a chain of oscillations, two of which are stable and one unstable. The stable oscillations are determined by regular dispersion laws. The unstable oscillation is determined by the dispersion law near the Morse point.

As is standard in the theory of nonlinear oscillations, we associate the linear oscillations corresponding to the regular points of the spectrum with linear monostable oscillations. It is convenient to represent them in the form of a linear monostable element (''limel''); they correspond to points 1 and 3 in Fig. 2a and the distributions of fields in Fig. 3 which correspond to these points. It is evident from Fig. 2b that limel 1 is a high-Q H_{03} mode of a confocal open cavity (we call it a large limel), and limel 3 corresponds to a low-Q H_{41} mode (small limel). Limels are noninteracting elements; they can be real or complex; their theory has been developed in detail.

Near a Morse point the linear modes interact, forming a hybrid unstable mode, which as the nonspectral parameter is varied, turns into two stable (ordinarily high- and low-Q) modes. We define the mode near the Morse point to be a linear bistable element (''libel''). Near a Morse point a libel transforms into two limels according to the scheme large limel→hybrid libel→small limel.

In summary, in linear oscillatory distributed systems (just as in lumped systems), together with noninteracting (limels) there also exist interacting (libels) oscillations. Clearly, this nature of libels (and limels) is determined by the dispersion properties of a linear oscillatory system. In this respect a libel is different from a flip-flop, though for libels also the transition from one stationary state to the other occurs as a result of an external action; the bistable state of a libel is a purely dispersion phenomenon.

2. In studying nonlinear evolutionary processes and, specifically, in constructing nonlinear evolutionary equations for a diffracted-radiation oscillator, an important fact is that the generation of oscillations is closely related to the state of open electrodynamic structures near Morse points. The first step in constructing a nonlinear evolutionary equation involves the nonlinear nonstationary characteristic modes that can arise in an open cavity near a Morse point. To this end, it is necessary to make use of the dispersion equation (13), solved for $\widetilde{\kappa}$, and replace the dimensionless characteristic frequency by the time differentiation operator, i.e., $\widetilde{\kappa}\sim id/d\tau$, where τ is a dimensionless time related to the true time $t=(l/c)\tau-t_0$ (t_0 is the time required to cover the distance l between the cavity mirrors). As a result, we obtain linear evolutionary equations for an open cavity near a Morse point in the form

$$\frac{du}{d\tau}+\alpha(\widetilde{\chi})u=0,$$

where $\alpha(\widetilde{\chi})=\pm\sqrt{\widetilde{\chi}^2+\delta}$, and the ± signs correspond to modes on the upper and lower branches of the dispersion curves in Fig. 2a (or Fig. 5). The function $u(\tau)$ is a complex-valued, twice continuously differentiable function in $\mathbb{R}^2$ of the variable τ. It must also be kept in mind that for distributed structures, the quantity u—the intensity of the electromagnetic field— depends on the spatial coordinates as well in the case of a two-dimensional open cavity, as one can see from Fig. 3.

If it is assumed that α depends on u, i.e., $\alpha(u)$ is a real analytic function defined on the semiaxis $\mathbb{R}^1$, then its Taylor series expansion for small u about the origin has the form $\alpha(u)=\alpha_0+\alpha_1u^2+\ldots$ (the fact that the coefficient of the linear term equals zero follows from the analyticity of $\alpha(u)$), where $\alpha_0=\alpha_0(\widetilde{\chi})$ and $\alpha_1=\alpha_1(\widetilde{\chi})$. Now the nonlinear evolutionary equation near a Morse point for an open cavity has the form

$$\frac{du}{d\tau}=\alpha_0(\widetilde{\chi})u-\alpha_1(\widetilde{\chi})u^3. \tag{15}$$

Here we have confined our attention to nonlinear oscillations consistent with the dispersion described by the upper dispersion curve in Fig. 2a (for the lower curve the sign on the right-hand side must be flipped). The most important aspect of Eq. (15) is the dependence of α_0 and α_1 on the control parameter $\widetilde{\chi}$, which governs the tuning of the cavity. The question of the duration of the nonlinear processes near the Morse point is equally important. That duration is probably determined by the scheme large limel →hybrid libel→small limel, and can be established by studying more carefully the dispersion curves in Figs. 2 and 3 together with Fig. 5; this will make it possible to find the time of ''approach'' to the Morse point (point 1 on the dispersion curve in Fig. 2a) and the time of ''exit'' from the region of the Morse point (point 3 in Fig. 2a). The resulting dwell time of an oscillatory system near a Morse point will make it possible to determine the initial (and final) data for the desired function $u(\tau)$ from Eq. (15).

