Quasiclassical solutions of the Klein–Gordon equation in a space–time with closed time-like curves

A. Yu. Neronov*)

M. V. Lomonosov Moscow State University, 119899 Moscow, Russia
(Submitted 9 April 1997)
Zh. Éksp. Teor. Fiz. **113**, 3–11 (January 1998)

A formulation of the Cauchy problem for the Klein–Gordon equation in a space–time with a wormhole is studied. It is shown that if closed time-like curves pass through the Cauchy surface, then a global solution exists in the high-frequency (quasiclassical) limit only if the initial data satisfy nonlocal consistency conditions. In the simplest case these conditions determine a discrete spectrum of frequencies (energies of a quasiclassical scalar particle). © *1998 American Institute of Physics.* [S1063-7761(98)00101-2]

1. INTRODUCTION

The behavior of classical and quantum fields against a background of solutions containing closed time-like curves of the Einstein equations has been investigated in a number of works[1,2] in connection with the possibility of the existence of space–time with a time machine, as proposed in Refs. 3–5, as well as with the problem of the self-consistency of physical laws in such a space–time.[6,7] The global solutions of the dynamical equations must satisfy the ''self-consistency principle,'' which eliminates paradoxes associated with the breakdown of causality in a space–time with closed time-like curves.

The nontrivial topology of space–time as well as the existence of closed time-like curves passing through the Cauchy surface impose constraints on the initial data for solving the Cauchy problem for the Klein–Gordon equation. This is because any solution determined in a neighborhood of the Cauchy surface can be continued to a global solution in all space–time.

The situation can be analyzed completely for a definite class of solutions of the Klein–Gordon equation, specifically, for high-frequency solutions. In this case there exists a technique, developed by Maslov (see, for example, Ref. 8), for constructing solutions in the large that makes it possible to construct global solutions over geometric objects (Lagrangian submanifolds) given in the phase space of a Hamiltonian dynamical system which is naturally associated with a partial differential equation.

We take as a model space–time the solution of Einstein's equations with a wormhole.[3] Asymptotically flat regions are sewn together so that the wormhole serves as a ''perpetual'' time machine. An identical space–time was used in Ref. 9 to analyze the behavior of a scalar field.

Global high-frequency solutions of the Klein–Gordon equation that are constructed by the Maslov method satisfy nonlocal conditions on the Cauchy surface expressing the correlation of the phases of the solution at different points on the surface. In the case when the solution is concentrated near a closed trajectory of a classical Hamiltonian system (the situation corresponding in the present case to the causal paradox), the solution can have only a discrete set of frequencies. This result is presented in Sec. 4. Section 2 contains the formulation of the problem and Sec. 3 contains information on the Maslov method.

2. FORMULATION OF THE PROBLEM

We are studying the Cauchy problem for a scalar field Φ

$$\left(g^{\mu\nu}\partial_\mu\partial_\nu - \frac{m^2}{\hbar^2} \right)\Phi(x)=0, \tag{1}$$

where $\hbar$ is Planck's constant (we choose units of measurement such that the speed of light satisfies $c=1$). We are studying the case when closed time-like curves pass through a three-dimensional Cauchy surface Σ^3 in a space–time M^4.

Let us consider as a model space–time containing closed time-like curves a solution of Einstein's equations that contains a wormhole.[1] We choose on the cylinder $R^2\times S^1$ a metric of the wormhole type[3] (Fig. 1), for example,

$$ds^2=-dt^2+dl^2+(b_0^2+l^2)(d\theta^2+\sin^2\theta d\phi^2). \tag{2}$$

Near $l\gg b_0$ and $l\ll -b_0$ there exist two asymptotically flat regions with the coordinates (X_+,Y_+,Z_+,T_+) and (X_-,Y_-,Z_-,T_-) where the metric is virtually identical to the Minkowski metric. We prescribe the gluing functions for the coordinate maps (X_+,Y_+,Z_+,T_+) and (X_-,Y_-,Z_-,T_-) as follows:

$$T_+=T_-+\delta T,\quad X_+=X_-,\quad Y_+=Y_-,$$
$$Z_+=Z_-+\delta Z. \tag{3}$$

Closed time-like curves start at some point (X_0,Y_0,Z_0,T_0), enter one mouth of the wormhole at time T and leave the other mouth at time $T-\Delta T$, preceding the moment T_0, and then connect up with the point (X_0,Y_0,Z_0,T_0) of the time-like curve directed into the future. The metric (2) is almost flat everywhere except near the mouths of the wormhole. If the distance between the mouths is much larger than the characteristic length of the handle [b_0 for metric (2)], then the following model can be used as a first approximation. Two hollow cylinders C_1 and C_2 of the form $R^1\times B^3$ (B^3 is a three-dimensional sphere) (see Fig. 2a), whose axes are also world lines of the centers of the mouths $X=0$, $Y=0$, $Z=0$ and $X=0$, $Y=0$, and $Z=\delta Z$, are cut out of Minkowski space–time. Then the edges of the cuts are glued together so

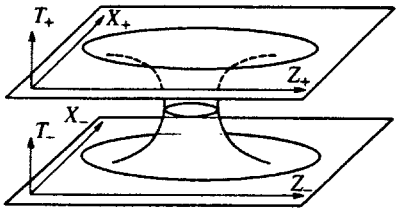


FIG. 1.

that the sphere $S^2=\{(X,Y,Z,T)\in C_1,\ T=T_0\}$ is glued to the sphere $S^2=\{(X,Y,Z,T)\in C_2,\ T=T_0+\delta T\}$.

We choose the Cauchy surface in such a space–time as follows. Let us consider a space-like section $T=T_0$. There exist time-like curves that do not cross this surface. They enter the mouth of the wormhole at time $T<T_0$ and exit at time $T>T_0$ (as, for example, the curve α does in Fig. 2a). Augmenting the surface $T=T_0$ with part of the cylinder C_1: $(X,Y,Z,T)\in C_1$:$T_0-\delta T<T<T_0$, we obtain the surface Σ^3 through which any (complete) time-like curve passes at least once. We choose Σ^3 as the Cauchy surface. Such a Cauchy surface was studied in Ref. 1 .

Since closed time-like curves pass through the surface Σ^3, the Cauchy data for a scalar field cannot be chosen arbitrarily. We shall analyze this effect for the example of high-frequency solutions of the Klein–Gordon equation.

It is natural to study the high-frequency asymptotic behavior of the solutions of the Klein–Gordon equation as a quasiclassical approximation of the quantum mechanics of a scalar relativistic particle. Indeed, Eq. (1) can be interpreted as an equation for the wavefunction $\Phi(x)$ of a scalar particle. The solution is chosen in the form of a rapidly oscillating phase term multiplied by a slowly oscillating amplitude

$$\Phi(x)=\exp\left[\frac{iS(x)}{\hbar}\right]\sum_{k=0}^{\infty}\hbar^k\Psi_k, \tag{4}$$

where $\hbar\to 0$ is a small parameter.

The metric (2) is stationary. Therefore the solution can be expanded in modes with constant frequency $\Phi=e^{i\omega t}\Psi$. In studying the high-frequency asymptotic behavior, we shall confine ourselves to the modes for which $\omega\to\infty$ as $\hbar\to 0$.

3. THE MASLOV METHOD OF CONSTRUCTING HIGH-FREQUENCY SOLUTIONS

The Maslov method[8] makes it possible to construct high-frequency solutions of Eq. (1) on the manifold M^4 in the whole based on analysis of the geometry of a classical Hamiltonian system corresponding to a partial differential equation. We present below the required information about this method.

The phase space of the Hamiltonian system corresponding to Eq. (1) is a cotangent bundle T^*M to the space–time manifold M^4 with the standard simplectic form $\Omega=dp_\mu\wedge dx^\mu$. The Hamiltonian has the form

$$H=g^{\mu\nu}(x)p_\mu p_\nu. \tag{5}$$

Substituting a function of the form (4) into Eq. (1) gives the Hamilton–Jacobi equation to zeroth order in $\hbar$:

$$H(x,\partial S/\partial x)=-m^2.$$

The function $S(x)$ prescribes in the phase space T^*M a four-dimensional surface according to the equation $\Lambda^4=\{(x,\ p(x)=\partial S/\partial x)\}$. This surface possesses the property that the simplectic form Ω vanishes when restricted to the surface Σ^4. The four-dimensional surfaces on which Ω vanishes in an eight-dimensional phase space are said to be Lagrangian. The Lagrangian surface Λ^4 determining the solution of the Hamilton–Jacobi equation can be constructed as follows. It is necessary to prescribe the initial values of $S(x)$ and $\partial S/\partial x^0$ for $x\in\Sigma^3$. This determines the surface $\Lambda^3=\{(x,p):\ x\in\Sigma^3,\ p=\partial S(x)/\partial x\}$. Next, extending from the points of Λ^3 trajectories of the Hamiltonian system (5) we obtain a four-dimensional surface Λ^4. In some neighborhood of the initial surface Σ^3 the surface Λ^4 can be uniquely projected onto $R^4=\{x\}$, i.e., $\Lambda^4=\{(x,\ p(x)=\partial S/\partial x\}$ for some function $S(x)$. However, Λ^4 does not necessarily project everywhere uniquely on $R^4=\{x\}$. Points where a unique projection breaks down form a cycle of singularities of the projection of the manifold Λ^4, and their projections onto $R^4=\{x\}$ are caustics of the trajectories of the dynamical system with Hamiltonian (5). Thus, generally speaking, the solution of the Klein–Gordon equation cannot be chosen everywhere in the form (4) (just as the quasiclassical expansion is valid far from turning points). Nonetheless, the solution can be continued beyond the caustic by constructing the Maslov operator. The solution beyond the points of the caustic is determined by the same Lagrangian submanifold Λ^4.

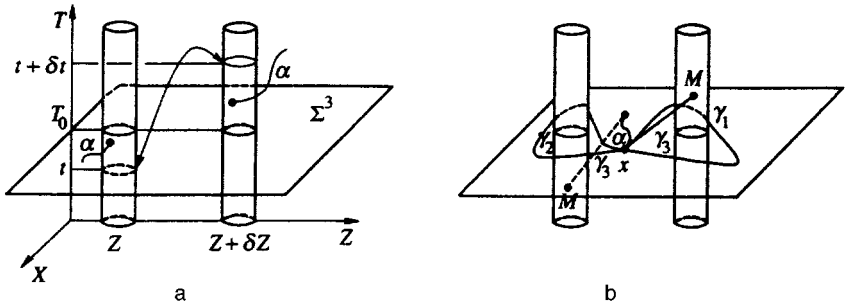


FIG. 2.

To construct the solution as a whole the global structure of the submanifold Λ^4 is important; this structure can be quite complicated. For example, closed curves that cannot be contracted into a point while remaining on the given Lagrangian manifold (nontrivial cycles) can exist on this submanifold.

To first order in $\hbar$ there arises for the function $\Psi_0(x)$ an equation called the transfer equation. In the region where Λ^4 projects uniquely onto $R^4=\{x\}$, the function $\Psi_0(x)$ can be regarded as a function on the submanifold Λ^4. Let

$$\tilde{\Psi}_0(y)=\sqrt{\frac{\partial y}{\partial x}}\Psi_0[x(y)],$$

where y are coordinates on the Lagrangian submanifold Λ^4. The transfer equation has the form $d\tilde{\Psi}_0/\partial\tau=0$, the condition that the function $\tilde{\Psi}_0$ is constant along the trajectory of the system (5) (τ is the time along trajectories in phase space). The expansion (4) is invalid at points in a neighborhood of and beyond caustics. However, it is possible to construct a function Ψ_0 that is constant along the trajectories of the system on the entire Lagrangian manifold Λ^4.

To obtain the solution as a whole, it is necessary to construct a special (canonical) atlas of maps of the Lagrangian submanifold Λ^4. Part of the solution of the Klein–Gordon equation is determined with the aid of the Maslov operator on the function Ψ_0 on Λ^4. For us it is important that in matching different parts of the solution obtained in this manner restrictions arise on the choice of the Lagrangian submanifold Λ^4. Specifically, the integral condition

$$\frac{1}{2\pi\hbar}\oint_\gamma p_a dx^a-\frac{1}{4}\,\mathrm{ind}\,\gamma=k, \tag{6}$$

which is called the quantization condition, must hold on each nontrivial cycle γ on Λ^4. Here ind γ is the index of the curve γ. It is calculated with the corresponding signs of the point of interaction of the curve γ with the cycle of singularities of the projection of the submanifold Λ^4. If on each nontrivial cycle of Λ^4 the condition (6) holds, then it is possible to construct an approximate (to different orders in $\hbar$) solution of the Klein–Gordon equation. The conditions (6) are multidimensional analogs of the Bohr–Sommerfeld quantization conditions. The Lagrangian submanifold satisfying these conditions is called a quantum submanifold.

We shall not present the explicit form of the Maslov operator, mapping the function $\tilde{\Psi}_0$ on the quantized Lagrangian submanifold Λ^4 into a global approximate solution of the Klein–Gordon equation on the manifold M^4 (see Ref. 8). We are interested only in the conditions for the existence of such a global solution, given by the expression (6).

4. FORMULATION OF THE CAUCHY PROBLEM IN A SPACE–TIME WITH CLOSED TIME-LIKE CURVES

The solution of the Cauchy problem in the high-frequency limit reduces to finding an appropriate initial submanifold Λ^3 in the phase space Π, projected onto the Cauchy surface Σ^3 in space–time. Extending integral trajectories of the Hamiltonian of a vector field with Hamiltonian (5) from the points Λ^3, we obtain a four-dimensional invariant (with respect to the Hamiltonian flux) Lagrangian submanifold Λ^4. Next, it is necessary to check that the quantization conditions (6) hold on the basic nontrivial cycles of this submanifold. We thereby limit the choice of Λ^4 and therefore the initial submanifold Λ^3. This means that we obtain restrictions on the choice of the initial data on the Cauchy surface Σ^3.

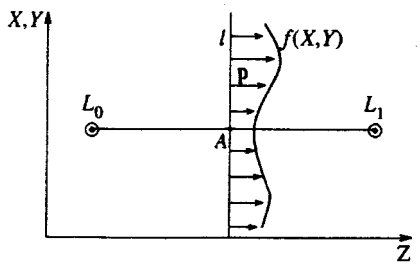


FIG. 3.

Let us consider the situation when the trajectories of the classical system enter the right-hand entrance into a wormhole (see Fig. 2b). Then, leaving the left-hand entrance at time T, preceding the time T_0 at which the Cauchy surface Σ^3 is given, the trajectory necessarily intersects Σ^3 once again. The first simple requirement of self-consistency of the initial data is that this trajectory must return to the initial surface Λ^3. Thus, the initial submanifold Λ^3 cannot be arbitrary, as happened in the absence of closed time-like curves.

Let us examine a natural method for ''matching'' the initial surface Λ^3 that makes it possible to obtain a consistent, in the sense indicated above, surface Λ^3. We start with an arbitrary surface $\Lambda_0^3=\{(x,p(x)=\partial S/\partial x, x\in\Sigma^3\}$. It is not, generally speaking, self-consistent. Let us construct the trajectories of the Hamiltonian system (5) and study the trajectories that pass through the wormhole and cross the surface $Q=\{(x,p)\colon x\in\Sigma^3\}$ at some time along the trajectory. This behavior of the trajectories determines a Poincaré map on the surface Q in which Λ_0^3 is embedded. The images $[x(\tau),p(\tau)]$ do not, generally speaking, belong to Λ_0^3. This means that there exists in Q another component $V_1^3=\Lambda^4\cap Q$ that is different from Λ_0^3. We can ''match'' the initial data for Σ^3 at this step, determining a new initial submanifold $\Lambda_1^3=\Lambda_0^3\cup V_1^3$. Repeating the above procedure for a new initial submanifold, we obtain a new correction of the initial data Λ_2^3, and so on. In the limit this procedure yields a consistent initial submanifold Λ_∞^3.

As an illustration of the fact that the curved submanifold Λ_∞^3 can be quite complicated, let us examine the example illustrated in Fig. 3. We are studying a massless scalar field. For it $p_0^2=|\mathbf{p}|^2$, so that only the spatial part of the momentum can be shown. Let the direction of $\mathbf{p}$ be the same as the direction between the entrances into the wormhole, and let the magnitude of the vector $\mathbf{p}$ on the surface $l=\{x\in\Sigma^3, Z=\mathrm{const}\}$ be given by some function $|\mathbf{p}|^2=f(X,Y)$. The number of images of the point A in Fig. 3 under a Poincaré map is infinite, provided that the parameters of the position of the wormhole are not chosen in a special manner. Points near A have a finite number of images. However, this number increases without limit as the point A is approached.

Thus, one can see that the structure of the limiting initial submanifold Λ^3_∞ is quite complicated.

Let us now confine our attention to the simplest case, where the initial submanifold Λ^3 consists of a single connected piece that projects uniquely onto Σ^3 and is immediately matched in the sense of the procedure described in the preceding paragraph, i.e., the images $[x(\tau),p(\tau)]$ of the points $[x(0),p(0)]$ under a Poincaré map lie in Λ^3.

The Lagrangian submanifold Λ^4 constructed on the initial submanifold Λ^3 determines the solution of the Klein–Gordon equation if the conditions (6) are satisfied on the basic nontrivial cycles. For the Lagrangian submanifold Λ^4 there exists a nontrivial cycle γ_3 (see Fig. 2b). It starts at some point $x\in\Lambda^3$, passes along a trajectory that enters the right-hand entrance of the wormhole, leaves the left-hand entrance, and reaches a point x' of the surface Λ^3, after which it is connected by some curve α, lying on Λ^3, with the point x.

Let us consider the quantization condition (6) on a nontrivial cycle γ_3. Let (t,l,θ,ϕ) be coordinates in a neighborhood of a mouth of the wormhole and let $(p_t,p_l,p_\theta,p_\phi)$ be the corresponding conjugate momenta. With the explicit form of the metric (2) the Hamiltonian (5) assumes the form

$$H=-p_t^2+p_l^2+\frac{1}{b_0^2+l_2}\left(p_\theta^2+\frac{1}{\sin^2\theta}p_\phi^2\right)=-m^2. \tag{7}$$

A Hamiltonian system with the Hamiltonian (7) is integrable. For simplicity, let us study trajectories with $\theta=\pi/2$, $p_\theta=0$. The first integrals are

$$E=p_t, \quad M=p_\phi. \tag{8}$$

The trajectories in phase space are determined by the expression

$$p_l(l)=\sqrt{E^2-m^2-\frac{M^2}{b_0^2+l^2}}. \tag{9}$$

Then the integral in the quantization condition (6) along the trajectory is

$$\begin{aligned}\oint_{\gamma_3}p_i dx^i&=\int_\alpha\frac{\partial S}{\partial x^i}dx^i+\int\left\{p_t\frac{dt}{d\tau}+p_l\frac{dl}{d\tau}+p_\phi\frac{d\phi}{d\tau}\right\}d\tau\\&=\int_\alpha\frac{\partial S}{\partial x^i}dx^i+\int\left\{-2E^2+2p_l^2+\frac{2M^2}{b_0^2+l^2}\right\}d\tau\\&=S(x)-S(x')-m^2\int_{l_-}^{l_+}\frac{dl}{p_l(l)}\\&=S(x)-S(x')-\frac{m^2}{E}\Delta T,\end{aligned} \tag{10}$$

where ΔT is a time shift into the past occurring as a result of propagation through the wormhole. Taking into consideration the explicit expression (9) for p_l, we find that ΔT can be expressed in terms of an elliptic integral

$$\begin{aligned}\Delta T&=\int\frac{dT}{d\tau}d\tau=\int\frac{\partial H}{\partial p_t}\frac{d\tau}{dl}dl\\&=-2E\int_{l_-}^{l_+}\frac{dl}{\sqrt{E^2-\tilde{m}^2-M^2/(b_0^2+l^2)}}.\end{aligned} \tag{11}$$

Substituting the expression (10) into the quantization condition (6), we find that this condition expresses the phase correlation between the solutions at the image and inverse image points under a Poincaré map:

$$\frac{1}{2\pi\hbar}\left[S(x)-S(x')-\frac{m^2}{E}\Delta T\right]+\text{ind }\gamma_3=k. \tag{12}$$

Let us consider the case which in the present problem is an analog of the situation arising in a causal paradox. Let the solution be concentrated near a trajectory passing through the wormhole and let the initial point x be close to its image x'. This means that an observer emits quasiclassical particles, so that these particles, passing through the hole, reach the observer at the same moment in time. For $x=x'$ the phases of the solution are automatically identical and the quantization condition (12) gives a condition for the energy of a quasiclassical particle

$$-\frac{m^2}{2\pi\hbar E}\Delta T+\text{ind }\gamma_3=k. \tag{13}$$

The energy E appears in the expression (11) for ΔT, so that the condition (13) is a quite complicated function of E and the parameters of the wormhole. In the simplest case, when the particle propagates along a line connecting the entrances into the wormhole, the angular momentum satisfies $M=0$ and the integral (11) can be calculated. Then the quantization condition (13) is

$$\frac{m^2\Delta Z}{2\pi\hbar\sqrt{E^2-m^2}}+\text{ind }\gamma_3=k, \tag{14}$$

where ΔZ is the distance between the entrances into the wormhole. If the Lagrangian manifold projects everywhere uniquely on the configuration space $R^4=\{x\}$, then the index γ_3 of the curve equals zero and the quantization condition determines a discrete energy spectrum of a quasiclassical particle

$$E=\pm m\sqrt{1+m^2\frac{(\Delta Z)^2}{4\pi^2\hbar^2k^2}}. \tag{15}$$

The interpretation of this quantization rule is as follows. An observer can emit a particle ''into the past,'' but then these particles must satisfy a self-consistency condition, which in the present case is that the energy spectrum be discrete.

When the mass equals zero the Klein–Gordon equation becomes a wave equation and describe propagation of light. The quantization condition (13) in this case is trivial. Thus, in the problem studied the Maslov method does not give any restrictions on the propagation of light along closed isotropic geodesics.

5. DISCUSSION

We have studied a formulation of the Cauchy problem for a scalar field against the background of a space–time M^4 containing closed time-like curves. The existence of closed time-like curves passing through the Cauchy surface Σ^3 imposes restrictions on the choice of initial data on this surface. In order to be able to continue the local solution of the Cauchy problem into a global solution, the initial data must satisfy nonlocal self-consistency conditions (12). The geometric meaning of these conditions is that the Lagrangian submanifold Λ^4 in the phase space Π of a classical Hamiltonian system corresponding to the Klein–Gordon equation determining a global solution in the high-frequency limit must be quantized in the sense of the Maslov theory. In the simplest case such nonlocal self-consistency conditions give the restrictions (15) on the frequency spectrum of the solutions of the Klein–Gordon equation or on the energy spectrum of a quasiclassical scalar relativistic particle described by this equation. Thus, signals (quasiclassical particles) sent ''into the past'' can have only a special form and a discrete energy.

We have studied only the simplest case of self-consistent initial data for Σ^3 that corresponds to the simplest type of geometry of a Lagrangian surface: The initial submanifold Λ^3 on Σ^3 consists of one connected piece $\Lambda^3=\{[x,p(x)], x\in\Sigma^3\}$. When Λ^3 consists of more than one piece, nontrivial cycles of a more complicated form appear on Λ^4 (for example, corresponding to trajectories returning to the initial connected piece after multiple passage through the wormhole). The quantization conditions (6) must hold on these cycles also. This gives new restrictions on the choice of initial data for Σ^3.

The problem of constructing quasiclassical solutions of the Dirac equation in a space–time with closed time-like curves can be studied in the same formalism. The effect, observed in the present problem, of the global structure of space–time on the formulation of the Cauchy problem for a partial differential equation can be analyzed on the basis of the Maslov method on a background of other solutions of the Einstein equations.

I wish to thank V. A. Berezin, A. M. Boyarskiĭ, and A. I. Shafarevich for a fruitful discussion of the question considered in this paper. This work was supported in part by the Russian Fund for Fundamental Research (Grant No. 97-02-l7064-a).

*)e-mail: aneronov@mech.math.msu.su

[1] J. Friedman, M. S. Morris, I. D. Novikov, F. Echeverria, G. Klinkhammer, K. S. Thorne, and U. Yurtsever, Phys. Rev. D **42**, 1915 (1990).
[2] F. Echeverria, G. Klinkhammer, and K. S. Thorne, Phys. Rev. D **44**, 1077 (1991).
[3] M. S. Morris and K. S. Thorne, Am. J. Phys. **56**, 395 (1988).
[4] I. D. Novikov, Zh. Éksp Teor. Fiz. **95** 769 (1989) [Sov. Phys. JETP **68**, 439 (1989)].
[5] M. S. Morris, K. S. Thorne, and U. Yutsever, Phys. Rev. Lett. **61**, 1446 (1988).
[6] E. V. Mikheeva and I. D. Novikov, Phys. Rev. B **47**, 1432 (1993).
[7] A. Carlini, V. P. Frolov, M. B. Mensky, and I. D. Novikov, Int. J. Mod. Phys. **4**, 557 (1995).
[8] V. P. Maslov and M. V. Fedoryuk, *Semi-classical Approximation in Quantum Mechanics*, D. Reidel, Dordrecht; Kluwer Academic, Boston (1981).
[9] J. L. Friedman and M. S. Morris, Phys. Rev. Lett. **66**, 401 (1991).

Translated by M. E. Alferieff

Structural transition in small gas-like clusters

D. I. Zhukhovitskiĭ

Institute of High Temperatures, Russian Academy of Sciences, 127412 Moscow, Russia
(Submitted 25 April 1997)
Zh. Éksp. Teor. Fiz. **113**, 181–190 (January 1998)

We propose a model, which is an alternative to the droplet model and presumes that the number of bonds between the atoms is a minimum, to describe highly excited clusters containing a small number of atoms. It is shown that at sufficiently high temperatures such a structure, which has the form of a system of spontaneously appearing chains of atoms (virtual chains), is realized with a greater probability than the close-packed structure. Analytic estimates are supported by the results of numerical molecular-dynamics simulations. © *1998 American Institute of Physics.* [S1063-7761(98)01201-3]

1. INTRODUCTION

The difficulties in the classical theory of nucleation[1,2] are known to be related to the poor applicability of the droplet model to the description of the properties of clusters consisting of a small number of molecules. The classical theory poses the problem of the kinetics of the vapor–liquid transition, i.e., it is assumed that the temperature is above the melting point. Such temperatures are high for the clusters found in a vapor, which are characterized by strong excitation of both single-particle and collective degrees of freedom for the motion of the molecules comprising a cluster. As numerical simulation has shown, the clusters do not resemble droplets and have the form of shapeless aggregates that are similar to a dense gas (see Refs. 3 and 4), this effect being stronger, the smaller the cluster size (the number of molecules in it). On the basis of the similarity to fluctuation clusters in a dense gas, such states are called gas-like, since the clusters cannot be characterized by a definite volume and density and since the short-range order characteristic of liquids is likewise lacking.

It is not surprising that attempts to improve the droplet model (see, for example, Refs. 5, 6, and 7) that employed it as a zeroth approximation for systematically calculating small corrections in powers of the reciprocal droplet radius, and were aimed at expanding its range of applicability to include smaller radii, were ineffective. The description of the state of clusters at relatively high temperatures requires a model that is an alternative to the droplet model and is not based on perturbative methods.

This paper examines the analytically simplest case of clusters consisting of atoms whose interatomic potential is an additive short-range potential. In such a system, each atom interacts only with its nearest neighbors. The droplet model can be regarded as a limiting case with the maximum number of bonds in the system. The other limiting case, which is, therefore, an alternative to the droplet model, is a system with the minimum number of bonds. In this model the cluster has the form of a set of connected chains. The sequence of atoms in each chain varies as the atoms move; therefore, these chains can be called virtual.

The reason for the appearance of the structure just described is the competition between states that have high binding energy and low statistical weight, and states with low binding energy and high statistical weight. As the temperature rises, the probability of the latter rises, and a transition from a compact to a gas-like structure occurs in the cluster. For this to happen, the energy difference between the compact and gas-like states must not be excessive. This is possible only for clusters containing $g<10$ atoms, since the number of bonds per atom in them is significantly smaller than in a continuous liquid, as a consequence of their finite size.

In contrast to a macroscopic system, the transition in a system with a finite number of particles occurs in a certain temperature range. The purpose of the present work is to investigate the structure of clusters in this range and to describe the transition from a compact to a gas-like structure as the temperature rises. To achieve this goal, we utilize both the analytic expression for the partition function of a gas-like cluster based on simplified modeling ideas for the particle interaction potential and its structure, and a numerical molecular-dynamics simulation of the evolution of clusters in a supersaturated vapor (an ensemble with constant temperature and pressure[4]).

The expression for the partition function enables us to write the equilibrium size distribution of the clusters and to use it to calculate the rate of homogeneous nucleation in the supersaturated vapor. Also, since small cluster sizes are considered, the Frenkel–Döring equations are inapplicable, and discrete equations (in the space of sizes) of the kinetics of nucleation must be used (see, for example, Ref. 8).

In Sec. 2 the limiting cases of a compact cluster and a virtual chain are examined, and their partition functions and the characteristic temperature of the structural transition are evaluated, and an interpolation formula for the potential energy is written. The methodology of the numerical experiment is discussed in Sec. 3, and its results are discussed in Sec. 4.

2. PARTITION FUNCTION OF A SMALL GAS-LIKE CLUSTER

We consider a cluster consisting of g atoms which interact by means of an additive pairwise potential $u(r)$. We evaluate the partition function of the cluster in the limiting cases of low and high temperatures. Let

 1063-7761/98/010101-06$15.00

$$u(r)=\begin{cases} +\infty, & r<a-r_0, \\ (M\omega_0^2/4)(r-a)^2-D_0, & a-r_0\leqslant r\leqslant a+r_0, \\ 0, & r>a+r_0, \end{cases} \tag{1}$$

where M is the mass of an atom, $\omega_0=(2/r_0)\sqrt{D_0/M}$ is the vibrational frequency of a dimer, and D_0 is the depth of the well. The potential (1) is finite and oscillatory in the region where it is negative; it is assumed that the length parameters a and r_0 satisfy $a/r_0\gg 1$, i.e., the potential is short-range.

We assume that a cluster is a system of atoms, each of which has at least one neighbor that belongs to the same cluster and is located at a certain distance not exceeding a certain value. For the potential (1) the maximum distance to the nearest neighbor should clearly be selected equal to $a+r_0$. At low temperature, the cluster has the structure corresponding to closest packing of the atoms. Adopting the assumption that the motion of the atoms is classical, which is valid for an argon-like system even at fairly low temperatures, we estimate the partition function of a cluster with $g \geqslant 3$ using the model of an Einstein crystal:[9]

$$Z_p^{(g)}=\frac{V}{\lambda\!\!\!^{-3}}Z_r^{(g)}Z_v^{(g)}\exp\left(\frac{D_g}{k_BT}\right),$$

$$Z_r^{(g)}=C_r(g)\left(\frac{a}{\lambda\!\!\!^{-}}\right)^3, \quad Z_v^{(g)}=C_v(g)\left(\frac{k_BT}{\hbar\omega_0}\right)^{3g-6}, \tag{2}$$

where V is the volume; $\lambda\!\!\!^{-}=\sqrt{2\pi\hbar^2/Mk_BT}$ is the thermal wavelength; $Z_r^{(g)}$ and $Z_v^{(g)}$ are, respectively, the rotational and vibrational partition functions; k_B is Boltzmann's constant; D_g is the ground-state energy of the cluster; and $C_r(g)$ and $C_v(g)$ are numerical factors determined by the close-packed structure (for example, for the structures of a right triangle and a tetrahedron we have $C_r(3)=C_r(4)=2\pi^2/3$, $C_v(3)=(4/3)\sqrt{2/3}$, and $C_v(4)=\sqrt{2}$).

Let us consider the other limiting case (high temperatures). We say that there is a bond between two atoms if their interaction potential is not zero. We call any subset of atoms in a cluster a virtual chain if the atoms can be numbered so that each ith atom, except the first and the last, is bound only to the $(i-1)$th and $(i+1)$th atoms belonging to that subset (and possibly to other atoms not belonging to that subset). The first atom is bound only to the second in a given chain, and the last is bound to the next-to-last atom. By definition, a circular configuration is not a virtual chain. We call an atom a branch point if it is bound not only to atoms of its own chain, but also to at least one atom belonging to another chain. A new chain appears when an atom is added to an atom that is not the last in a chain. It is not difficult to show that the cluster with the fewest of bonds is a set of virtual chains that are joined to one another by one bond at branch points; all the chains, except the first, have one free end. It is obvious that a cluster containing g atoms has $g-1$ bonds.

Let us assume that the probability of states with a more than the minimum number of bonds is negligibly small. Unlike polymer molecules, the sequence and number of atoms in virtual chains are not constant, as a consequence of the additive interatomic potential. Therefore, we are dealing with virtual, rather than real, chains. The assumption just made imposes geometric constraints on the region of phase space in which the atoms of a cluster can be found. We number them in the following manner. We select a virtual chain with two free ends and number the atoms in it $1,2,\ldots,n_1$ from one end to the other. Then we select one of the branch points in the first chain and assign the number n_1+1 to the atom of a second chain forming a bond with that point. We continue the numbering until we reach the free end of the second chain, which is given the number n_1+n_2. We then select another branch point, and so on. As a result we obtain N virtual chains with n_j atoms in the jth chain, $\Sigma_{j=1}^{N}n_j=g$.

In the virtual-chain approximation, the potential energy of a cluster can be written in the form

$$U_c=\sum_{i=1}^{n_1-1}u(r_{i+1\,i})+u(r_{n_1+1\,n_1})+\sum_{i=n_1+1}^{n_1+n_2-1}u(r_{i+1\,i})$$

$$+\ldots+\sum_{i=g-n_N+1}^{g-1}u(r_{i+1\,i}), \tag{3}$$

where $r_{i+1\,i}=|\mathbf{r}_{i+1}-\mathbf{r}_i|$ is the bond length and $\mathbf{r}_i$ is the coordinate of the ith atom.

To calculate the total partition function of a cluster in this approximation

$$Z_c^{(g)}=\lambda\!\!\!^{-3g}\int\ldots\int{}'\exp\left[-\frac{U_c}{k_BT}\right]d\mathbf{r}_1\ldots d\mathbf{r}_g, \tag{4}$$

where the prime on the integral signifies that the integration is carried out over the region in which only physically different states are realized, we make the replacement of variables

$$\begin{aligned}
&\mathbf{r}_1=\mathbf{q}_1,\\
&\mathbf{r}_2=\mathbf{q}_1+\mathbf{q}_2,\\
&\ldots\\
&\mathbf{r}_{n_1+1}=\mathbf{r}_{b_1}+\mathbf{q}_{n_1+1},\\
&\mathbf{r}_{n_1+2}=\mathbf{r}_{b_1}+\mathbf{q}_{n_1+1}+\mathbf{q}_{n_1+2},\\
&\ldots\\
&\mathbf{r}_g=\mathbf{r}_{b_{N-1}}+\mathbf{q}_{g-n_N+1}+\ldots+\mathbf{q}_g,
\end{aligned} \tag{5}$$

where $\mathbf{r}_{b_1}$ is the radius vector of the first branch point. The partition function (4) can be factored and expressed in terms of the partition function of a dimer $Z_c^{(2)}$:

$$Z_c^{(g)}=\frac{V}{\lambda\!\!\!^{-3g}}\int\ldots\int{}'\prod_{i=1}^{g-1}\exp\left[-\frac{u(q_i)}{k_BT}\right]d\mathbf{q}_1\ldots d\mathbf{q}_{g-1}$$

$$=\frac{V}{\lambda\!\!\!^{-3g}}\left\{\int{}'\exp\left[-\frac{u(q)}{k_BT}\right]d\mathbf{q}\right\}^{g-1}=\frac{V}{\lambda\!\!\!^{-3}}[Z_c^{(2)}]^{g-1}. \tag{6}$$

In the special case of an absence of branch points, Eq. (6) corresponds to the partition function of a macromolecule in the standard Gaussian model of a polymer chain.[10] It follows from (6) that the mean potential energy of a cluster depends on its size: $U_g=\langle U_c\rangle=(g-1)U_2$, where U_2 is the mean potential energy of the dimer at the same temperature.

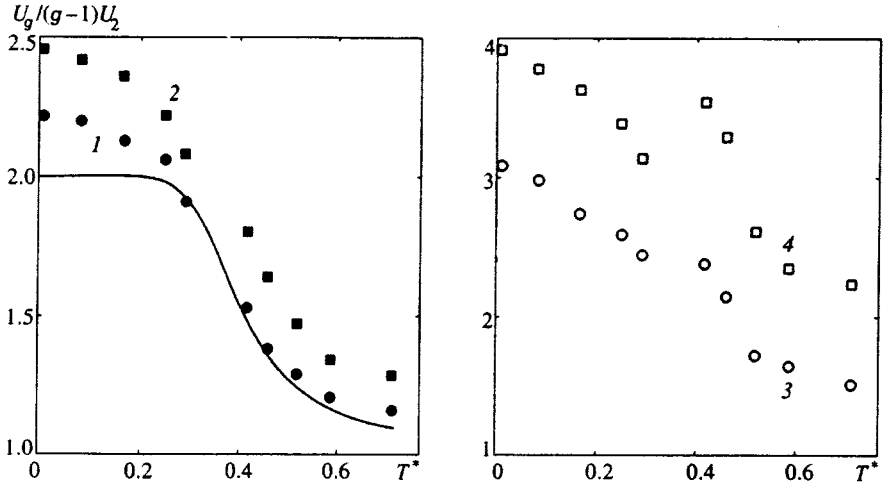


FIG. 1. Dependence of the mean potential energy of a cluster on heat-bath temperature for various cluster sizes: *1*—$g=5$, *2*—$g=7$, *3*—$g=14$, *4*—$g=50$. Curve—calculation based on Eq. (8).

Relations (2) and (6) specify the ratio between the probabilities of structures with the minimum ($P_{\min}$) and the maximum number of bonds ($P_{\max}$):

$$\frac{P_{\min}}{P_{\max}}=\frac{[Z_c^{(2)}]^{g-1}}{Z_r^{(g)}Z_v^{(g)}}\exp\left(-\frac{\Delta E_g}{k_BT}\right)$$

$$=\frac{\pi^{g-1}}{C_rC_v}\left(\frac{a}{r_0}\right)^{2g-5}\left(\frac{2D_0}{\pi k_BT}\right)^{g-2.5}\exp\left(-\frac{\Delta E_g}{k_BT}\right), \quad (7)$$

where $\Delta E_g=D_g-(g-1)D_0$ and the estimate $Z_c^{(2)}\simeq Z_r^{(2)}Z_v^{(2)}=\pi(a/\lambda\!\!\!^{-})^2(k_BT/\hbar\omega_0)$. It follows from (7) that the transition under consideration is smooth and occurs in a certain temperature range. For example, assuming for simplicity that only states with the minimum and the maximum number of bonds are realized and using (7), we find the temperature dependence of its mean potential energy

$$\frac{U_4}{3U_2}=1+\frac{1}{1+\left(\frac{T_0}{T}\right)^{3/2}\exp\left[\frac{3D_0}{k_B}\left(\frac{1}{T_0}-\frac{1}{T}\right)\right]}. \quad (8)$$

This dependence is shown in Fig. 1. It is seen that the transition from the compact state to the gas-like state has a considerable width.