3. The next step in constructing a nonlinear evolutionary equation for a diffracted-radiation oscillator is to take account of the electron flux accompanying excitation of an open cavity and the continuous generation process. To this end, we examine qualitatively the behavior of an electron flux in an open electrodynamic structure of a diffracted-radiation oscillator.

The open system of an oscillator can be excited[3] by the diffracted radiation that results from the uniform rectilinear motion of an electron flux near a diffraction grating. For subsequent analysis, it is convenient to represent the diffraction grating as a periodic sequence of open waveguides, which in the simplest case is simply the segment of a slow-wave system covering part of the bottom mirror of the cavity. The dispersion relation for the system is

$$\frac{c}{v_{\mathrm{ph}}}=\sqrt{1+\frac{g^2}{d^2}\tan^2\left(\pi\frac{h}{\lambda_0}\right)},$$

where v_{ph} is the phase velocity of the surface waves; d and g are the period and width of the grating; h is the depth of the channels; and λ_0 is the wavelength in free space. The dispersion curve for this equation is similar to that presented in Fig. 2a; dispersion types are exchanged at the Morse point (i.e., as $v_{\mathrm{ph}}\to c$): anomalous dispersion is replaced by normal dispersion. For an electron flux with such dispersion, the mixing of electrons trapped by the wave is reminiscent of the motion of a liquid along a pipe with expanding diameter. At points 1 and 3 in Fig. 2a, when the energies are equal, the slow wave of the diffraction grating entrains a different number of electrons from the flux. If the indicated points 1 and 3 correspond to the same increment in the potential (i.e., a constant electric field μ is applied to the electron flux), then according to momentum conservation and the foregoing conditions, the change in the velocity of the electrons at point 1 in Fig. 2a will be greater than at point 3.

At the Morse point itself, a state of the nonlinear oscillatory process arises in which the two opposing types of dispersion occur. The parametric effect, accompanied by a frequency shift near the Morse point, comes into play (this phenomenon has been investigated experimentally[12]). Therefore, a constant electric field applied to the electron flux orders its motion, and also produces the prerequisities for the formation of diffraction radiation, which, on account of the parametric effect near a Morse point of an open electrodynamic system, is amplified and transformed into cw electromagnetic emission.

The foregoing qualitative arguments concerning the role of the electron flux in the operation of the oscillator can be represented phenomenologically by introducing a constant parameter μ into the nonlinear equation (15). The result is

$$\frac{du}{d\tau}=\alpha_0(\tilde{\chi})u-\alpha_1(\tilde{\chi})u^3+\mu. \qquad (16)$$

The presence of the constant coefficient μ in Eq. (16), which is the second control parameter of the oscillator, is fundamental, since Eq. (16) differs radically from Eq. (15), if for no other reason than that Eq. (15) is a homogeneous nonlinear evolutionary equation and Eq. (16) is an inhomogeneous equation.

6. QUALITATIVE ANALYSIS OF NONLINEAR EVOLUTIONARY EQUATIONS

Let us examine Eq. (16) from the standpoint of its structural stability. In this regard, we note once again that the electrodynamic structure of the diffracted-radiation oscillator is open, i.e., it is a dissipative system. Energy is dissipated by an open system not as a result of thermal losses, but as a result of purely diffractive losses. In analyzing Eq. (16), bifurcations can therefore be studied on the basis of a simple dissipative system. It must also be borne in mind that Eq. (15)—and therefore Eq. (16)—differ from the known equations,[13] in that the control parameters depend on quantities that already appeared in the solution of linear spectral problems, and that control the tuning of the cavity (the quantity $\chi=l/a$).

1. We first qualitatively consider the nonlinear evolutionary equation (15). In this equation the rate of change of u is a linear function of the parameter $\gamma=\alpha_0(\tilde{\chi})\alpha_1^{-1}(\tilde{\chi})$. The stationary state of Eq. (15) can be determined from the equation $-u_s^3+\gamma u_s=0$, the trivial solution of which is $u_0=0$; the two nontrivial solutions are $u_\pm=\pm\sqrt{\gamma}$ ($u_\pm$ for which $\gamma>0$ will also be physically meaningful). The solutions $u_\pm$ merge with u_0 at $\gamma=0$, and branch from it for $\gamma>0$.

This is the well-known phenomenon of bifurcation. The stability of oscillations in an open system can be studied completely, bearing in mind that Eq. (15) admits an exact solution; depending on the sign of γ, the solution u_0 is globally asymptotically stable (for $\gamma<0$) or unstable (for $\gamma>0$). The solutions $u_\pm$ are simply asymptotically stable. Clearly, $u_\pm$ appear as a result of bifurcation at the moment when the state $u_0=0$ becomes unstable, although the solutions $u_\pm$ themselves are stable. Near the bifurcation point $\gamma=0$, the continuous dynamics result in singularities of the solution $u_\pm$, i.e., $u_\pm$ are not analytic functions of γ, which is what determines the qualitative, bifurcation-induced change in the behavior of the open system of a diffracted-radiation oscillator.

These arguments follow from an analysis of classical bifurcation.[13] In the the nonlinear dynamics studied here, the problem is complicated by the fact that the classical control parameter γ is itself controlled by $\tilde{\chi}$, i.e., $\gamma=\gamma(\tilde{\chi})$. The qualitative analysis of bifurcations in the open system of an oscillator must therefore take account of the fact that in Eq. (15) γ depends on $\tilde{\chi}$, and this relation must be tied to the behavior of the system under the conditions of a linear stationary process.

We should then bear in mind that bifurcations themselves are manifest in the dispersion behavior near the Morse point. This means that the character of the solutions $u_\pm(\gamma,\tau)$ is determined not only by temporal evolution ($\tilde{\kappa}\sim i d/d\tau$, $\tilde{\kappa}=\pm\sqrt{\tilde{\chi}^2+\delta}$), but also by $\tilde{\chi}$. In this regard, it is necessary to turn once again to Figs. 2a and 3, which can aid in establishing the initial conditions through which $\alpha_0(\tilde{\chi})$ and $\alpha_1(\tilde{\chi})$, and hence γ, can be established for the nonlinear process under study. The nonlinear equation (15) can then be investigated further. We merely note that on the basis of a study of the free nonlinear oscillations in the open system of the oscillator, according to Eq. (15), the role of limels and libels becomes clear: they constitute a linear physical ‘‘prehistory’’ of the bifurcations that emerge, and to some extent they predetermine them.

2. We now examine Eq. (16). In contrast to Eq. (15), in Eq. (16) we have a problem concerning a nonlinear process with two control parameters. The conditions of stationarity of Eq. (16) are given by the cubic equation

$$-u_s^3+\gamma u_s+\overline{\mu}=0, \quad \overline{\mu}=\mu\alpha_1^{-1}(\tilde{\chi}).$$

The roots of this equation are

$$u_{s1}=p+q, \quad u_{s2,3}=-\frac{1}{2}(p+q)\pm\frac{\sqrt{3}}{2}(p-q),$$

where

$$p=\left(-\frac{1}{2}\overline{\mu}+\sqrt{\frac{1}{4}\overline{\mu}^2+\frac{1}{9}\gamma^3}\right)^{1/3},$$

$$q=\left(\frac{1}{2}\overline{\mu}-\sqrt{\frac{1}{4}\overline{\mu}^2+\frac{1}{9}\gamma^3}\right)^{1/3};$$

any value of the root can be taken for p, and the value of the root for which $pq=-\gamma/3$ is then taken for q. For us, it is important that the cubic equation possesses three real solutions, and we can try to merge three solutions into one real solution by adjusting γ and $\overline{\mu}$. The relationships among the two-parameter regimes of the nonlinear evolutionary process can be determined with the aid of the equation $4\gamma^3+27\overline{\mu}^2=0$, which implies the existence of three real roots of the cubic equation that end up at the point $\gamma=\overline{\mu}=0$, where the dependence of γ on $\overline{\mu}$ has a cusp-type singularity.