The characteristic temperature T_0 in (8), at which $P_{\min}=P_{\max}$, is called the transition temperature by convention. In the general case, T_0 is the solution of the transcendental equation

$$\ln\frac{a}{r_0}+\frac{1}{2}\ln\frac{2D_0}{\pi k_BT_0}+(2g-5)^{-1}\ln\frac{\pi^{g-1}}{C_rC_v}$$

$$=\frac{\Delta E_g}{(2g-5)k_BT_0}. \quad (9)$$

For typical values of the parameters, the second and third terms on the left-hand side of Eq. (9) are of order unity, but the first term is greater than unity by virtue of the definition of a short-range potential. Since for the latter $\Delta E_g/g\to 5D_0$ as $g\to\infty$ (Ref. 11), it is not difficult to show for $a/r_0\sim 10$ and a temperature below the critical value that $P_{\min}/P_{\max}<1$, i.e., the structure of a large cluster is compact. Since the value of ΔE_g for a short-range potential increases sharply when the first shell of a cluster is filled ($2\leqslant g\leqslant 13$ (Ref. 11)), it can be expected in this case that $P_{\min}/P_{\max}>1$, i.e., light clusters are gas-like, the characteristic size for onset of the structural transition being $g\sim 10$.

For a trimer, Eq. (9) reduces to

$$\sqrt{\beta}e^{-\beta}=\frac{8}{9}\sqrt{\frac{\pi}{3}\frac{r_0}{a}}, \quad (10)$$

where $\beta=D_0/k_BT_0$. For oscillations near the bottom of the Lennard–Jones 6–12 potential we have $a/r_0=6$, and from (10) it follows that $k_BT_0\simeq 0.434D_0$. When $g=4$, the transition temperature is also determined by Eq. (10), with the numerical factor $(\pi/9)^{1/6}$ on the right-hand side. For a tetramer $k_BT_0\simeq 0.416D_0$, which is close to the transition temperature for a trimer.

It follows from (6) that any thermodynamic function of a small cluster that is a linear functional of $\ln Z_c^{(g)}$ is proportional to $g-1$. This makes it possible to construct a simple interpolation formula for the size dependence of any thermodynamic function that is additive for a macroscopic substance. For example, for the mean potential energy of a cluster we have

$$U_g=(g_0-1)(U_2-\bar{u})+(g-1)\bar{u}, \quad (11)$$

where g_0 is the number of atoms on the surface of the cluster and

$$\bar{u}=\frac{1}{2}\lim_{g\to\infty}\left[\sum_{j=1}^{i_0-1}u(r_{i_0j})+\sum_{j=i_0+1}^{g}u(r_{i_0j})\right]$$

is the potential energy of an atom in a continuous liquid (i_0 is the number of the ''central'' atom, which is located closest to the center of mass of the cluster). The expression for the chemical potential of a cluster, which was previously postulated (Eq. (2) in Ref. 12), can be obtained in a similar manner. Using the model in Ref. 12, we write

$$g_0 = 3\Omega(g-g_0)^{2/3} + 3\Omega\lambda(g-g_0)^{1/3} + \Omega\lambda^2,$$

$$\Omega = \frac{4\pi}{3}\frac{\sigma_0 r_l^2}{U_2 - \overline{u}}, \tag{12}$$

where $\lambda = (z/\Omega - 3/4)^{1/2} - 3/2$, z is the coordination number in the liquid, $\sigma_0 = \sigma_f - T(d\sigma_f/dT)$, σ_f is the surface tension of a plane surface of the liquid, $r_l = (3/4\pi n_l)^{1/3}$, and n_l is the concentration of atoms in the continuous liquid.

We write (11) in the form that is characteristic of the droplet model:

$$U_g = 4\pi\sigma(g)R_g^2 + (g-1)\overline{u},$$

$$\sigma(g) = \frac{\sigma_0}{3\Omega g^{2/3}}[g_0(g) - 1], \tag{13}$$

where $R_g = r_l g^{1/3}$. It follows from the first relation in (13) that

$$\sigma(g) = \frac{1}{(36\pi)^{1/3}}\left(\frac{n_l}{g}\right)^{2/3}[U_g - (g-1)\overline{u}]. \tag{14}$$

The assumptions underlying the model in Ref. 12 and the relations obtained above can be tested by a numerical molecular-dynamics simulation.

3. NUMERICAL SIMULATION

The goal of the numerical simulation was to study ensemble-averaged quantities that characterize the structure of a cluster. For this purpose we selected a realistic short-range interatomic interaction potential with $r_c = 1.6a$, which is greater than the mean interatomic distance, but less than twice the distance:

$$u(r) = \begin{cases} v(r) + v(2r_c - r) - 2v(r_c), & r \leqslant r_c, \\ 0, & r > r_c, \end{cases} \tag{15}$$

where $v(r) = D_0[(a/r)^{12} - 2(a/r)^6]$. The form of the potentials $u(r)$ and $v(r)$ differs only in the vicinity of $r = r_c$; at this point $u(r)$ is continuous, along with the derivative. The small value of the cutoff parameter permits the use of the same criterion for assigning an atom to a cluster as in Sec. 2.

The (P,T)-ensemble technique[4] was used to simulate the evolution of a cluster in a vapor with constant pressure and temperature. As in Ref. 4, the cluster temperature was stabilized by introducing an additional Berendsen ''frictional force.'' [13] The cluster temperature was evaluated from the formula $T = [M/3k_B(g-1)]\Sigma_{j=1}^{g}(\mathbf{v}_j - \mathbf{v}_{\rm cm})^2$, where $\mathbf{v}_{\rm cm}$ is the center-of-mass velocity. The evolution of a cluster was investigated at various values of the heat-bath temperature $T^* = k_B\overline{T}/D_0$. The radius of the spherical cell was $10a$, and at $T^* > 0.42$ the number of vapor atoms in the cell was set to 40–50. Under these conditions the size of clusters with an initial size $g \leqslant 460$ decreases as a result of the evaporation of atoms from their surface. Each numerical realization was repeated many times. At $T^* < 0.3$, evaporation from the clusters was so insignificant that no vapor atoms were generated at the cell boundary.

To analyze the structure of a cluster in a numerical experiment, it is convenient to introduce the concept of a simple virtual chain. By definition, a subset of atoms forms a simple virtual chain if there is a way to label them so that for each ith atom the $(i-1)$th and $(i+1)$th atoms are its nearest neighbors, and the first and last atoms each have only one neighbor at a distance not exceeding r_c. If there is a closed ring, we assume that the first and last atoms are the ones separated by the largest distance. This definition is applicable to any potential. In analogy to (3), we can define the potential energy of a system of simple virtual chains by the relation $U_{\rm sc} = \Sigma_i u(r_{i+1\,i})$, where only the interaction energies with the two nearest neighbors are taken into account, and interactions at the ends of the chains are disregarded.

Another important parameter of the structure is $U_{\min}$, which is the sum of the $g-1$ lowest (largest in absolute value) pairwise interaction energies, out of the total number $g(g-1)/2$ of such interactions. Here $U_{\rm sc}$ and $U_{\min}$ are the upper and lower estimates of the total energy of the bonds when a cluster is represented as a system of virtual chains. If a cluster consists of one chain, then $U_{\rm sc}$ and $U_{\min}$ are clearly close to the total potential energy of the cluster U_g, and $U_g/(g-1)U_2$ is close to unity.

4. RESULTS AND DISCUSSION

The temperature dependence of the potential energy U_g obtained as a result of averaging is presented in Fig. 1 for several cluster sizes. It is seen that in the temperature range $0.25 < T^* < 0.5$ there is a significant decrease in the ratio $U_g(T)/(g-1)U_2(T)$. However, while for $g \gtrsim 10$ this ratio is appreciably greater than unity even at high temperatures, for $g \lesssim 10$ it decreases to values close to unity. We note the satisfactory agreement between the theoretical estimate (Eq. (8)) and the data from the numerical experiment. Thus, at high enough temperatures, the potential energy of small clusters corresponds approximately to the minimum number of bonds.

The calculations show that at $T^* = 0.71$ and $g < 8$ we have $(U_{\rm sc} - U_g)/k_BT \leqslant 1.4$ and $(U_{\min} - U_g)/k_BT \leqslant 0.5$, i.e., the true potential energy can be replaced by the approximate potential energy U_c in the Hamiltonian of the system. Conversely, at large g we have $(U_{\rm sc} - U_g)/k_BT \gg 1$, which suggests a transition to the compact structure. Figure 2 presents the ratio of the potential energy, calculated in various approximations, to the energy of a system of virtual chains $(g-1)U_2$ as a function of cluster size. At small g this ratio is close to unity (curves *1–3*). If the energy of the interatomic interaction in the droplet model (the maximum number of bonds) is estimated to be $U_p = U_2(T)U_g(0)/U_2(0)$, where $U_g(0)/U_2(0)$ is the number of bonds, we find that $(U_g - U_p)/k_BT \gg 1$ in the range of sizes in Fig. 2, i.e., the energy of a close-packed cluster differs strongly from U_g (curve *4*). The calculation shows that U_g approximates U_p at $g \sim 10^2$. This is associated with a transition to the compact structure. Thus, the mean potential energy of the small clusters is far closer to the energy of a system of virtual chains than to the energy of a macroscopic droplet.

The calculations suggest that the mean number of bonds in simple virtual chains N_c at large g does not depend on the cluster size, while at $g < 20$ it increases sharply with decreasing g and reaches a maximum at $g = 7$, after which N_c re-

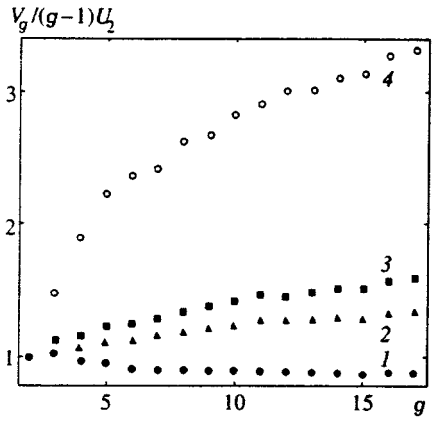


FIG. 2. Size dependence of the potential energy of a cluster calculated in various approximations: *1*—$V_g = U_{sc}$, *2*—U_{min}, *3*—U_g, *4*—U_p. $T^* = 0.71$.

mains close to $g-1$. This means that at small g the probability of states with a single virtual chain is high. A calculation of the size dependence of this probability reveals that states with a single virtual chain are dominant for small clusters, while they are essentially absent for $g > 9$. The values of this probability for a trimer and a tetramer are similar, in agreement with the estimates in Sec. 2. Typical configurations of small clusters with no more than one branch point that are observed in the numerical experiment are shown in Fig. 3.

To test the interpolation formula (11) it is convenient to compare the value of σ determined directly from U_g using (14) with the value calculated from Eq. (13) (Fig. 4). The following values of the parameters were used: $\overline{u} = 3.264D_0$, $n_l = 0.544a^{-3}$, and $z = 9$. The asymptotic value $\sigma_0 = 0.904D_0/a^2$ is reached essentially by $g = 400$. The parameter Ω was varied to achieve the best fit to the experimental data. The value obtained, $\Omega = 0.794$, is typical of many substances, and is consistent with the thermodynamic model in Ref. 12. It is clear from Fig. 4 that the curve faithfully describes the numerical experiment. Since the identity of (12), (13), and (14) is a direct consequence of the interpolation formula (11), Fig. 4 confirms its validity.

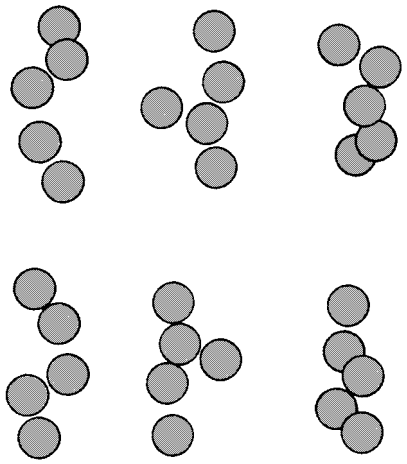


FIG. 3. Configuration of clusters with virtual-chain structure. Two configurations contain one branch point, and the others contain none.

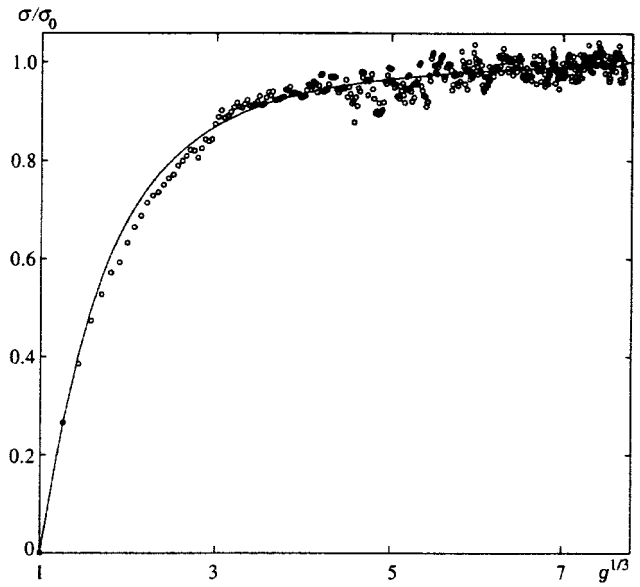


FIG. 4. Size dependence of the effective surface tension at $T^* = 0.46$. Curve—calculation based on Eqs. (12) and (13), points—value of σ corresponding to the potential energy of the cluster determined in a numerical experiment (Eq. (14)).

Figure 5 presents the radial distribution function $G(r)$ for the central atom. We note the following special features. For a large cluster ($g = 430$) it has the same form as for a continuous liquid, and maxima that correspond to the first three coordination spheres. The size effect begins to show up at $g = 60$: the third maximum vanishes, and the height of the first two maxima decreases. At $g < 18$ a plateau forms instead of the second maximum, suggesting a rapid weakening of the correlations. At the same time, the number of atoms in the region corresponding to the first coordination sphere decreases sharply. For example, when $g = 6$, it is 2.46. This behavior can be attributed to the emergence of virtual chains, in which each atom correlates with only two nearest neighbors. This phenomenon is similar to the weakening of the correlations in a freely articulated Gaussian chain.[10] It would

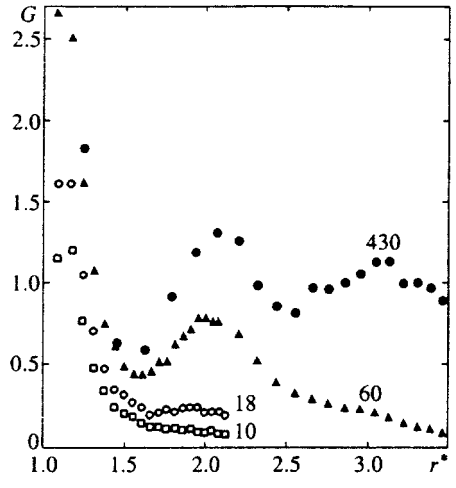


FIG. 5. Radial distribution function for the central atoms of clusters of various sizes. $T^* = 0.46$; $r^* = 2^{1/6}r/a$. The cluster sizes are indicated near the curves.

be interesting to compare the distribution functions shown in Fig. 4 with the distribution functions that might be determined in a real experiment.

Thus, at sufficiently high temperatures ($T^*>0.4$), a cluster containing no more than a dozen or so atoms exists in a special, gas-like state and has the form of a set of virtual chains. Three ranges of cluster sizes can be identified: $2 \leqslant g<10$, $10\leqslant g\leqslant 300$, and $g>300$. In the first range, clusters are systems of virtual chains, in the third they acquire the properties of macroscopic droplets, and the second range is transitional.

[1] I. I. Frenkel', *Kinetic Theory of Liquids*, Clarendon Press, Oxford (1946).
[2] Ya. B. Zel'dovich, Zh. Éksp. Teor. Fiz. **12**, 525 (1942).
[3] R. S. Dumont, S. Jain, and A. G. Basile, J. Chem. Phys. **102**, 4227 (1995).
[4] D. I. Zhukhovitskiĭ, Zh. Éksp. Teor. Fiz. **109**, 839 (1996) [JETP **82**, 451 (1996)].
[5] F. P. Buff, J. Chem. Phys. **23**, 419 (1955).
[6] A. Dillman and G. E. A. Meier, J. Chem. Phys. **94**, 3872 (1991).
[7] I. J. Ford, A. Laaksonen, and M. Kulmala, J. Chem. Phys. **99**, 764 (1993); V. I. Kalikman and M. E. H. van Dongen, J. Chem. Phys. **103**, 4250 (1995).
[8] J. L. Katz, H. Saltsburg, and H. Reiss, J. Colloid Interface Sci. **21**, 560 (1966).
[9] M. R. Hoare and P. Pal, Adv. Phys. **24**, 645 (1975).
[10] A. Yu. Grosberg and A. R. Khokhlov, *Statistical Physics of Macromolecules*, American Institute of Physics, New York (1994).
[11] B. M. Smirnov, Usp. Fiz. Nauk **162**, 97 (1992) [Sov. Phys. Usp. **35**, 37 (1992)].
[12] D. I. Zhukhovitskii, J. Chem. Phys. **101**, 5076 (1994).
[13] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, J. Chem. Phys. **81**, 3684 (1984).

Translated by P. Shelnitz

On the nature of turbulence

L. N. Pyatnitskiĭ

Institute of High Temperatures, Russian Academy of Sciences, 127412 Moscow, Russia
(Submitted 10 June 1997)
Zh. Éksp. Teor. Fiz. **113**, 191–203 (January 1998)

A concept of turbulence is presented that is based on the results of an investigation of the structure of a gas flow in a tube with a square cross section in front of a nonsteady-state moving flame front. It is shown that a region of elevated pressure, consisting of alternating condensations, is formed in the gas flow near the tube walls. These condensations are the sources of waves which form a distribution of velocity fluctuations in the gas flow over a wide range of amplitudes, frequencies and directions. The dynamics of the perturbations at the walls and the configuration of the wave in the gap make it possible to consider the fluctuations in the flow as pseudochaotic and to use statistical methods to describe them. © *1998 American Institute of Physics.* [S1063-7761(98)01301-8]

1. INTRODUCTION

The turbulent flow of a gas or liquid is characterized by chaotic fluctuations of the thermodynamic parameters of the medium. Turbulence has been studied for more than a hundred years but the basis of the phenomenon, the question of its cause and the chaotization mechanism of the motion are still open.

Prandtl[1] explained this phenomenon by the formation of vortices near the wall and their breakup into finer eddies when they are ejected into the primary flow. But the idea was not embodied in a specific mechanism. Another approach to the problem is based on an analysis of the conditions for the loss of stability of the system of hydrodynamic equations by the small-perturbation method (see Refs. 2 and 3).

The formalism of the small-perturbation method as applied to the Navier–Stokes and continuity equations makes it possible to determine the conditions for the appearance of turbulence but not its mechanism. Correspondingly, turbulence in the main flow turns out to be an abstraction that is not related to the physical nature of the phenomenon. Then, by resorting to a number of hypotheses that take account of the nonlinearity of the equations of hydrodynamics[4] turbulence is considered as a stochastic phenomenon and statistical methods are used to describe it. Attempts have also been made on this basis to describe the development of turbulence; concepts of chaotic dynamics, such as bifurcation and strange attractors[5] (see also Refs. 6 and 7) are used for this purpose.

The small-perturbation method can probably be considered the most successful today. Variations of the thermodynamic parameters (velocity v, pressure p, density ρ, temperature T) are found in it in the form of a harmonic solution of the same system of hydrodynamic equations but linearized with respect to small variations. In this situation the question of the loss of flow stability is considered only with respect to the effect of perturbations along the stream lines,[2] ignoring transverse effects. The following plausible arguments serve, at first glance, as the basis for such an assumption. The perturbations of the parameters are small compared with their average values. Moreover, the diameter of the flow is much smaller than its length and the steady-state distribution of the parameters is established rapidly in the transverse direction.

Meanwhile, the velocity fluctuations observed in the presence of turbulence have an arbitrary orientation. And since no motion variations (or fluctuations) in themselves occur, the acting forces, which are local pressure gradients in a given case, must be analyzed. It is also obvious that the pressure fluctuations originate in the boundary of the flow. But the pressure perturbation appearing within a small volume can propagate throughout the entire flow only in the form of a wave, for which all propagation directions are equivalent — a circumstances which is ignored in the small-perturbation method when the equations are linearized. Finally, the pressure distribution over the cross section becomes steady-state in the time during which a sound wave is at least damped. This time considerably exceeds the characteristic durations of turbulent processes and therefore the distribution of the parameters over a cross section must be regarded as stationary. It is interesting to point out that H. A. Lorentz,[8] attempting to explain the turbulization phenomenon, assumed a form of perturbation consistent only with the continuity equation and not necessarily satisfying the original flow equations.

A turbulence concept is presented below according to which it is precisely small pressure gradients of arbitrary orientation that are responsible for the velocity fluctuations in the flow. Within this model there is no need to refer to nonlinear interactions to understand the nature of turbulence; it is sufficient to use the linear wave equation that describes the propagation of a sound wave in the medium. In general features it reduces to the following. The drag exerted on the gas flow by the wall causes a reduction in the flow velocity v and, correspondingly, an increase in the pressure p. Flow perturbations adjacent to the wall propagate as sound waves throughout the entire cross section. The reflection from the walls of waves with different directions causes a nonuniform pressure distribution at the flow boundary, in which short-lived renewable local perturbations are formed. The waves of these perturbations fill the flow and create a network of irregular parameter fluctuations in it, which look chaotic. The drag can of course also produce waves of finite amplitude in

the medium: simple or shock waves, which does not alter the overall picture.

2. PERTURBATION WAVE IN A BOUNDED SPACE

Proceeding to an analysis of the process in which we are interested, we compare the behavior of a local sonic perturbation under the conditions of spherical symmetry and a bounded space. The problem is well-known in the first formulation[2] and it is solved in terms of the wave equation for the potential ϕ of the perturbation v' of the velocity v:

$$\frac{\partial^2\phi}{\partial t^2}=c^2\,\frac{l}{r^2}\,\frac{\partial}{\partial r}\left(r^2\,\frac{\partial\phi}{\partial r}\right), \tag{1}$$

where c is the velocity of sound. We assume that the perturbations of the velocity, density and other thermodynamic parameters of the medium are small, $v'\ll c$, $\rho'\ll\rho,\ldots.$, and the medium is at rest, $v=0$.

The solution of this equation is

$$\phi=\frac{\delta(ct-r)}{r}, \tag{2}$$

where $\delta(ct-r)$ is a delta function. It describes the propagation of an elementary wave, which is a spherical surface of radius $r=ct$ with a decaying field moving away from the coordinate origin. A perturbation of arbitrary type, in the form of the initial distribution $f|_{t=0}(a)$ or the complex pulse $f|_{r=0}(c\tau)$ or a combination of them, for example, can be composed of elementary waves. Then at a distance $r>a\gg c\tau$ the perturbation will have the form of a spherical layer whose thickness and profile are given by the function f.

Let us trace the propagation of the perturbation in a bounded space for the example of an elementary wave in a gap of height d, formed by two parallel walls. The $x-y$ plane coincides with the lower wall, the z axis is directed vertically, and the center of the perturbation is located at the coordinate origin. In the diagram of Fig. 1a the gap walls in the $x-z$ plane (the bottom portion of the diagram) are designated by the heavy lines.

The location of the elementary wave at an arbitrary instant of time t in the free half-space (in the $y=0$ plane in the diagram) is shown shaded. Within the gap the wave remains spherical only until its first collision with the wall, $ct<d$. Then it changes its configuration first because of reflection from the upper wall, then from the lower, etc. The regions of the sphere surface, filling the gap between the walls after each reflection, are separated by lines (or planes) in the diagram, located at distances that are multiples of d from the upper boundary of the flow.

In traversing the path $r=ct$ the wave is reflected n times from the walls:

$$n=\mathrm{In}\left(\frac{z}{d}\right)=\mathrm{In}\left(\frac{\sqrt{(ct)^2-x^2-y^2}}{d}\right). \tag{3}$$

Here the operation of taking the integer part of a number is denoted by In, and the sequence $n=0, 1, 2,...,$ m constitutes the reflection number, where $n=0$ corresponds to unbounded space and m is the maximum number of reflections,

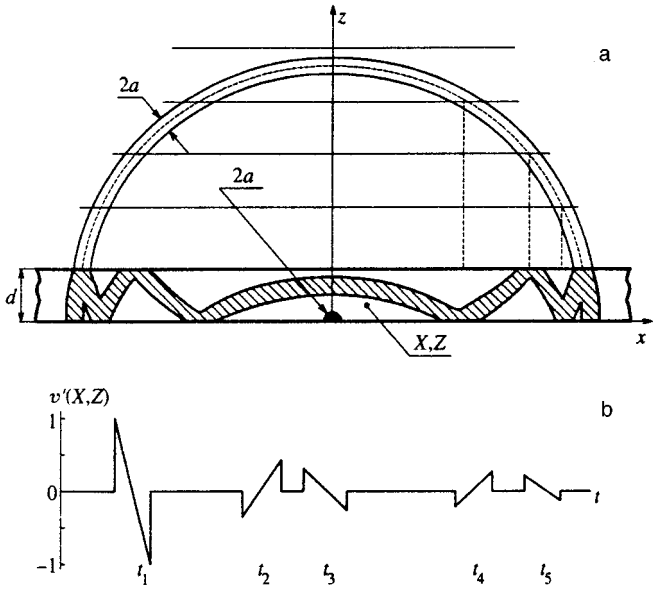


FIG. 1. Wave configuration (a) and velocity fluctuations (b) in gap between walls.

for which the fluctuations of the parameters created by the wave amount to significant values for the particular problem, about which more will be said below.

By knowing n, it is easy to establish the relationship between the coordinates and the propagation directions at corresponding points of the wave in the half-space and in the gap. Let us assume the point $r=(x^2+y^2+z^2)^{1/2}$ of a spherical wave with polar θ and azimuthal φ angles corresponds to the point $R=(X^2+Y^2+Z^2)^{1/2}$ with ϑ_n and φ_n of the wave in the gap. Then,

$$x=X,\quad y=Y,\quad z_n=2d\ \mathrm{In}\left(\frac{n+1}{2}\right)+(-1)^nZ, \tag{4}$$

$$\cos\ \vartheta_n=(-1)^n\,\frac{z_n}{ct}=(-1)^n\ \cos\ \vartheta,\quad \tan\ \varphi_n=\frac{y}{x}=\tan\ \varphi. \tag{5}$$

Note that the wave passes through the point R many times and the times at which this passage occurs form the sequence t_n, which can be characterized by the ''frequency'' $\nu_n(R,\vartheta)$. We find from Eq. (3) for the nth reflection:

$$t_n=\frac{z_n}{c\ \cos\ \vartheta_n}=\frac{\sqrt{x^2+y^2}}{c}\ \sqrt{1+\frac{z_n^2(Z)}{x^2+y^2}},$$

$$\nu_n=\frac{1}{t_{n+1}-t_n}. \tag{6}$$

It follows from this formula that the arrival of the wave at the arbitrary point R is not adjustable in terms of recurrence, nor in terms of amplitude, nor in terms of propagation direction. Thus, even a single perturbation in the gap creates a field of irregular fluctuations.

Let us illustrate their frequency characteristics at the point $Z=0.5d$ by examples for two limiting positions of the observation point—near $(X^2+Y^2)^{1/2}\ll d$ and far $(X^2+Y^2)^{1/2}\gg d$:

$$\nu_1 \simeq \frac{c}{d}, \quad \nu_2 \approx \frac{c}{d}\frac{R}{d}\frac{1}{n+1}. \tag{7}$$

In the first case the frequency of the fluctuations changes little but, as will be shown below, the fluctuations damp out quickly. In the second case in the first reflections the parameters fluctuate with a greater frequency; however, as n increases the frequency of the fluctuations decreases. Thus, the fluctuations have a broad frequency spectrum.

3. PARAMETER FLUCTUATIONS

Let us now consider the behavior of the fluctuations in the wave. The velocity v' (by definition) and the density ρ' are related to the potential ϕ by the expressions

$$v' = \frac{\partial \phi}{\partial r}, \quad \rho' = -\frac{\rho}{c^2}\frac{\partial \phi}{\partial t}. \tag{8}$$

Let us dwell on the simple case, assuming to be specific that at the initial time $t=0$ the perturbation has the form of a bubble with the small diameter $2a$ $(a \ll d)$ with its center at the coordinate origin. Inside the bubble the gas is compressed so that $\rho' = \Delta = \text{const}$ and $v' = 0$. Outside the bubble $v' = \rho' = 0$.

Since the wave field according to Eq. (2) decreases quite rapidly with distance $\propto 1/r$, absorption can be ignored. However, upon reflection from a wall (generally speaking from any surface where the impedance ρc experiences a sudden change) additional losses appear. Without getting into details for now, we take them into account by introducing the loss factor α, and we will look for the perturbation field for chosen initial conditions with Eq. (3) taken into account. At the point **R** of the gap, corresponding to the point **r** of the spherical wave, we have

$$v' = 0 \quad \text{for} \quad \frac{r-a}{c} > t > \frac{r+a}{c}$$

$$v' \approx (-1)^n (1-\alpha)^n \frac{c\Delta}{2\rho}\frac{r-ct}{r} \quad \text{for} \quad \frac{r-a}{c} < t < \frac{r+a}{c} \tag{9}$$

for the velocity fluctuations v' (to within terms of order a^2/r^2). Equations (9) and (3)–(5) define the configuration of the wave in the gap and the velocity fluctuations $v'(t)$ at the center of the flow $Z \in [2a, (d-2a)]$. A single perturbation with diameter $2a$ creates in the gap a moving wave layer of thickness $2a$. The diagram in Fig. 1a, where the layer in the gap is shaded, demonstrates the unique features of its configuration in the gap and in the upper half-space.

The character of the fluctuations at the center of the flow is illustrated in Fig. 1b for the example of the velocity $v'(t)$ at the point $R(X, Y=0, Z)$, marked in the diagram of Fig. 1a. The velocity in the wave changes discontinuously; this is caused by the singularity at the boundary of the artificially introduced initial condition ($\Delta = \text{const}$ for $r \leqslant a$). A real perturbation at subsonic velocities should not contain such a singularity. In linear theory, however, this distribution has no effect on the duration of the fluctuations nor on the sequence of the fluctuations.

As a result of the decay of the wave as it propagates a question arises with regard to the distance r^* and the number of reflections $n=m$, for which the fluctuations of the variables can be considered to be significant. By measuring r in units of d so that $\mathscr{R}^* = r^*/d$, and the damping v' in fractions of $v'(d)$, $\mu = [v'(r^*)/v'(d)]$, we find from Eqs. (9) and (3) a transcendental equation for m and the relationship for $\mathscr{R}^*$ in a given direction ϑ:

$$m = \text{In}\left((1-\alpha)^m \frac{\cos\vartheta}{\mu}\right), \quad \mathscr{R}^* = \frac{(1-\alpha)^m}{\mu}. \tag{10}$$

According to Eq. (10), the choice of μ determines the number of reflections m and the distance $r^* = d(1-\alpha)^m \mu^{-1}$, within which the velocity fluctuations v' affect the fluctuations of the medium. It is obvious that the path length r^* for a particular wave will depend on the observation angle ϑ. Let us consider an example. Take $d = 3$ cm, $\mu = 0.01$, and $\alpha = 0.1$ and 0.3. We compare the length r^* for these values of α for three angles: $\vartheta = 0$, 45 and 89.4°. The first of these corresponds to the direction at right angles to the gap while the third corresponds to a direction that is nearly parallel to the X, Y plane when the wave decays without experiencing a single reflection. For the given set of α and ϑ we obtain:

$\alpha = 0.1$: $m = 16; 14; 0$ and, correspondingly,

$r^* \approx 56;\ 69;\ 300$ cm;

$\alpha = 0.3$: $m = 7; 6; 0$ and, correspondingly,

$r^* \approx 22;\ 29;\ 300$ cm.

By introducing the loss factor α upon reflection, we eliminated fluctuations near the wall from consideration. Meanwhile, at its surface we have $v = v' = 0$ and the character of the influence of the wave on the flow is altered. Actually, as one approaches the wall, more precisely in the region $Z \notin [2a, (d-2a)]$, the velocity fluctuations decrease but the rest of the parameters, in turn, fluctuate more strongly. Estimating the amplitude of the variables, let us say the pressure, we use the expression for the energy density in a plane traveling wave in the form[2] $\varepsilon = \rho v'^2$. When the losses associated with reflection are small and $\alpha \ll 1$, the character of the fluctuations obeys the following conditions:

$$v' \to 0, \quad p' \to (2-\alpha)\overline{\rho v'^2}\cos^2\vartheta, \quad \text{for } Z \to 0 (Z \to d), \tag{11}$$

i.e., near the wall the pressure fluctuations increase by almost a factor of two. The region of thickness $2a$ near the wall plays the role of a boundary layer in which the interaction of the spherical incident and spherical reflected waves is not reduced to a linear superposition.

Let us mention one other unique feature of the reflection of a spherical wave related to the refracted wave. Usually the velocity of sound c_w in the wall material is greater than its value c in the gas. In this case a secondary wave[2] with the surface of a truncated cone also arises in the gas in addition to the spherical waves. The larger circumference of the cone is located at the wall surface and is joined to the edge of the front of the spherical refracted wave. The smaller circumference forms a locus of points at which the conical surface of the secondary wave is tangent to the spherical surface of the

reflected wave. The normal to these surfaces at the tangency point intersects the line joining the center point of the forward wave (in the gas) and its virtual reflection in the wall (or specular reflection) at the angle ϑ_0. It amounts to several degrees for a metallic wall.

The secondary wave, unlike the spherical reflected wave, decays along a given direction in inverse proportion to the square of the distance.[2] Therefore, it must have a greater influence on fluctuations near the wall than far from it, where the spherical waves exert the primary influence on the fluctuational characteristics of the medium.

As seen from this analysis, the field of the fluctuations in the variables is a complex space-time function and it is characterized by a broad spectrum of frequencies. It is obvious, however, that the picture of the ''chaotization'' of the motion of the medium cannot be considered to be stochastic. This set of relations for the fluctuations is rather similar to the summary of rules and resembles the solution of the problem of cellular automatons, belonging to the class of irreducibles. This statement is fully valid for the fluctuation field produced by a single perturbation. In reality the fluctuation field is formed under the conditions of repeatedly appearing perturbations. Therefore, it is of interest to compare the obtained results with the actual process for the appearance of turbulence; this will be done in the next section. Before this, however, it would be advisable to mention the problem of dispersion.

It was implicitly assumed above that dispersion is absent in the medium, and the dependence of the frequency ω of the oscillations on the wave number k in the dispersion equation has the simple form $\omega = ck$. It can turn out to be more complicated, and the dispersion equation can have several branches as, for example, in a plasma. The discontinuity of the dispersion equation at the boundary of such a medium will play the role of a wall, and the reflection of a wave from such a wall can serve as the source for the appearance of other wave modes, the number of which depends on the number of branches of the dispersion equation. Thus, the described picture of chaotization is supplemented by a ''multiplication'' of the types of perturbations.

4. EXPERIMENTAL BASIS

An experimental investigation has importance for checking those assumptions which were made in the analysis of the properties of the fluctuation field formed by a single perturbation. A second aspect of the problem also exists, however, that refers to the formation process itself of the primary perturbations. Since no universally accepted method of describing this process exists at this stage of the investigations, its experimental observation is of interest. But this has its own difficulties. They are related primarily to the possibility of visualizing weak inhomogeneities in a gas flow in the initial stage of motion. It is convenient to observe this stage in a tube while the gas flow is being established in front of a piston. In the experiments presented later the flow structure, which is created by a nonsteady-state moving flame (the analog of a piston), is investigated where the intensity of the processes becomes accessible for recording in view of the nonsteady-state nature of the propagation of the flame-piston.

The investigations were performed in a tube with a 28.6×28.6 mm square cross section with polished walls. It consisted of standard sections, in one of which 18-cm long optical glass plates served as the side walls for visualizing the distribution pattern of optical inhomogeneities in the flow. Displacement of the sections made it possible to record the flow status in any portion of the tube. Combustion was initiated in a CH_4+4O_2 mixture by a short (1 μsec) inductive electric discharge. The tube had a length (1.5 m) for which waves, reflected from the end opposite the ignition, had no effect on the gas flow in front of the flame until the detonation.

The flow structure was recorded by the schlieren method. In this method the distribution of the density gradient $\nabla\rho$ (and in the absence of chemical reactions also the pressure gradient ∇p) is visualized in the direction of the normal $\mathbf{N}$ to a Foucault knife edge, and the excess or deficiency of the luminance of the image point is determined by the sign of the scalar product $\mathbf{N}\cdot\nabla\rho$ at the conjugate point of the object. A high-speed movie camera made it possible to obtain photographs of the visualized pattern of the distribution of flow inhomogeneities at exposure times from 100 to 0.5 μsec. The experimental procedure and conditions have been described elsewhere.[9]

The onset of combustion produces the first pressure perturbation, traveling with the velocity of sound $c \sim 350$ m/sec. It sets the gas into motion. The gas flow is bounded at the front by the first perturbation and at the rear by the flame-piston. The equivalent velocity u of the gas piston in the flow depends on the normal burning rate U, the ratio of the flame surface S to the tube cross-sectional area F, and the expansion coefficient of the gas γ during combustion ($u \ll c$):

$$u = U\frac{S}{F}(\gamma - 1). \tag{12}$$

The chosen mixture has $\gamma = 10$, $U = 2.3$ m/sec, and for a flame with a planar front $u \sim 20$ m/sec. The shape of the front surface duplicates the velocity distribution of the flow in front of the flame and exactly reflects the structure of the flow, making it possible to monitor its character. In this case the surface area S of the flame and, according to Eq. (12), the velocity u are variable quantities. Thus, if $(S/F) \sim 0.02$ holds immediately after ignition, then at later stages, when the gas is slowed down by the wall, the flame surface exceeds the tube cross section by severalfold. Correspondingly, the equivalent piston velocity u can amount to 100 m/sec and more.