The dependence of u_s on $\overline{\mu}$ for fixed γ can be traced out. As is well known, this leads to an S-shaped curve related to the multiplicity of solutions as a function of the parameter $\overline{\mu}$. Two branches are stable simultaneously. The bistability arising near the points $\overline{\mu}_1$ and $\overline{\mu}_2$ causes hysteresis.

The function $u_s(\gamma)$ with fixed $\overline{\mu}$ yields two curves, one of which has a limit-point singularity. Correspondingly, stable and bistable states of the oscillatory process in an open electrodynamic system of the diffracted-radiation oscillator are obtained.

It is important that while Eq. (15) determines a symmetric bifurcation, Eq. (16) yields no such bifurcation for any $\overline{\mu}$, however small. Therefore $\overline{\mu}$ plays the role of a ‘‘destroyer’’ of symmetric bifurcation.

The arguments presented above for Eq. (16) are well known,[13] with the exception of the fact that the two control parameters in Eq. (16) themselves depend on the control parameter $\tilde{\chi}$. Results concerning the efficiency of a diffracted-radiation oscillator therefore relate to precisely this problem.

7. CHARACTERISTIC FEATURES OF THE NONLINEAR OPERATING REGIME OF A DIFFRACTED-RADIATION OSCILLATOR

The ''historical'' development of coherent cw emission by an oscillator involves three successive stages: the formation of a electronic noise cloud as electrons are extracted from a hot cathode; conversion of the cloud in the near-cathode region into a chaotic flux of particles; and, finally, conversion of these particles with the aid of positive feedback (i.e., an open electrodynamic system) and an applied constant electric field into a highly organized assemblage of particles, which produce diffracted radiation that feeds the open cavity of the oscillator. The analysis of Eq. (16) concerns the latter stage.

1. To understand the role of Eq. (16) in the study of nonlinear processes in an oscillator, it is necessary to consider the behavior of the electron flux in the interaction space of an open electrodynamic system. Above all, it must be borne in mind that at millimeter and especially submillimeter wavelengths, the electron flux must be flat and very thin, and its density must be high. On account of Coulomb repulsion, space-charge waves, and other reasons, strong fluctuations of electrons arise in such a flux, the electrons precipitate onto the diffraction grating, and so on. All this makes it necessary to employ, together with the ''pulling'' constant field applied to the electron flux, a constant longitudinal magnetic field that focuses the flux.

Nonetheless, the electrons must pass near the diffraction grating (in a real oscillator working at $\lambda = 2$ mm, the gap between the grating and the electron flux is ~ 0.1 mm). Since the current is ~ 0.1 A and the applied voltage is 3–5 kV, the thermal power dissipated by the grating (the total dissipation area is $\sim 1.5\ \mathrm{cm}^2$) is ~ 500 W. For the device as a whole, this is a very difficult thermal regime. Forced (water) cooling of the oscillator is required. Even so, thermal fluctuations at the grating and the bottom mirror of the oscillator are such that at short millimeter and especially submillimeter wavelengths they lead to appreciable mechanical fluctuations in the body of the operating device, which ultimately influence the quantity $\chi = l/a$ and therefore γ and $\bar{\mu}$ in Eq. (16).

Therefore, even in a stationary state of the oscillator, the amplitude and frequency of the oscillations vary continuously as a result of many internal and external destabilizing factors, whose effect on the microwave signal of the oscillator is all the weaker, the more wideband the open systems used in the oscillator are. Ordinarily, the fluctuation parameters of the oscillator are determined by analyzing the excitation of bulk ''noise'' oscillations and their action on the electron flux. Noise waves are studied as a result of the changes in all of the above-indicated parameters of the oscillator. The problem consists of establishing the relationship between the indicated processes and the structural stability of the dynamic system, determined by Eq. (16).