Fragments of the high-speed schlieren cinematography of the process are shown in Fig. 2. The picture-taking rate and the field of view of the device were chosen in such a manner as to ensure the temporal and spatial resolution of the pattern of the process. The full tube cross section in a region with a length of 8 cm lies within the frames in Fig. 2. The crosspiece for mounting the optically homogeneous glass is visible within the limits of this region (the vertical

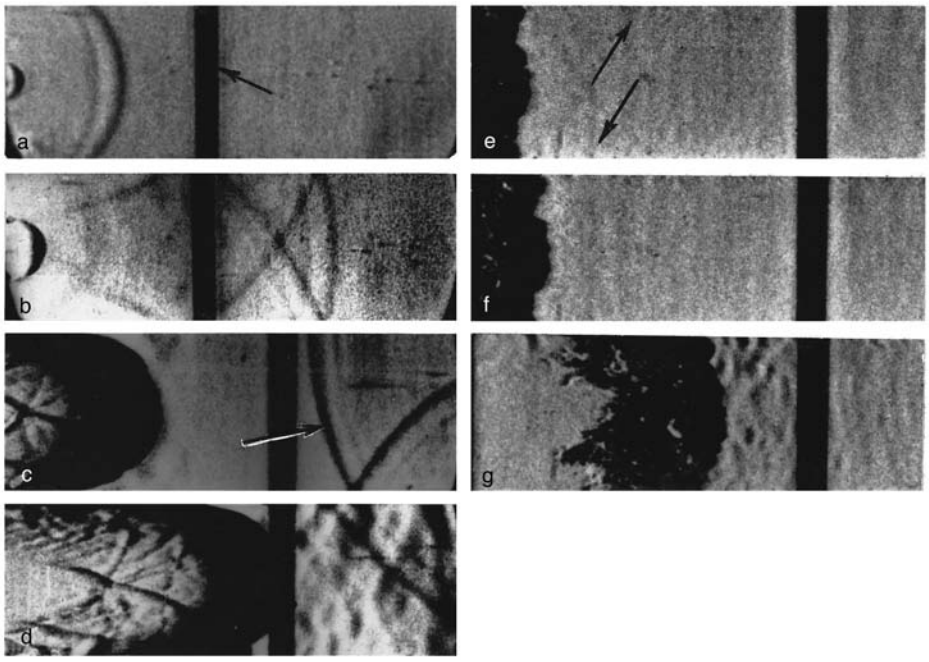


FIG. 2. Oscillograms of the pressure during flame propagation in tube.

dark band in the frames, denoted by the arrow in frame a). The time t is measured from the moment of ignition and the distance x is measured from the end cap of the tube (to the left outside the frame in the photographs). If the flame front is located at the point $x=\mathscr{X}$, then the extent of the gas flow in front of the flame has a length $\ell=ct-\mathscr{X}$.

The photographs of Figs. 2a and 2b visualize the propagation process of the first perturbation produced near the end of the tube by a weak 3-mm diameter spark. They are exposed at the times $t=52$ and 165 μsec. The volume of the burning gas is still small and it can be assumed that $\mathscr{X}\approx 0$ and $\ell\approx x\approx 1.9$ and 6.0 cm. Both wave configurations graphically demonstrate that all propagation directions of the perturbation are equivalent; in this situation the variation of the parameters due to the action of the perturbation is a prolonged process.

Thus, the assumptions that the perturbation exerts an influence only in the direction of the flow and the transverse pressure gradient can be ignored are incorrect. It is also clear that the wave moves almost at the velocity of sound c (3% faster) and has the shape of a spherical layer with a thickness ~ 3 mm, equal to the diameter of the spark, and the structure of the layer is preserved upon reflection from the walls. The same result was obtained for air except that the wave, not supported by the motion of a ''piston,'' decayed rapidly, resulting in a low-contrast image on the photographic paper that is difficult to discern even after the wave has moved by a distance of only 4–5 cm.

Frames c and d refer to the times $t=1.3$ and 1.4 msec when $\mathscr{X}=13$ and 15.5 cm and $\ell\approx 31$ cm. At this time only a portion of the length ℓ, corresponding to 5 and 3 cm, enters the field of view. The accelerated propagation of the flame creates in the flow the structure of a chaotic density distribution, resembling large-scale turbulent fluctuations. Because of the high temperature in the combustion products the sensitivity of the schlieren method is reduced by an order of magnitude and visualization of the perturbations becomes ineffective, but the flame front reacts to them, which transforms the combustion into the turbulent regime. Thus, small-scale fluctuations are also present in the flow in front of the flame although because of their smaller amplitude they are less noticeable if not invisible.

The essential feature of this experiment is that the tube is mounted with a vertical shift of the two adjacent sections so that a projection with a height of ~ 0.3 mm is formed in the $x=22$ cm cross section. This ''irregularity'' is a localized source of strong waves following one another at an interval $\tau\sim 15$ μsec. In the photographs they move from the upper right corner and in frame c the first of these waves is denoted by the arrow. As seen from a comparison of photographs c and d, the ''irregularity'' drastically chaoticizes the flow. The schlieren cinematography of this stage of flow formation confirms the idea, that it is precisely the superposition of the forward and reflected waves that gives rise to the fluctuation velocity field in the flow.

The next three photographs are obtained for the conditions when the motion of the gas in front of the flame occurs at a displacement $\ell\sim 40$ cm, which makes it possible to trace the evolution of the development of the flow structure. Frames e and f ($t\approx 1.9$ msec, $\mathscr{X}\approx 29$ cm, $\ell\approx 37$ cm) reflect the structure in the region where the flame is moving with a constant velocity and it generates almost no new waves. Frame g ($t\approx 2.05$ msec, $\mathscr{X}\approx 32$ cm, $\ell\approx 38$ cm), on the other hand, is recorded at the moment when the flame is accelerated. The different stages of flow formation demonstrate its different properties.

To begin with, while the flame and the gas in front of it move uniformly, as seen from photographs e and f, small-scale fluctuations prevail in the flow. They appear in sharper relief, of course, in the flame structure. Thus, here we observe a process which in the existing theory of turbulence can be called a transfer of energy from large-scale fluctua-

tions to small-scale. Moreover, other inhomogeneities in the flow are more easily observed against such a relatively quiescent background: a dark region near the upper wall and light region near the lower (denoted by the arrows in photograph e).

This important result is explained in the following manner. During the exposure of frames e, f, and g the Foucault knife edge was positioned with a certain tilt so that the vector **N** had a vertical component in the z axis direction. For the chosen orientation of the Foucault knife edge the dark region corresponds to a positive value of the density gradient $\nabla\rho > 0$ and the light region to negative $\nabla\rho < 0$. Consequently, the gas condenses near both walls. Moreover, the drastic change in the luminance at the boundaries of both of these regions means that the component $N_z(\partial\rho/\partial z)$ makes the primary contribution to the scalar product $\mathbf{N}\cdot\nabla\rho$. And since N_z is small, the magnitude of the gradient $(\partial\rho/\partial z)$ must be large, and the density and pressure in the near-wall regions are markedly higher than in the rest of the cross section.

However, these regions are inhomogeneous and contain discrete condensations with typical size $2a \approx 2$ mm. The distribution of inhomogeneities changes rapidly and is usually not discernible in the next frame; this makes it possible to estimate the upper limit of the lifetime of the near-wall condensation. For a picture-taking rate of 62.8×10^3 frames/sec it does not exceed ~ 16 μsec, whereas the sound wave will traverse the shortest path between walls in both directions in a time that is considerably longer than 170 μsec. It can be assumed that the pressure in the condensation is uniformly distributed over the entire cross section during the wave damping, which according to Eq. (10) exceeds each of these quantities by one or two orders of magnitude. In view of this the pressure in the near-wall condensations cannot be equalized over the cross section, and the waves originating from them form a conglomerate of local perturbations in the volume.

The form of the inhomogeneities near the walls is also visible in the photographs of the uniform flow. According to Eq. (9) each type of inhomogeneity, propagating as a spherical perturbation, should increase in size (or diameter) by 1.4 mm during the frame exposure time (2 μsec). In reality, however, the wave layer is deformed and acquires the shape of elongated formations up to 3 mm long, tilted in the flow direction. This is particularly evident in frame g, where the flame and flow start to be accelerated again and the processes become more intense. The deformations of the perturbations are easily explained by the fact that the drift of different portions of the wave during different stages of its development depends on the velocity distribution $v(Z)$, which increases progressively with distance from the wall. A modeling of the propagation of the wave layer (9) in the flow $v(Z)$ near the wall gives precisely such a perturbation contour. Let us add that acceleration of the flow produces additional compression waves which, forming a new $v(Z)$ distribution, intensify the boundary processes.

It becomes obvious from the foregoing analysis of the results of the schlieren cinematography that the condensations near the walls serve as sources of flow perturbations. The question of the mechanism of formation of the condensations, which, by the way, is directly related to the formation of the $v(Z)$ distribution, is still unanswered, however. The most important of the original parameters of the condensations are not so much the geometrical shape and dimensions $(a \ll r)$ as their distribution $F_s(X,Y,Z,t)$ over the surface of the gap walls and with time. The photographs of Figs. 2c and 2d showed that the ''irregularity'' in the fixed position generates perturbations with a period $\tau \approx 15$ μsec. Unfortunately, the equipment does not permit investigating the details of the mechanism; therefore, let us consider the process of the emission of pressure waves by an unsteady moving flame which is similar but more intense and therefore more accessible for recording.

The oscillograms in Figs 3a–3e represent a continuous recording of the excess pressure $p' = p'_x(t)$ at the points $x = 3$, 19, 35, 51 cm. The first of these refers to the propagation region in which the flame is still far from spanning the tube cross section and shows basically the overall increase of pressure in the tube during combustion. The photographs of Figs. 2a and 2b can serve as an illustration here. The next oscillograms are obtained for the developed flame; the pressure peak in this case on each curve corresponds to the time at which the combustion zone passes through the cross section in which the measuring probe is located. The strong increase of the pressure on oscillogram b is recorded in the cross section where, according to the photographs of Figs. 2c and 2d, the flame undergoes an acceleration. The oscillogram of Fig. 3c is obtained in the region along which, as shown in the photograph of Figs. 2e and 2f, the flame is moving with a constant velocity. Finally, the oscillogram of Fig. 3d corresponds to the photograph of the flame (Fig. 2f) during the new acceleration.

The character of the oscillograms and a comparison of them yield the pattern of wave emission. The pressure maximum occurs in the flame localization region. The energy release in the combustion contributes to a pressure increase in the flame localization zone, from which its excess tends to be concentrated over the flow in the form of compression waves. The maximum pressure increase depends on the ratio of the energy release rate to its characteristic outflow time. In the flame acceleration region this is also supplemented by an overall intensification of the burning, as a result of which the pressure increases to a maximum. Wave emission occurs nonmonotonically, which is evident in the discrete nature of the compression waves and is caused by the ''fluctuations'' of energy release due to the effect of variations of the surface S of the flame front, as follows from Eq. (12). The surface of the front, in turn, is acted upon by the flow perturbations propagating in the cross section; this also gives rise to fluctuations in the energy release. It is a unique wide-band self-oscillation process.

In exactly the same way a slowing of the flow is accompanied by a pressure increase near the wall and by wave emission. Therefore, the velocity profile $v(Z)$ develops when the flow and waves interact. This, in turn, is a forward wave moving along the wall. The wave arriving from the second, opposite, wall and after that the secondary and reflected waves, arising at the first wall, also play a role. All waves have different amplitudes and orientation. Therefore, in ac-

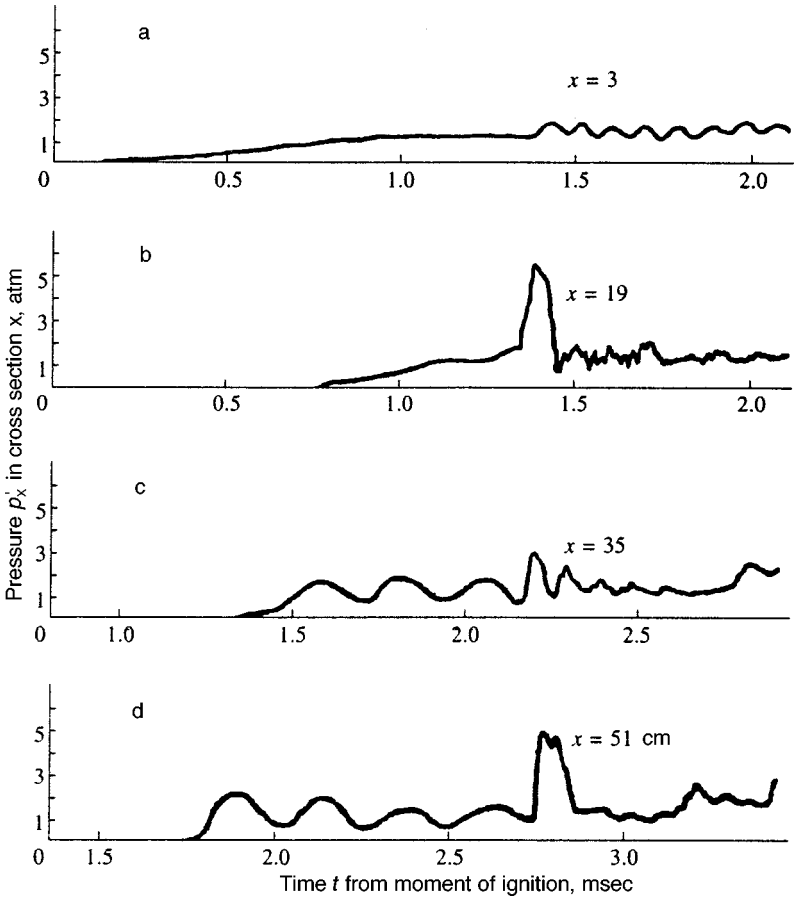


FIG. 3. Oscillograms of the pressure during flame propagation in tube.

cordance with Eq. (11) the reflection conditions for them and individual portions of them are different. As pointed out, the role of the secondary wave is evident over short distances near the walls, whereas the forward and reflected spherical waves not only take part in the formation of the source $F_s(X,Y,Z,t)$ and the profile $v(Z)$ but also produce a fluctuational field in the flow itself. Then the process is repeated for each reflection (within the limits $n \leqslant m$) but with a smaller intensity.

Due to the influence of this interaction the initial smoothness of the deceleration process breaks down very quickly even for the case of an ideally smooth wall, and local ''flickering'' perturbations form in this zone. The boundary perturbations themselves are sources of waves and, completing the cycle of continual renewal of the source $F_s(X,Y,Z,t)$, undamped fluctuations of the parameters in the flow are thereby maintained. The intensity of each wave falls off as they propagate in the gap, and when the flow parameters are constant, the fluctuation field arrives at the steady-state condition. The level of the fluctuations depends, all other things being equal, on the intensity of the cyclic renewal process of the source $F_s(X,Y,Z,t)$, which is determined by the average flow velocity value.

The discussions about the boundary source of pressure fluctuations and the mechanism for its renewal are confirmed qualitatively by the results. This process has not been investigated quantitatively. It is clear, however, that if the emission of intense waves, formed during the burning, was considered[10] on the basis of the theory of the so-called simple (and shock) wave, then the small-amplitude approximations are sufficient for describing the action of perturbations near the walls on the flow. It follows from a comparison of the energy aspects of the combustion and the deceleration of the gas by the wall, and also from the fact that it is hard to detect perturbations in front of the flame in the flow, that they have a small amplitude and the proposed approach is justified.

5. CONCLUSION

The experimental data and the analysis of the conditions for the propagation of a sound wave in the gap between two walls that have been presented above make it possible to propose reasons for the nature and mechanism of turbulence.

Movement of the medium arises because of a pressure gradient, which when flow is established in a wall-bounded space sometimes does not have a planar front, and its total magnitude is composed of a set of compression waves, propagating at different angles to the general flow direction.

Retardation of the flow by the wall raises the pressure in the boundary region. The pressure distribution throughout the entire volume occurs by means of an emission of sound waves from this region. At a sufficient distance (compared with the source dimensions) they assume a spherical shape since all directions of perturbation propagation in the medium are equivalent. The equalization time of the pressure over the cross section is determined by the damping of the perturbation wave in the medium. This time is considerably

greater than the propagation time of the wave at right angles to the cross section and, especially, the time involved in the interaction of the waves with the flow during formation of the velocity profile $v(Z)$ in the boundary region. Wave propagation in the flow causes fluctuations of the thermodynamic parameters, including the velocity. Near the wall the velocity fluctuations become small, but the fluctuations of the other parameters, such as pressure and density, increase.

The primary distinction of the flow state at the flow boundary is the interaction of the forward and reflected waves, in which the reflection process plays an important role. Waves producing a pressure gradient and providing a longitudinal translational movement of the medium are included above all in this process. Forward spherical waves follow from them, which appear during the slowing of the flow at the start of its motion. They propagate over the entire flow and, reflected from the opposite wall, return to the region where they appeared. A thin layer of these waves moves along the wall without reflection but they interact with the higher density medium with the distribution $v(\rho)$. The spherical reflected and secondary waves take part in the interaction during the reflection of each wave. The secondary wave damps out quickly and does not exert a strong influence on the velocity fluctuations in the flow.

As a result, a pressure fluctuation field is formed in the boundary region. The extremal points of this field form perturbations which create a velocity fluctuation field at the center of the flow as well as a pressure fluctuation field at the boundary of the flow, forming the velocity profile $v(Z)$ and supporting the undamped cyclicity of the process. Its intensity depends on the average flow velocity and the damping time of the wave perturbation in the medium. From this viewpoint the boundary layer corresponds to a region in which the pressure fluctuations prevail over velocity fluctuations.

The distribution dynamics of the perturbations over the wall surface and with time is apparently supported by calculation and it can be considered determinate. The problem itself, like the problem of the emission of compression waves by a flame, belongs to the class of irreducibles. At the same time, the character of the flow at the center of the stream makes it possible to consider the fluctuations as pseudochaotic and to use statistical methods for describing them.

Within the framework of this mechanism, in which an alternation of phases dominates because of multiple reflections of the waves from the boundaries, an explanation of turbulence properties requires no postulates whatsoever, and the role of such parameters as the Reynolds number, viscosity of the medium, wall roughness, the cascade process for the transfer of energy from large-scale fluctuations to small-scale, the spectral characteristics of the fluctuations, the turbulence of the inflowing stream acquires a natural physical explanation.

[1] L. Prandtl, *Mechanics of Viscous Liquids. Aerodynamics*, ed. by V. F. Durand (Oborongiz, Moscow, 1939), p. 47.
[2] L. D. Landau and E. M. Lishitz, *Fluid Mechanics*, 2nd edition (Pergamon Press, Oxford, 1987; Nauka, Moscow, 1986), p. 3.
[3] H. Schlichting, *Formation of Turbulence* (Foreign Literature Press, Moscow, 1962), p. 7.
[4] A. N. Kolmogorov, Izv. Akad. Nauk SSSR, Ser. Fiz. **6**, 3 (1972).
[5] D. Ruelle and F. Takens, Commun. Math. Phys. **20**, 167 (1971).
[6] A. S. Monin, Usp. Fiz. Nauk **125**, 97 (1978) [Sov. Phys. Usp. **21**, 429 (1978)].
[7] M. I. Rabinovich, Usp. Fiz. Nauk **125** 123 (1978) [Sov. Phys. Usp. **21**, 443 (1978)].
[8] H. A. Lorentz, *Über die Enstehung turbulenter Flussigkeitsbewegungen und über den Einfluss dieser Bewegungen bei der Stromung durch Rohren. Abh. theor. Physik* (Teubner, Leipzig, 1907), p. 43.
[9] L. N. Pyatnitskiĭ, Dokl. Akad. Nauk SSSR **144**, 1262 (1962) [Sov. Phys. Dokl. **7**, 495 (1962)].
[10] L. N. Pyatnitskiĭ, Dokl. Akad. Nauk SSSR **176**, 631 (1967).

Translated by Eugene R. Heath

Nuclear β-decay and photoproduction of e^+e^- pairs in intense electromagnetic fields with a complex configuration

V. N. Rodionov

Moscow State Geological Prospecting Academy, Department of General Physics, 117873 Moscow, Russia
(Submitted 19 May 1997)
Zh. Éksp. Teor. Fiz. **113**, 21–42 (January 1998)

The effect of an intense electromagnetic field formed by the superposition of a constant magnetic field and a laser-type field on nuclear β-decay and on pair production by two γ-rays with different polarizations is studied. Time integral representations are obtained for the total probabilities of these processes without restrictions on the strengths of the fields making up the configuration. Despite the different nature of these reactions, in the nonrelativistic limit these expressions contain similar dependences on the field characteristics and the differences reduce to different power-law singularities in the behavior of the integrands at zero. At low fields, complete asymptotic expansions of the probabilities of these processes, including perturbation theory terms and oscillatory field contributions, are obtained in terms of parameters characterizing the fields. It is shown that the oscillatory corrections can be enhanced owing to the effect of a combination external field. The analysis of the probabilities of the various processes given here in terms of nonlinear functions of the field is illustrated by numerical calculations and graphs. © *1998 American Institute of Physics.* [S1063-7761(98)00301-1]

1. INTRODUCTION

As is well known, theoretical studies of the decay and creation of particles in intense external fields have been under way for a long time. (See Refs. 1–7 and the numerous papers cited there.) The distinctive feature of these processes is that they take place in the absence of a field. Thus, the contribution of an external field to these processes shows up through invariant parameters, which, besides the field characteristics, include the maximum energy transferred to the product particles. It is easy to see that, from the standpoint of possible experiments for isolating the effect of the field against the background of the characteristics of the processes *in vacuo*, studies of nonrelativistic reactions are of the greatest interest since the role of the external field becomes greater as the energy is lowered.

Based on studies of the effect of plane wave fields F on the decay of elementary particles,[6–16] it has been established that in the nonrelativistic limit at sufficiently low frequencies the fundamental parameter is

$$\chi_E = F/F_0, \tag{1}$$

where the characteristic field is

$$F_0 = (2I)^{3/2}\frac{m^2}{e}, \tag{2}$$

e and m are the charge and mass of the electron, and I is the maximum kinetic energy of a decay electron expressed in terms of its rest energy. (Here and in the following a system of units with $\hbar = c = 1$ is used.) In the region $\chi \gg 1$ the effect of the field dominates, since in this case the processes develop over characteristic times and lengths which depend on the field and their probabilities have an essentially nonlinear dependence on the field.[1–5]

In the limit $\chi \ll 1$ the effect of the field shows up as corrections to the probabilities for the unperturbed process; however, here a nontrivial field dependence is also possible.[6,9,12,13,16] It should be noted that a different type of electromagnetic field acting on the process generally leads to a different value of the parameter characterizing the field. Thus, in particular, in the case where a constant magnetic field H affects decay, the fundamental parameter has the form

$$\chi_H = H/H_0 = \frac{\mu}{2I}, \tag{3}$$

where $\mu = eH/m^2$, and the characteristic field in the nonrelativistic case is given by[10,16]

$$H_0 = 2I\frac{m^2}{e}. \tag{4}$$

The difference between the parameters (2) and (4) can be shown qualitatively based on the following simple considerations: in the case of a constant plane wave field, with an electric component among its constituents, the parameter χ_E is given by the work done by the field over a distance $\lambda_d/2$ (λ_d is the minimum De Broglie wavelength of the decay electron) divided by the maximum kinetic energy released in the decay.

In a constant magnetic field H, which performs no work, the role of the analogous parameter χ_H is played by the ratio of the energy of the shift in the frequency of the oscillations of the charged oscillator in the magnetic field,[1]

$$\Delta\varepsilon = \frac{eH}{2m}, \tag{5}$$

to the maximum energy of a decay electron.

In the case of an electromagnetic wave field, the parameter (1) is supplemented by a parameter containing the frequency ω of the field:[1–20]

$$\xi = \frac{eF}{\omega m (2I)^{1/2}}. \tag{6}$$

This parameter (6) can be interpreted as the ratio of the so-called oscillator momentum of the electron in the wave[1] to the maximum momentum acquired by the electron during decay. The difference in the field parameters defined in terms of the energy or momentum shows up additionally in the fact that the quantities (1) and (3) are proportional to Planck's constant, while ξ is a purely classical parameter. In view of its definition, the role of ξ can show up noticeably prior to integration over the momenta of the charged particles of these reactions. In this sense, the dependence of the probabilities of these processes on the parameter (6) is limited by their differential distributions, but the total reaction probabilities are determined exclusively by χ_E or by a combination with (6), which indicates an explicit frequency dependence which vanishes as ω goes to 0. Thus, the strong dependence of the total decay probabilities in an intense radiation field on ξ reported in some papers[17,18] was completely wrong, as first pointed out in Ref. 8.

Evidently, for electromagnetic waves with a more complicated structure, the number of field parameters can only increase. Thus, in particular, in the case where decay processes are affected by an external field containing both a constant magnetic field and an electromagnetic wave field,[16] the probabilities of these processes also depend on the parameter

$$\delta = 1 - \frac{eH}{m\omega},$$

which represents the relative detuning of the wave frequency ω from the cyclotron frequency $\omega_H = eH/m$.

It is also extremely important that in electromagnetic fields $F \ll F_0$, besides power law expansions in terms of the small parameter χ, the total probabilities also contain so-called oscillatory contributions[1–6,9,13,16,20] with an essential singularity at the point $F = 0$ that was first pointed out in Ref. 6. Such behavior by these processes requires the use of adequate methods for studying them, since these effects cannot, in principle, be obtained by perturbation theory techniques. In this regard it should be noted that, despite the popularity of accounting for the effect of external fields using the Furry representation, each specific form of the field configuration, in general, requires the development of specific techniques for studying it. (See Refs. 1–8, for example.)

A study has been made[13] of the formation of a nonrelativistic e^+e^- pair by two γ-rays in the presence of a circularly polarized wave and of low-energy β-decay in an electromagnetic field with the same structure. These processes shared a dependence of their reaction probabilities on the field characteristics prior to integration with respect to time in the time integral representation. The differences showed up only in different power law dependences for the time dependence in the integrands. This was explained by a difference in the phase volumes of the product particles from these reactions.

We have recently investigated[16] nuclear β-decay in an electromagnetic field with a complicated structure. The main result of that paper was a relatively compact integral representation for the total probability of β-decay taking into account the effect of the superposition of a constant magnetic field and the field of a plane electromagnetic wave propagating along it. It was also shown there[16] that in the nonrelativistic limit the total probability can be represented in the form of a single integral with respect to time, essentially without restriction on the strengths of the fields included in the superposition. Previously[19] we have obtained the probability of electron-positron pair production by an external photon propagating counter to an electromagnetic wave in a constant magnetic field. The orientation of the magnetic field coincided with the propagation direction of the wave and the probability of the process was written in the form of a double integral, in which one of the integrations corresponded to the Fock-Schwinger proper time representation.

The single-photon e^+e^- pair production reaction studied in Ref. 19 can be generalized to the case of two-photon pair production, so it becomes possible to make a direct comparison of the probabilities of creation of a nonrelativistic pair and of nuclear β-decay in an electromagnetic field with an extremely complicated structure.

In this paper we study the probabilities of processes owing to interactions of a different nature, but their field dependence has a number of similar features. We have studied expressions for the probabilities of both allowed and forbidden β-decay processes, as well as of pair formation by photons with different polarizations. The analytical expressions for the probabilities are analyzed over a wide range of variation of the parameters characterizing the field, and a number of asymptotic expressions are obtained which illustrate the existence of ''monotonic'' as well as ''oscillatory'' contributions to the field dependences.

It should be noted that forbidden β-decay in an intense electromagnetic filed has been studied previously in Ref. 15, which refers to yet earlier papers devoted to this topic that turned out to be wrong. In this paper we make a detailed analysis of the effect of a strong electromagnetic wave on the unique β-transitions of the first forbiddenness class and make estimates involving two possible mechanisms for the external field interaction associated with direct transitions of the nucleus from the initial (i) to the final (f) state, as well as for a transition through a virtual nuclear state. The second mechanism involves the removal of the prohibition owing to the external field and depends strongly on the energies $\Delta\varepsilon_1 = \varepsilon_1 - \varepsilon_i$ or $\Delta\varepsilon_2 = \varepsilon_2 - \varepsilon_f$, where the excited states of nuclei 1 and 2 have a momentum and parity that permit an allowed β-transition. The quantities $\Delta\varepsilon_{1,2}$ are actually on the order of tens or hundreds of keV and can be dominant in a resonance situation, when the energies of the wave photons are comparable to the transition energies. If the field is not resonant, in the case of real laser intensities the main contribution to the total probability is from transitions of the first type. Retaining the approach used to examine forbidden β-

decay in Ref. 15, we have evaluated the effect of direct transitions owing to the effect of the constant field on this process.

2. PROBABILITIES IN A MAGNETIC FIELD AND IN A WAVE FIELD

The earlier work[19] examined the formation of electron-positron pairs by circularly polarized external photons with energy ω' propagating counter to the photons in a circularly polarized wave in a uniform constant magnetic field. To derive expressions for the probability of this process it was assumed that the wave propagated along the magnetic field. If we assume that in an electromagnetic field with this configuration a photon ω'' propagates counter to the photon ω', then after the required calculations and transformations of the parameters, the basic formula for the probability of single-photon pair creation[19] yields an analogous expression for the two-photon process in the form of a double integral. Let us discuss the case of pair formation near the threshold, assuming that

$$\omega'+\omega''=2m+Im, \tag{7}$$

where, as before, Im represents the maximum kinetic energy released in the process, i.e., we shall consider the nonrelativistic limit in accordance with $I\ll 1$. In this case we can take the integral over the energy variable essentially without restrictions on the strengths of the electromagnetic fields which make up the configuration. The probability of pair production depends significantly on the polarization of the γ-photons. The calculations for the case in which the circular polarizations of the photons coincide yield the probability

$$W_{3/2}=G_{3/2}\mu\int_{-\infty}^{\infty}\frac{\cot(\mu\rho)}{\rho^{1/2}}\,e^{iS}d\rho, \tag{8}$$

where

$$S=\frac{2I}{\lambda\delta}\,\xi^2\left\{x\left(\frac{\delta}{\xi^2}-\frac{1}{2-\delta}\right)+\frac{2(1-\delta)}{\delta(2-\delta)^2}\,\frac{\sin[x(2-\delta)]\sin(x\delta)}{\sin[2x(1-\delta)]}\right\},$$

$$x=\frac{\rho\lambda}{2},\quad \lambda=\frac{\omega}{m},\quad \mu=\frac{eH}{m^2}, \tag{9}$$

and the remaining notation has been introduced before.

For γ-photons with opposite polarizations, the probability of pair production is given by

$$W_{5/2}=G_{5/2}\mu^2\int_{-\infty}^{\infty}\frac{e^{iS}}{\rho^{1/2}\sin^2(\mu\rho)}\,d\rho. \tag{10}$$

In Eqs. (8) and (10) the constants in front of the integrals are related to the corresponding vacuum probabilities $W_0(\nu)$ for the processes:

$$G_\nu=W_\nu(0)\exp\left[\frac{i(\nu-1)\pi}{2}\right]\frac{I^{1-\nu}}{2\Gamma(1-\nu)}, \tag{11}$$

where $\nu=3/2$ and $\nu=5/2$ correspond to pair production by photons with the same polarizations and when the polarizations of the γ-photons are different, respectively. Equation (11) also retains its significance for $\nu=9/2$ and $\nu=11/2$, which correspond to allowed and forbidden β-decay in these cases.

For the case of allowed β-decay ($\nu=9/2$) with a small energy release, the probability can be written in the form[16]

$$W_{9/2}=G_{9/2}\left(\frac{\mu}{2}\right)\int_{-\infty}^{\infty}\frac{\cot(\mu\rho/2)}{\rho^{7/2}}\,e^{iS}d\rho, \tag{12}$$

where

$$S=\xi^2\,\frac{2I}{\lambda\delta}\left\{x\left(\frac{\delta}{\xi^2}-1\right)+\frac{1-\delta}{\delta}\,\frac{\sin x\,\sin(x\delta)}{\sin[x(1-\delta)]}\right\}.$$

Note that in these integrals all the singularities are shifted into the upper half plane, i.e., the parameters receive a small negative imaginary contribution which subsequently goes to zero in the limit.

It is also easy to see from a comparison of Eqs. (8), (10), and (12) that in the case of an electromagnetic field with a complicated configuration, we are dealing, as before, with very similar expressions for the integrands in the time representations of the probabilities for the different processes. These analogies become even more evident when we study the various limiting values of the field parameters upon which the probabilities depend.

Thus, in the limit $\mu\to 0$, $\delta\to 1$ Eq. (8) corresponds to the probability of formation of a nonrelativistic e^+e^- pair by two γ-photons with the same polarization in the presence of a circularly polarized wave and coincides with a similar expression that was studied in detail in Ref. 13. There an analogy was noted between the probabilities of pair formation and low-energy β-decay of tritium in the field of an electromagnetic wave. The field dependence under the integral sign was common to both, while the difference reduced to the characteristic replacement of the square of the charge, e^2, by $2e^2$ and a change in the power-law singularity at zero on integrating with respect to time ($t^{-3/2}$ for pair production and $t^{-9/2}$ for β-decay).

It should be noted that similar substitution rules for the limit $\mu\to 0$, $\delta\to 1$ (which corresponds to the complete elimination of the constant magnetic field from this configuration) can be obtained in Eq. (10), but the power-law factor in the time dependence for the case of pair production by γ-photons of opposite polarization is different:

$$W_{5/2}=G_{5/2}\left(\frac{\lambda}{2}\right)^{3/2}\int_{-\infty}^{\infty}\frac{dx}{x^{5/2}}\exp\left\{i\,\frac{2I}{\lambda}\left[x(1-2\xi^2)+2\xi^2\,\frac{\sin^2 x}{x}\right]\right\}. \tag{13}$$

If we make the replacement

$$2\xi^2\to\xi^2,$$

then the formal dependence of the exponent on ξ remains the same as in the case of β-decay in the field of a circularly polarized wave. Note that, as before, the singularity in Eq. (13) at zero is bypassed below.

Another limiting case, $\xi\to 0$, which corresponds to eliminating the field of the electromagnetic wave from the

superposition of electromagnetic fields, yields the probabilities for e^+e^- pair production by two γ-photons in a constant magnetic field. Here if we place no restrictions on the magnetic field strength, Eqs. (8)–(11) yield

$$W_\nu = G_\nu \mu \int_{-\infty}^{\infty} \frac{\cot(\mu\rho)}{\rho^{\nu-1}}\, e^{i\rho I} d\rho, \tag{14}$$

where $\nu = 3/2$ corresponds to pair formation by two photons with identical polarizations and $\nu = 9/2$, to β-decay. In this case a qualitative convergence of the expressions is attained by making the replacement $\mu/2 \to \mu$ or substituting $2e$ in the case of pair production in place of e for β-decay. It is interesting that in both the wave and the magnetic field, a transition to the pair production process leads to a formal increase in the interaction between the charge and field; however, the degrees of increase are different.

Two-photon pair production in a magnetic field has been studied previously.[19–23] The quantity of interest in this case, however, is not the probability but the cross section for the given reaction. Retaining the nonrelativistic approximation ($I \ll 1$), we obtain the following cross section for pair production by photons with identical polarizations from Eq. (14):

$$\sigma^+ = \frac{1}{2}\, \mu r_0^2 \sqrt{\pi} e^{-3\pi i/4} \int_{-\infty}^{\infty} \frac{\cot(\mu\rho)}{\rho^{1/2}}\, e^{i\rho I} d\rho, \tag{15}$$

where r_0 is the classical radius of the electron.

The analogous expression for pair production by two γ-photons with opposite polarizations has the form

$$\sigma^- = r_0^2 \mu^2 \sqrt{\pi} e^{3\pi i/4} \int_{-\infty}^{\infty} \frac{d\rho}{\rho^{1/2} \sin^2(\mu\rho)}\, e^{i\rho I}. \tag{16}$$

In the limit $\mu \ll I$, on expanding $\cot x$ at zero and including the contributions to the integral from the poles $x = \pi n$, where $n = \pm 1, \pm 2, \cdots$, from Eq. (15) we obtain

$$\sigma_H^+ = \sigma_-^+ + \sigma_\sim^+, \tag{17}$$

where

$$\sigma_-^+ = \sigma_0^+ \sum_{k=0}^{\infty} \frac{\Gamma(3/2) 2^{2k} B_{2k}}{(2k)!\,\Gamma(3/2-2k)} \left(\frac{\mu}{I}\right)^{2k}, \tag{18}$$

and

$$\sigma_\sim^+ = \frac{\sigma_0^+}{2} \left(\frac{2\mu}{I}\right)^{1/2} \zeta\left(1/2, \left\{\frac{I}{2\mu}\right\}\right). \tag{19}$$

Here the B_{2k} are the Bernoulli numbers, $\zeta(1/2,\{v\})$ are the generalized Riemann zeta functions,[24] $\{v\}$ is the fractional part of the number v,

$$\sigma_0^+ = 2\pi r_0^2 I^{1/2} \tag{20}$$

is the pair production cross section in vacuum in the nonrelativistic limit (the Breit-Wheeler cross section), and σ_-^+ and $\sigma_\sim^+$ are the so-called monotonic and oscillatory contributions to the cross section.

Limiting ourselves to a few terms in the expansion (18), we obtain

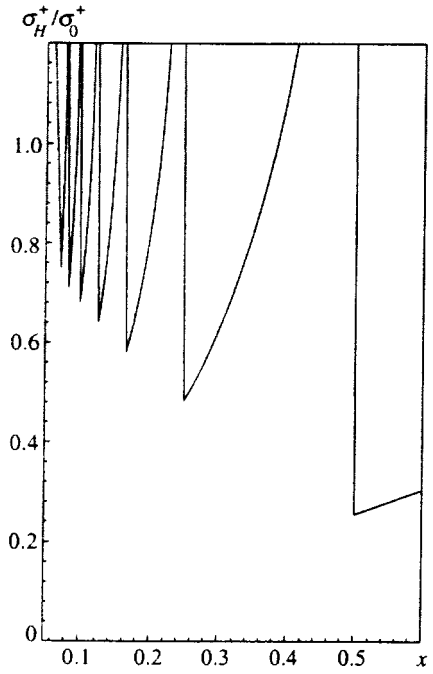


FIG. 1. The cross section for pair production by two γ-photons with identical polarizations, σ_H^+, normalized to the vacuum cross section as a function of the magnetic field strength $x = \mu/I$.

$$\sigma_-^+ = \sigma_0^+ \left\{ 1 - \frac{\mu^2}{12I^2} + \frac{\mu^4}{48I^4} - \frac{\mu^6}{32I^6} + \frac{143\mu^8}{1280I^8} + \cdots \right\}. \tag{21}$$

The first term in the expansion (21) coincides with a result of Ref. 20 obtained by other means. It is extremely interesting that the root singularity in the oscillatory contribution noted previously[20,21] can be studied numerically using the approach we have developed. Figure 1 shows a plot of σ_H^+/σ_0^+ as a function of the parameter $x = \mu/I$, which manifests the explicit oscillatory behavior of the cross section with a divergence at the points $I/2\mu \to n - 0$, where $n = 1,2,\cdots$. For these values of the parameters μ and I the product electron and positron have a zero component of the momentum along the magnetic field. In this approach we have neglected the Coulomb interaction between the e^+ and e^-, but under the conditions described here it turns out to be the most important process, and the behavior of the cross section under resonance conditions requires further study. In Ref. 20 the cross section for pair production in a magnetic field was estimated taking into account the broadening of the Landau levels owing to synchrotron radiation. Note, also, that at the points $2\mu/I \to 1/n + 0$ the cross section σ_H^+/σ_0^+ is finite and takes on a decreasing sequence of values with a minimum of 0.25 at the point $\mu = I/2$, after which the oscillations in the cross section cease and a linear rise is observed with increasing field:

$$\sigma_H^+ = \sigma_0^+ \frac{\mu}{2I}. \tag{22}$$

When the photons have opposite polarizations, the easily derived expansion

$$\frac{1}{\sin^2 z}=\sum_{n=0}^{\infty}(-1)^n\frac{(1-2n)2^{2n}B_{2n}}{(2n)!}z^{2n-2}, \tag{23}$$

in Eq. (16) yields

$$\sigma_H^-=\sigma_-^-+\sigma_\sim^-, \tag{24}$$

where

$$\sigma_-^-=\sigma_0^-\sum_{n=0}^{\dddot{=}}\Gamma(5/2)\frac{2^{2n}(1-2n)B_{2n}}{(2n)!\Gamma(5/2-2n)}\left(\frac{\mu}{I}\right)^{2n}, \tag{25}$$

and

$$\sigma_\sim^-=\sigma_0^-\frac{3}{4}\left[\left(\frac{2\mu}{I}\right)^{1/2}\zeta\left(\frac{1}{2},\left\{\frac{I}{2\mu}\right\}\right)-\left(\frac{2\mu}{I}\right)^{3/2}\times\zeta\left(-\frac{1}{2},\left\{\frac{I}{2\mu}\right\}\right)\right]. \tag{26}$$

The notation follows the same principles as that employed for σ^+ in Eqs. (24)–(26), with

$$\sigma_0^-=\frac{8}{3}\pi r_0^2 I^{3/2}. \tag{27}$$

Retaining only some of the terms in the expansion in Eq. (25), we have

$$\sigma_-^-=\sigma_0^-\left(1-\frac{\mu^2}{4I^2}+\frac{3\mu^4}{80I^4}-\frac{5\mu^6}{96I^6}+\frac{231\mu^8}{1280I^8}+\cdots\right). \tag{28}$$

The expansion term $\sim\mu^2/I^2$ in Eq. (28) was cited previously in Ref. 20, in which the oscillatory contribution (26) was also calculated. However, that result[20] does not contain the term owing to $\zeta(-1/2),\{v\})$, whose contribution is decisive when μ reaches or exceeds $I/2$. In this case, the cross section σ^- goes to zero, as indicated by the plot of σ_H^-/σ_0^- as a function of $x=\mu/I$ in Fig. 2.