2. We have already noted that in Eq. (16) the role of the parameter $\bar{\mu}$, even for $\bar{\mu} = \mathrm{const}$, reduces to destruction of the symmetric bifurcation diagram. In this case, the indistinguishable stable branches $u_\pm$ of the solution of Eq. (15) $(\bar{\mu} = 0)$ with $\bar{\mu} \neq 0$ become separated, and as γ increases, the system remains on the branch u_+, provided that a fluctuation which would transform it into the state u_- does not appear. Thus, in Eq. (16) with $\bar{\mu} \neq 0$ and γ varying between u_+ and u_-, there appears a gap $\Delta(\bar{\mu})$ whose nature is determined by $\bar{\mu}$. For a nonlinear dynamic system, the quantity $\Delta(\bar{\mu})$, originating from Eq. (16), makes it possible to describe the behavior of the system near the bifurcation point.[13]

To what extent is the oscillator ''sensitive'' in the presence of the above-indicated fluctuations to a change in $\bar{\mu}$, i.e., $\Delta(\bar{\mu})$? As shown in Ref. 13, for small fluctuations we have $\Delta(\bar{\mu}) = 3(4\bar{\mu})^{1/3}/2$ for $\bar{\mu} = \mathrm{const}$. The fact that we have $\Delta \sim \bar{\mu}^{1/3}$ is decisive, since the electric field applied to the electron flux can fluctuate by $\bar{\mu} \sim 10^{-6}$, which yields $\Delta \sim 10^{-2}$, and therefore in Eq. (16) this quantity is capable of competing with the macroscopic phenomena, i.e., it can be comparable to γ in the gap Δ, i.e., γ_Δ. In Ref. 13 it is established that $\gamma_\Delta = 3(\bar{\mu}^2/4)^{1/3}$ (γ_Δ is the minimum distance between the solutions u_+ and u_- of Eq. (15)), and this quantity yields a difference between u_+ and u_- for $\gamma = 0$, which corresponds to a bifurcation.

It is known experimentally from a 4-mm diffracted-radiation oscillator that spurious modulation of the output power and the emitted spectrum arise in the 1–100 kHz band. These distortions are due to all the fluctuations mentioned above. Significantly, in this case, $\gamma_\Delta \sim 10^{-2}$.

Thus, in Eq. (16) for $-u_s^3 + \gamma u_s + \bar{\mu} = 0$ it can be assumed that in actually operating oscillators γ and $\bar{\mu}$ have reasonable orders of magnitude, making it possible to establish stable operation of the device. These arguments are all valid for the dispersion region far from a Morse point (Figs. 2a or 5 far from the points 1 and 3).

However, the main feature of the theory developed here is that the analysis of Eq. (16) in application to a diffracted-radiation oscillator must be conducted keeping in mind that γ and $\bar{\mu}$ are functions of $\tilde{\chi}$. It is obvious that everything that we have said for the case when γ and $\bar{\mu}$ do not depend on $\tilde{\chi}$ can be tied to the operation of the oscillator with $\tilde{\chi} = \mathrm{const}$. If, however, $\tilde{\chi}$ varies as a result of diverse fluctuations in an open system and the electron flux, then the analysis of Eq. (16) becomes much more complicated. The difficulty is somewhat eased by the fact that a preliminary analysis of the operation of the oscillator can be performed using the results of Secs. 1 and 2 associated with linear stationary processes in the device. Equation (16), however, must substantiate the behavior of the open system and electron flux under conditions when the cavity exhibits high-Q oscillations, which are displayed in Fig. 3 (case 1), or low-Q oscillations (see Fig. 3, case 3) with the existence of hybrid oscillations (Fig. 3, case 2). This means that in the scheme large limel→hybrid libel→small limel, which reflects the linear stationary process in the cavity near the Morse point, Eq. (16) should confirm that millimeter and submillimeter diffracted-radiation oscillators operate in the large limel regime. Equation (16) itself reflects the ''struggle'' between large and small limels (via a hybrid libel), and shows that the large limel prevails.

Thus, for correctly chosen γ and $\bar{\mu}$, which depend on

$\widetilde{\chi}$, the oscillations arising in a nonlinear process near bifurcation are globally asymptotically stable. This ensures high short-time stability (up to 10^{-10}) of oscillator operation, a low level of phase noise, and small parasitic deviation of the frequency. Implementation of certain frequency-locking and power-stabilization methods makes it possible to achieve long-term stability of the oscillator frequency of up to 5×10^{-10}, and changes in the output signal power can be reduced to $\leqslant 1\%$ (at 4 mm, this stability remains for several hours of operation at a level of continuous output power of the oscillator of up to 10 W). These qualities of diffracted-radiation oscillatros make them useful in the development of ultrahigh-resolution spectrometers, in the diagnostics of thermonuclear plasma, dielectric measurements in absorbing objects, spectroscopy of polarized nuclear targets, high-coherence radar, and other applications.