Asymptotic expressions for the probability of β-decay in an intense magnetic field (i.e., the case $\nu=9/2$) were calculated in detail in our earlier paper.[16] Here we limit ourselves to illustrating the field dependence of $W_{9/2}/W_{9/2}(0)$ in Fig. 3. The dotted curve corresponds to a calculation with rough accuracy. Increased computational accuracy leads to an extremely interesting stepwise dependence. This kind of behavior in the total probabilities for the processes in a magnetic field is a consequence of including (with a reduction in the field strength for constant energy release during decay) all the new channels and new partial channels characterized by an increase in the number of Landau levels.

In this regard, it is interesting that for pair production processes by photons with different polarizations it is possible to obtain expressions which explicitly take into account the appearance of new contributions as the field strength is reduced,

$$\sigma_H^+=\sigma_0^+\left[\frac{\mu}{2I}+\frac{\mu}{I}\sum_{n=1}^{N}\frac{1}{(1-2\mu n/I)^{1/2}}\right], \tag{29}$$

and

$$\sigma_H^-=\sigma_0^-\frac{3\mu^2}{I^2}\sum_{n=1}^{N}\frac{n}{(1-2\mu n/I)^{1/2}}, \tag{30}$$

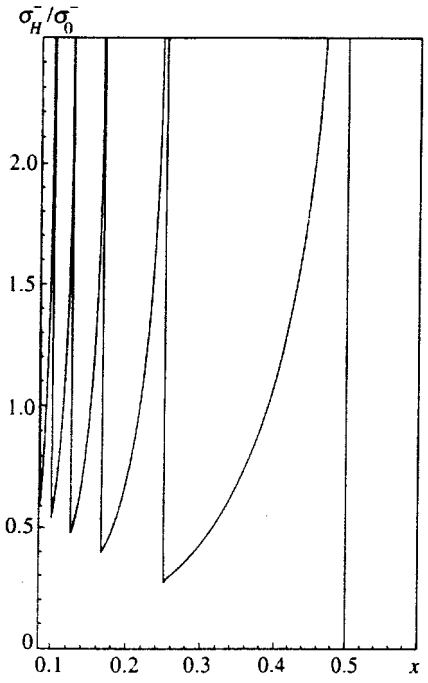


FIG. 2. The cross section for pair production by two γ-photons with opposite polarizations, σ_H^-, normalized to the vacuum cross section as a function of the magnetic field strength $x=\mu/I$.

where $N=[I/2\mu]$ is the integral part of the number $I/2\mu$. These last formulas indicate that, as μ/I reaches 1/2 the cross section σ^+ increases linearly as the field rises, while σ^- goes identically to zero.

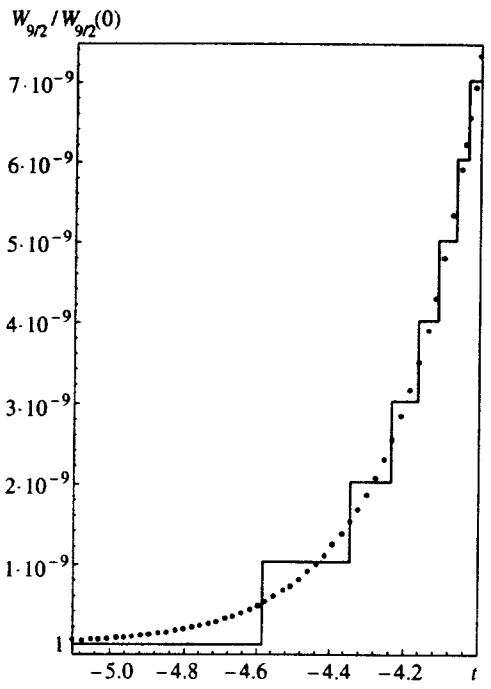


FIG. 3. The probability W_H of β-decay normalized to the vacuum probability as a function of the magnetic field $t=\log(\mu/I)$. The dotted curve is a rough estimate.

3. PROBABILITIES FOR LOW MAGNETIC FIELDS WITH ARBITRARY PLANE WAVE FIELD STRENGTHS

The region of relatively low magnetic fields, $\mu \ll I$, is characterized by the fact that in this case the main contribution to the integrals (8)–(10) is from the neighborhood of the zero point. Given this, we expand the integrands, retaining terms up to and including x^7:

$$S_i = \frac{2I}{\lambda}\left(x - \frac{\xi^2}{3}\, x^3 + \xi^2 x^5 A_i + \xi^2 x^7 B_i \right), \tag{31}$$

where $i = 1,2$. S_1 applies to β-decay and S_2 to pair production processes. The functions A_i and B_i depend on the magnetic field and wave frequency, which are combined in the parameter δ:

$$A_1 = -\frac{\delta^2}{45} + \frac{4\delta}{45} - \frac{1}{45}, \quad A_2 = -\frac{2\delta^2}{15} + \frac{4\delta}{15} - \frac{4}{45}, \tag{32}$$

and

$$B_1 = -\frac{2\delta^4}{945} + \frac{4\delta^3}{315} - \frac{23\delta^2}{945} + \frac{4\delta}{315} - \frac{2}{945},$$

$$B_2 = -\frac{17\delta^4}{315} + \frac{68\delta^3}{315} - \frac{284\delta^2}{945} + \frac{32\delta}{189} - \frac{32}{945}. \tag{33}$$

It is quite remarkable that in Eq. (31) the terms proportional to x^3 do not contain any dependence on field parameters other than ξ^2. Beginning with x^5, besides the dependence on ξ^2, the parameter δ, which characterizes the dependence of the probabilities on the frequency characteristics of the combined field, ''comes into play.'' If we retain only the terms up to x^3, inclusively, in the exponent while the remaining terms are written in the form of an expansion, then we can obtain a series which contains the limit of a constant crossed field as the zeroth approximation.

Writing the trigonometric functions in the preexponential factors in the integrals of Eqs. (8)–(10) as expansions in the neighborhood of the zero point yields yet another series containing other parameters. It is important that both these and the other estimates of the integrals (8)–(10) can be made assuming $\mu \ll I$ without restrictions on the parameter χ. Since the expressions obtained by simultaneous expansion of the exponential and trigonometric functions are so cumbersome, we shall do this in two stages. First, we examine the contributions to the integral of Eq. (8) owing to the expansion of $\cot x$ about zero, retaining only terms up to and including x^3 in the exponential. In this case the integral (8) can be written in the form

$$W_{3/2} = G_{3/2} I^{1/2} \chi^{1/3} \int_{-\infty}^{\infty} \frac{dz}{z^{3/2}} \exp[\, i(z\chi^{-2/3} - z^3/3)\,] \times \left(1 - \frac{\mu^2 z^2}{3I^2\chi^{4/3}} - \frac{\mu^4 z^4}{45 I^4 \chi^{8/3}} \right). \tag{34}$$

Using the notation

$$I_{(\nu)} = \int_{-\infty}^{\infty} dz\; z^{\nu} \exp[\, i(z\chi^{-2/3} - z^3/3)\,], \tag{35}$$

for the probability of pair production, we shall have

$$W_{3/2} = G_{3/2} I^{1/2} \chi^{1/3} \left\{ I_{(-3/2)} - \frac{\mu^2}{3I^2\chi^{4/3}}\, I_{(1/2)} - \frac{\mu^4}{45 I^4 \chi^{8/3}}\, I_{(5/2)} \right\}, \tag{36}$$

where

$$I_{(-3/2)} = \frac{2iA}{2^{2/3}\chi^{1/3}} \left(\sqrt{y}\,\Phi^2 + \frac{\Phi'^2}{\sqrt{y}} \right), \tag{37}$$

$$I_{(1/2)} = i 2^{1/3} A \Phi \Phi', \tag{38}$$

and

$$I_{(5/2)} = \frac{Ai}{2}\,(4y\Phi\Phi' - \Phi^2), \quad A = \frac{4 e^{i\pi/4}}{2^{1/3}\sqrt{\pi}}, \tag{39}$$

while Φ and Φ' are the Airy function and its derivative with respect to the argument $(-y)$, with $y = 1/(2\chi)^{2/3}$. Thus, for the probability of pair production by two γ-photons with the same polarizations we have

$$W_{3/2} = \frac{W_{3/2}(0)}{\pi} \left\{ K_0^+ - \frac{\mu^2}{3I^2}\, K_1^+ + \frac{\mu^4}{90 I^4}\, K_2^+ \right\}, \tag{40}$$

where

$$K_0^+ = \sqrt{y}\,\Phi^2 + \frac{\Phi'^2}{\sqrt{y}}, \quad K_1^+ = \frac{1}{\chi}\,\Phi\Phi', \tag{41}$$

and

$$K_2^+ = \frac{1}{\chi^2} \left(\sqrt{y}\,\Phi^2 - \frac{2}{\chi}\,\Phi\Phi' \right). \tag{42}$$

Note that when the oscillatory terms are neglected as $\chi \to 0$

$$K_0^+ \to \pi, \quad K_1^+ \to \frac{\pi}{4}, \quad K_2^+ \to \frac{15}{8}\,\pi, \tag{43}$$

and Eq. (40) yields the expansion (21) characterizing two-photon pair production in a magnetic field.

Similar calculations for $\nu = 5/2$ and $\nu = 9/2$ give

$$W_{5/2} = G_{5/2} I^{3/2} \chi \left\{ I_{(-5/2)} + \frac{\mu^2}{3I^2\chi^{4/3}}\, I_{(-1/2)} + \frac{\mu^4}{15 I^4 \chi^{8/3}}\, I_{(3/2)} \right\}, \tag{44}$$

and

$$W_{9/2} = G_{9/2} I^{7/2} \chi^{7/3} \left\{ I_{(-9/2)} - \frac{\mu^2}{12I^2\chi^{4/3}}\, I_{(-5/2)} - \frac{\mu^4}{720 I^4 \chi^{8/3}}\, I_{(-1/2)} \right\}, \tag{45}$$

where

$$I_{(-5/2)} = -\frac{2^{4/3} A}{3\chi} \left(\sqrt{y}\,\Phi^2 + \frac{\Phi'^2}{\sqrt{y}} - \chi\Phi\Phi' \right), \tag{46}$$

$$I_{(-1/2)} = A\Phi^2, \quad I_{(3/2)} = A 2^{-1/3} \sqrt{y} \left(\sqrt{y}\,\Phi^2 - \frac{\Phi'^2}{\sqrt{y}} \right).$$

$$I_{(-9/2)}=\frac{16A}{105\,2^{2/3}\chi^{7/3}}\left[\sqrt{y}\Phi^2\left(1+\frac{21}{4}\chi^2\right)+\frac{\Phi'^2}{\sqrt{y}}\times\left(1+\frac{15}{4}\chi^2\right)-\chi\Phi\Phi'\right].$$

Finally, for the probabilities with $\nu=5/2$ and $\nu=9/2$, we obtain

$$W_{5/2}=\frac{W_{5/2}(0)}{\pi}\left\{K_0^-+\frac{\mu^2}{2I^2}K_1^-+\frac{\mu^4}{20I^4}K_2^-\right\},\tag{47}$$

and

$$W_{9/2}=\frac{W_{9/2}(0)}{\pi}\left\{K_0+\frac{35}{12}\left(\frac{\mu}{2I}\right)^2K_1-\frac{7}{24}\left(\frac{\mu}{2I}\right)^4K_2\right\},\tag{48}$$

where

$$K_2^-=\chi^{-2}\left(\sqrt{y}\Phi^2-\frac{\Phi'^2}{\sqrt{y}}\right),$$

$$K_0^-=K_1=\sqrt{y}\Phi^2+\frac{\Phi'^2}{\sqrt{y}}-\chi\Phi\Phi',\quad K_1^-=K_2=\sqrt{y}\Phi^2,$$

$$K_0=\sqrt{y}\Phi^2\left(1+\frac{21}{4}\chi^2\right)+\frac{\Phi'^2}{\sqrt{y}}\left(1+\frac{15}{4}\chi^2\right)-\chi\Phi\Phi'.$$

Note that in the limit $\chi\to0$ when the oscillatory terms are neglected, Eqs. (47) and (48) also lead to the expansion for a purely magnetic field. (See Eq. (28) and the corresponding formula in Ref. 16.)

4. ASYMPTOTIC ESTIMATES OF THE PROBABILITIES FOR $\chi\ll1$

We now consider the situation in which, along with contributions $\sim(\mu/2I)^{2k}$, where $k=0,1,2,\cdots$, to the process probabilities there will be contributions owing to the expansion in terms of the parameters of a plane wave field. This corresponds to $\xi\gg1$ and, if we restrict ourselves in this case to the main contributions from the expansion in terms of $\mu/2I$ and expand the exponent in Eqs. (8)–(12) beginning with terms containing the parameter δ, then, for example, for the probability of β-decay we can obtain

$$W_{9/2}=G_{9/2}\left(\frac{\lambda}{2}\right)^{7/2}\int_{-\infty}^{\infty}\frac{dx}{x^{9/2}}\exp\left[i\,\frac{2I}{\lambda}\left(x-\frac{\xi^2x^3}{3}\right)\right]\times\left[1+i\,\frac{2I}{\lambda}\,\xi^2(x^5A_1+x^7B_1)\right].\tag{49}$$

The characteristic integrals in Eq. (49) can also be reduced to Airy functions, whose properties in this limit have been analyzed in detail elsewhere.[13] In our earlier paper[16] a method was developed for investigating similar integrals for the case of β-decay in an intense electromagnetic field based on using integral Mellin transforms.[5,7,8,16,24] Using that method, we can obtain closed formulas for the asymptotic behavior of the bilinear combinations of Airy functions $\Phi^2(-y)$, $\Phi'^2(-y)$, and $\Phi(-y)\Phi'(-y)$ as $y\to\infty$, where $y=1/(2\chi)^{2/3}$. Thus, it is possible to compare directly the results obtained by different methods for studying decay processes and pair production in the overlap region for the parameters of a purely plane wave field and for the parameters characterizing the superposition of a wave field and magnetic field. Here an analysis of pair production by two oppositely polarized γ-photons, which was not discussed in Ref. 13, may be of independent interest.

Without dwelling on the details of these rather lengthy transformations, we write down the final results in the form

$$\Phi^2(-y)=\frac{\pi}{2\sqrt{y}}\sum_{n=0}^{\infty}(-1)^n\chi^{2n}\left\{\frac{\Gamma(3n+1/2)}{\Gamma(1/2)\Gamma(n+1)3^n}+a_{2n}\sin\frac{2}{3\chi}-\chi a_{2n+1}\cos\frac{2}{3\chi}\right\},\tag{50}$$

$$\Phi'^2(-y)=\frac{\pi\sqrt{y}}{2}\sum_{n=0}^{\infty}(-1)^n\chi^{2n}\left\{\left[1+\frac{3}{2}\chi^2(6n+1)\times(2n+1)\right]\frac{\Gamma(3n+1/2)}{\Gamma(n+1)\Gamma(1/2)3^n}-\left(a_{2n}+\frac{9}{4}\chi^2d_n\right)\sin\frac{2}{3\chi}+\frac{3\chi}{2}\times\left(f_n-\frac{9}{4}\chi^2g_n\right)\cos\frac{2}{3\chi}\right\},\tag{51}$$

and

$$\Phi(-y)\Phi'(-y)=\frac{\pi}{2}\sum_{n=0}^{\infty}(-1)^n\chi^{2n}\times\left\{\chi\,\frac{\Gamma(3n+3/2)}{\Gamma(1/2)\Gamma(n+1)3^n}-\frac{3}{2}\chi b_n\sin\frac{2}{3\chi}-\left(a_{2n}+\frac{9}{4}\chi^2c_n\right)\times\cos\frac{2}{3\chi}\right\},\tag{52}$$

where we have introduced the notation

$$b_n=\frac{2}{3}a_{2n+1}-\left(2n+\frac{1}{3}\right)a_{2n},\quad c_n=\frac{4}{3}a_{2n+1}\left(n+\frac{2}{3}\right),$$

$$d_n=\left(\frac{16}{3}n+\frac{28}{9}\right)a_{2n+1}-\left(\frac{16}{3}n+\frac{2}{3}+8n^2\right)a_{2n},$$

$$f_n=-\left(8n+\frac{2}{3}\right)a_{2n}+\frac{2}{3}a_{2n+1},$$

$$g_n=\frac{16}{9}(3n+2)(n+1)a_{2n+1},$$

and the coefficients a_m are calculated using the formula[16]

$$a_m=\sum_{k=0}^{2m}\frac{(2k-1)!!}{3^{2m-k}(2m-k)!(2k)!!}\,\frac{\Gamma(3m-k+1/2)}{\Gamma(1/2)}.\tag{53}$$

With these expressions it is easy to calculate the asymptotic variations in the probabilities of the processes being studied here. In particular, for the case of β-decay, Eqs. (31) and (49) imply that

$$W_{9/2}=\frac{W_{9/2}(0)}{\pi}\left\{\sqrt{y}\Phi^2\left(1+\frac{21}{4}\chi^2+\frac{105}{16}B_1\frac{\chi^4}{\xi^4}\right)+\frac{15}{4}\chi^2\frac{\Phi'^2}{\sqrt{y}}\left(1+O\left(\frac{\chi^2}{\xi^6}\right)\right)-\chi\Phi\Phi'\times\left(1+\frac{105}{8}\frac{\chi^2}{\xi^2}A_1+\frac{105}{8}\frac{\chi^2}{\xi^4}B_1\right)\right\}. \tag{54}$$

On substituting the expansions (50)–(52) in Eq. (54), for the limit $\chi\ll 1$ we obtain

$$\frac{W_{9/2}}{W_{9/2}(0)}=1+\frac{35}{8}\chi^2+\frac{35}{128}\chi^4-\left(\frac{35}{32}\right)^2\chi^6-\frac{105}{32}\frac{\chi^4}{\xi^2}A_1+\frac{1575}{128}\frac{\chi^6}{\xi^4}B_1+\left(\frac{105}{16}\chi^4-\frac{603575}{1536}\chi^6-\frac{35}{64}\frac{\chi^4}{\xi^2}A_1+\frac{175}{64}\frac{\chi^4}{\xi^4}B_1\right)\times\sin\frac{2}{3\chi}+\left(-\frac{3115}{64}\chi^5+\frac{105}{16}\frac{\chi^3}{\xi^2}A_1+\frac{105}{16}\frac{\chi^3}{\xi^4}B_1\right)\cos\frac{2}{3\chi}. \tag{55}$$

Note that for $\delta=1$, which corresponds to a transition to the case of the field of an electromagnetic wave, Eq. (55) yields an expression which reproduces the results of Ref. 13 in the first terms of the expansion (up to and including χ^4). For fixed values of δ equal to 1/2 and 0, Eq. (55) transforms to the analogous expansions of Ref. 16. Here we point out that for $|\delta|\gg 1$ the parameters A_i increase as δ^2 with magnetic field, while the B_i increase as δ^4. This may mean that for $|\delta|\simeq\xi$ the main contribution to the oscillatory terms will be determined by corrections containing the frequency characteristics of the field.