8. CONCLUSIONS

The approach developed here for studying diffracted-radiation oscillators transfers to other microwave devices in vacuum electronics: backward-wave tubes, klystrons, magnetrons, cyclotron-resonance masers, and free-electron lasers. The scheme for constructing the nonlinear evolution equations (16) and (15) for these devices is the same as for diffracted-radiation oscillators: solution of the linear spectral problem for a given structure; search for the Morse critical point; construction of the dispersion relation near the Morse point; construction and analysis of a nonlinear evolutionary equation; reconciliation with experiment. The difference will lie in the search for the control parameters. The minimum number of parameters, just as in a diffracted-radiation oscillator, is two: one controls the structure and the other controls the electron flux. Just as in a diffracted-radiation oscillator, these parameters themselves depend on the properties of the structure and the electron flux.

The method proposed for studying nonlinear evolutionary equations can be extended to physical, chemical, biological, and other areas of investigation of complex nonlinear processes.

This work was supported by the Soros International Educational Program in the Exact Sciences (ISSEP) under Grant No. EPU052028.

[1] V. P. Shestopalov, Dokl. Akad. Nauk SSSR **261**, No. 5 (1981) [Sov. Phys. Dokl. **26**, 1149 (1981)].

[2] V. G. Kurin, B. K. Skrynnik, and V. P. Shestopalov, Zh. Éksp. Teor. Fiz. **110**, 450 (1996) [JETP **83**, 240 (1996)].

[3] V. P. Shestopalov, *Diffraction Electronics* [in Russian], Khar'kov University Press, Khar'kov (1976).

[4] V. P. Shestopalov, *Spectral Theory and Excitation of Open Structures* [in Russian], Naukova dumka, Kiev (1987).

[5] V. P. Shestopalov, *The Riemann–Hilbert Method in Diffraction Theory* [in Russian], Khar'kov University Press, Khar'kov (1971).

[6] V. P. Shestopalov, *Morse Critical Points of Dispersion Equations* [in Russian], Naukova dumka, Kiev (1992).

[7] B. V. Shabat, *Introduction to Complex Analysis. Part II. Functions of Several Variables*, American Mathematical Society, Providence, RI (1992) [Russian original, Nauka, Moscow (1976)].

[8] V. I. Arnol'd, A. N. Varchenko, and S. M. Gusein-Zade, *Singularities of Differentiable Maps*, Birkhauser, Boston (1985) [Russian original, Nauka, Moscow (1982)].

[9] M. Morse, Trans. Am. Math. Soc. **33**, 72 (1931).

[10] I. K. Kuz'michev and V. P. Shestopalov, Dokl. NAN Ukrainy **12**, 87 (1996).

[11] Yu. A. Loskutov and A. S. Mikhaĭlov, *Introduction to Synergetics* [in Russian], Nauka, Moscow (1990).

[12] V. G. Kurin, B. K. Skrynnik, and V. P. Shestopalov, Pis'ma Zh. Tekh. Fiz. **21**, 5 (1995) [*sic*].

[13] I. Prigogine and G. Nicolis, *Exploring Complexity*, W. H. Freeman, San Francisco (1995).

Translated by M. E. Alferieff

Erratum: Realization of a heavily doped and fully compensated semiconductor state in a crystalline semiconductor with a deep impurity band [JETP 84, 309–316 (February 1997)]

M. I. Daunov, I. K. Kamilov, and A. B. Magomedov

Institute of Physics, Dagestan Science Center, Russian Academy of Sciencees, 367003 Makhachkala, Russia

[S1063-7761(97)02406-2]

The following corrections were reported by the authors. On page 309 in the right-hand column, in the 10th line from the bottom, the equation should read as follows: $N_d{}^+ = N_{d\mu}{}^+ + N_{dr}{}^+$. On the same page and the same column, in the 9th line from the bottom, the equation should read as follows: $N_a = N_{ar}{}^- + N_{a\mu}{}^- + N_A$. Also on the same page and the same column, in the 2nd line from the bottom, the equation should read as follows: $N = N_d{}^+ - (N_{a\mu}{}^- + N_{ar}{}^-)$.

In the first part of Table II, in the 11th column, in the second line from the bottom, the number should read as follows: 0.02.

1063-7761/97/061236-01$10.00