By analogy, for pair production we obtain

$$W_{3/2}=\frac{W_{3/2}(0)}{\pi}\left\{\sqrt{y}\Phi^2\left[1+\frac{A_2}{2\xi^2}-\frac{B_2}{\xi^4}\left(\frac{7}{4}\chi^2-\frac{1}{2}\right)\right]+\frac{\Phi'^2}{\sqrt{y}}\times\left(1-\frac{A_2}{2\xi^2}-\frac{B_2}{2\xi^4}\right)+\chi\Phi\Phi'\left(\frac{3}{2}\frac{A_2}{\xi^2}+5\frac{B_2}{\xi^4}\right)\right\}, \tag{56}$$

and

$$W_{5/2}=\frac{W_{5/2}(0)}{\pi}\left\{\sqrt{y}\Phi^2\left(1-\frac{3A_2}{4}\frac{\chi^2}{\xi^2}-\frac{21B_2}{8}\frac{\chi^2}{\xi^4}\right)+\frac{\Phi'^2}{\sqrt{y}}\times\left(1+\frac{15B_2}{8}\frac{\chi^2}{\xi^4}\right)-\chi\Phi\Phi'\left(1-\frac{3A_2}{2\xi^2}-\frac{3B_2}{2\xi^4}\right)\right\}. \tag{57}$$

When $\chi\ll 1$, Eq. (56) yields

$$\frac{W_{3/2}}{W_{3/2}(0)}=1+\frac{1}{8}\chi^2-\frac{105}{128}\chi^4+\frac{25025}{1024}\chi^6+\frac{A_2}{\xi^2}\left(\frac{105}{32}\chi^4-\frac{45045}{256}\chi^6\right)-\frac{B_2}{\xi^4}\frac{10395}{28}\chi^6+\cos\frac{2}{3\chi}\times\left[-\frac{1}{2}\chi+\frac{1225}{576}\chi^3-\frac{57482425}{995328}\chi^5+\frac{A_2}{\xi^2}\times\left(-\frac{17}{24}\chi-\frac{1225}{20736}\chi^3\right)+\frac{B_2}{\xi^4}\left(-\frac{59}{24}\chi+\frac{1925}{20736}\chi^3\right)\right]+\sin\frac{2}{3\chi}\left[-\frac{17}{24}\chi^2+\frac{199115}{20736}\times\chi^4+\frac{A_2}{\xi^2}\left(\frac{1}{2}+\frac{35}{576}\chi^2-\frac{86975}{995328}\chi^4\right)+\frac{B_2}{\xi^4}\times\left(\frac{1}{2}-\frac{385}{576}\chi^2+\frac{9625}{995328}\chi^4\right)\right]. \tag{58}$$

Note that, as in the case of β-decay, for $\delta=1$ Eq. (58) yields a result that characterizes the effect of the electromagnetic wave field. These corrections to the probability for the unperturbed e^+e^- pair production process by two γ-photons with identical polarizations in this limit in the first terms of the expansion (monotonic contributions up to χ^4/ξ^2 and oscillatory to χ) also agree with the analogous results from Ref. 13. It should, however, be emphasized that the transition to the pair production case is characterized by an enhanced field effect owing to the specific features of the integral dependence in Eq. (12) compared to that in Eq. (8). This shows up through the fact that the oscillatory contributions exceed the perturbation theory corrections, as noted in Ref. 13. Including the total field in this case leads to a still clearer demonstration of the possibility of separating the oscillatory contributions, since it follows from Eq. (58), in particular, that for $|\delta|\sim\xi$ the term proportional to $\sin(2/3\chi)$ can, in general, be dominant.

Similar calculations for the case of pair production by two photons with different polarizations yield the result

$$\frac{W_{5/2}}{W_{5/2}(0)}=1-\frac{\chi^2}{8}+\frac{35}{128}\chi^4-\frac{5005}{1024}\chi^6+\frac{A_2}{\xi^2}\left(-\frac{45}{32}\chi^4+\frac{10395}{256}\chi^6\right)+\frac{B_2}{\xi^4}\frac{2835}{128}\chi^6+\cos\frac{2}{3\chi}\left[\frac{35}{16}\chi^3-\frac{805805}{13824}\chi^5+\frac{A_2}{\xi^2}\left(-\frac{3}{4}\chi+\frac{25}{384}\chi^3-\frac{9625}{663552}\chi^5\right)+\frac{B_2}{\xi^4}\left(-\frac{3}{4}\chi-\frac{35}{384}\chi^3-\frac{89425}{663552}\chi^5\right)\right]+\sin\frac{2}{3\chi}\left[-\frac{3}{4}\chi^2+\frac{3745}{384}\chi^4+\frac{A_2}{\xi^2}\left(-\frac{15}{16}\chi^2+\frac{575}{13824}\chi^4\right)+\frac{B_2}{\xi^4}\left(-\frac{35}{16}\chi^2+\frac{665}{13824}\chi^4\right)\right]. \tag{59}$$

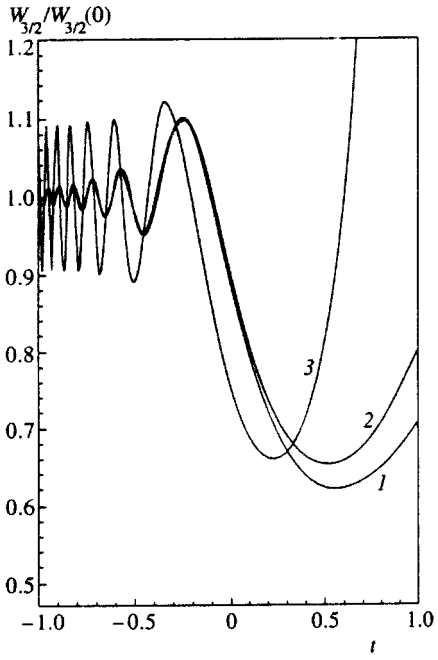


FIG. 4. The normalized probability of pair production, $W_{3/2}$ as a function of the parameter $t=(2/3)\log(2\chi)$ for different values of γ: $\gamma^2=0$ (*1*), 0.1 (*2*), 1 (*3*).

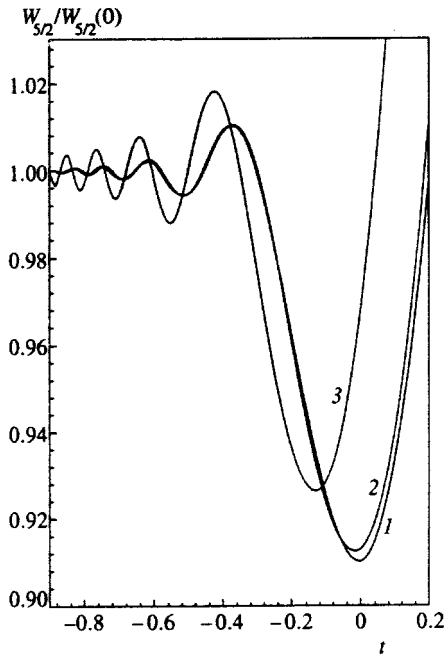


FIG. 5. The normalized probability of pair production, $W_{5/2}$, as a function of the parameter $t=(2/3)\log(2\chi)$ for different values of γ; $\gamma^2=0$ (*1*), 0.1 (*2*), 1 (*3*).

A comparison of Eqs. (58) and (59) reveals clearly the decreasing oscillatory field dependence ($\chi \ll 1$) of the probability for $\nu=5/2$ compared to $\nu=3/2$. Thus, in particular, in the limit $\delta \to 1$ (electromagnetic wave field) we can see that if the main term characterizing the oscillatory behavior for $\nu=3/2$ is proportional to

$$-\frac{1}{2}\,\chi\,\cos\frac{2}{3\chi}, \tag{60}$$

then for $\nu=5/2$ we have

$$-\frac{3}{4}\,\chi^2\,\sin\frac{2}{3\chi}. \tag{61}$$

Including the dependence on the characteristics of the total field has shown that when $\delta^2 \sim \xi^2 \gg 1$ the main oscillatory dependence for $\nu=5/2$ is concentrated in the term proportional to

$$-\frac{3}{4}\,\chi\,\frac{A_2}{\xi^2}\,\cos\frac{2}{3\chi}. \tag{62}$$

Figures 4, 5, and 6 show plots of the probabilities $W_\nu/W_\nu(0)$ for $\nu=3/2$, 5/2, and 9/2 as functions of the plane wave field parameter χ for different values of $(\delta/\xi)^2$. It should be noted that for $|\delta| \gg 1$ and $\xi \gg 1$ a new parameter,

$$\gamma^2=\frac{\delta^2}{\xi^2}, \tag{63}$$

arises which characterizes the complete probabilities of all these processes. It is easy to see that for $|\delta| \gg 1$ the new parameter represents the ratio of twice the energy of the frequency shift of the charged oscillator in the magnetic field to the work of the field at the electron Compton wavelength,

$$\gamma^2 \simeq \frac{\omega_H^2}{\omega^2}\,\frac{\omega^2 m^2}{e^2E^2}=\frac{H^2}{E^2}. \tag{64}$$

Thus, in this limit γ is independent of the frequency characteristics of the field, but is determined by the ratio of the strength of the constant magnetic field to the field amplitude of the electromagnetic wave. From the standpoint of possible

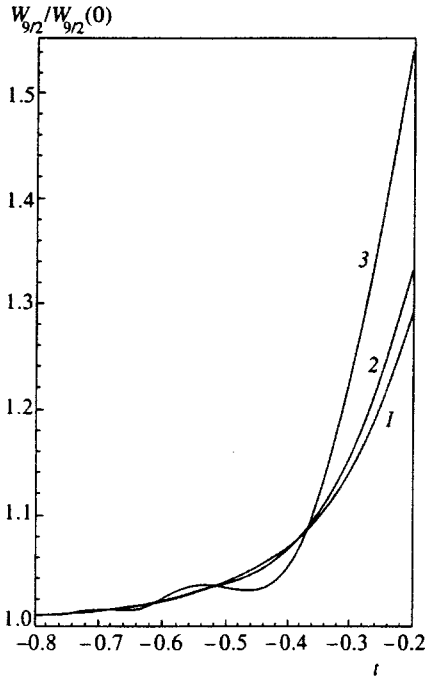


FIG. 6. The normalized probability of β-decay, $W_{9/2}$, as a function of the parameter $t=(2/3)\log(2\chi)$ for different values of γ; $\gamma^2=0$ (*1*), 10 (*2*), $10^{1.5}$ (*3*).

experiments, there may be special interest in the region $E\sim H$, where the oscillatory contributions increase and may greatly exceed the perturbation theory corrections. A comparison of the probabilities for the processes with $\nu=3/2$, 5/2, and 9/2 also shows that those with lower values of ν are more subject to the influence of the field. This is also indicated by the graph in Fig. 4, which corresponds to e^+e^- pair production by two photons with identical polarizations.

There is also interest in making an asymptotic expansion of the probabilities for the different values of ν and $\chi\ll 1$ using Eqs. (40), (47), and (48). With the aid of the expansions (50)–(52) it is easy to obtain a representation of W_ν in the form of a series in powers of χ with coefficients that depend on the parameter μ/I. Limiting ourselves to the first terms of the expansion, we have

$$\frac{W_{3/2}}{W_{3/2}(0)}=1-\frac{\mu^2}{12I^2}+\frac{\mu^4}{48I^4}+\left(\frac{1}{8}+\frac{35\mu^2}{96I^2}-\frac{77\mu^4}{128I^4}\right)\chi^2-\left(\frac{105}{128}+\frac{5005\mu^2}{512I^2}\right)\chi^4+\sin\frac{2}{3\chi}\left[\frac{\mu^4}{216I^4}\frac{1}{\chi^2}-\frac{\mu^2}{72I^2}-\frac{115\mu^4}{186624I^4}+\left(-\frac{17}{24}+\frac{3115\mu^2}{62208I^2}\right)\chi^2+\frac{199115}{20736}\chi^4\right]+\cos\frac{2}{3\chi}\left[\frac{\mu^4}{90I^4}\frac{1}{\chi^3}+\left(\frac{\mu^2}{6I^2}-\frac{5\mu^4}{5184I^2}\right)\frac{1}{\chi}+\left(-\frac{1}{2}+\frac{35\mu^2}{1728I^2}\right)\chi+\frac{1225}{576}\chi^3\right], \tag{65}$$

$$\frac{W_{5/2}}{W_{5/2}(0)}=1-\frac{\mu^2}{4I^2}+\frac{3\mu^4}{80I^4}+\left(-\frac{1}{8}+\frac{5\mu^2}{32I^2}-\frac{63\mu^4}{128I^4}\right)\chi^2+\left(\frac{35}{128}-\frac{1115\mu^2}{512I^2}+\frac{45045\mu^4}{2048I^4}\right)\chi^4+\sin\frac{2}{3\chi}\left[-\frac{\mu^4}{20I^4}\frac{1}{\chi^2}-\frac{\mu^2}{4I^2}+\frac{\mu^4}{5760I^4}+\left(-\frac{3}{4}+\frac{205\mu^2}{1152I^2}\right)\chi^2+\frac{3745}{384}\chi^4\right]+\cos\frac{2}{3\chi}\times\left[-\frac{\mu^4}{240I^4}\frac{1}{\chi}+\left(\frac{5\mu^2}{48I^2}-\frac{133\mu^4}{41472I^4}\right)\chi+\left(\frac{35}{16}-\frac{22715\mu^2}{41472I^2}\right)\chi^3-\frac{805805}{13824}\chi^5\right], \tag{66}$$

and

$$\frac{W_{9/2}}{W_{9/2}(0)}=1+\frac{35\mu^2}{48I^2}-\frac{7\mu^4}{768I^4}+\left(\frac{35}{8}-\frac{35\mu^2}{384I^2}+\frac{35\mu^4}{6144I^4}\right)\chi^2+\left(\frac{35}{128}+\frac{1225\mu^2}{6144I^2}-\frac{2695\mu^4}{32768I^4}\right)\chi^4+\sin\frac{2}{3\chi}\left[-\frac{7\mu^4}{768I^4}+\left(-\frac{35\mu^2}{64I^2}+\frac{1435\mu^4}{221184I^4}\right)\chi^2+\left(\frac{105}{16}+\frac{131075\mu^2}{18432I^2}\right)\chi^4-\frac{603575}{1536}\chi^6\right]+\cos\frac{2}{3\chi}\left[\frac{35\mu^4}{9216I^4}\chi+\left(\frac{1225\mu^2}{768I^2}-\frac{159005\mu^4}{7962624I^4}\right)\chi^3-\frac{3115}{64}\chi^5\right]. \tag{67}$$

With the aid of these formulas we can see that the expansions in terms of the parameters χ and μ/I are in complete agreement with the corresponding terms in the expansions in a constant plane wave field, as well as in a constant magnetic field, taken separately. Furthermore, in Eqs. (65)–(67) there are combination terms which in the monotonic part of the expansion have a regular character with respect to the parameter χ for all three processes considered here. The oscillatory contributions to the pair production processes, on the other hand, contain terms which are proportional to inverse powers of the plane wave parameter $\chi\ll 1$:

$$\nu=\frac{3}{2}:\quad \frac{\mu^4}{I^4}\left(\frac{1}{90\chi^3}\cos\frac{2}{3\chi}+\frac{1}{216\chi^2}\sin\frac{2}{3\chi}\right), \tag{68}$$

and

$$\nu=\frac{5}{2}:\quad \frac{\mu^4}{I^4}\left(-\frac{1}{20\chi^2}\sin\frac{2}{3\chi}-\frac{1}{240\chi}\cos\frac{2}{3\chi}\right). \tag{69}$$

Thus, we conclude that in a combination field when the allowable values of the parameters $\mu\ll I$ and $\chi\ll 1$ are taken into account, situations may arise in which the oscillatory contributions will predominate over the perturbation theory corrections, but the character of the enhancement in the contributions (68) and (69) is different from that of Eqs. (58) and (59).

5. FORBIDDEN β-DECAY IN A PLANE WAVE ELECTROMAGNETIC FIELD

Forbidden β-decay in an intense electromagnetic wave field has been examined previously.[15] Two mechanisms for the reduction in the decay half life of nuclei in an external field were evaluated, one owing to an increase in the phase volume of the states of the product electron and the other owing to absorption by the nucleus of photons from the wave and a change in the selection rules for a forbidden transition into an allowed one. It was shown[15] that, despite a number of previous estimates[18] in which it was predicted, for example, that the probability of the forbidden transition $^{113}\mathrm{Cd}\rightarrow{}^{113}\mathrm{In}$ should increase by 12 orders of magnitude, at the currently attainable electromagnetic field intensities the removal of the prohibition by the external field would not lead to any noticeable reduction in the half life of the nuclei. Nevertheless, it was established that for forbidden β-decay the effect of an external field on the product electron is more significant than in the case of allowed β-transitions.

In this regard, it is interesting to estimate the probability of forbidden β-decay in a plane wave electromagnetic field

for the case in which the frequency dependence is unimportant. It is easy to see that in this limit ($\omega \to 0$) the main contribution to the probability will originate in the influence of the external field on the electron. If we limit ourselves to the case of the unique β-transitions of the first forbiddenness class and assume that the interaction of the parent and daughter nuclei with the external field can be neglected entirely, after a number of transformations the probability reduces to a form that was studied in detail in Section 3:

$$W_{11/2}=G_{11/2}I^{9/2}\chi^3\left[I_{(-11/2)}-\frac{2i}{3}I_{(-5/2)}\right], \tag{70}$$

where

$$I_{(-11/2)}=-\frac{2i}{9}\left[I_{(-5/2)}-\chi^{-2/3}I_{(-9/2)}\right].$$

Substituting the expressions $I_{(-5/2)}$ and $I_{(-9/2)}$ in Eq. (70), we finally obtain the following for the forbidden β-decay probability:

$$W_{11/2}=\frac{W_{11/2}(0)}{\pi}\left[\sqrt{y}\Phi^2\left(1+\frac{161}{4}\chi^2\right)+\frac{\Phi'^2}{\sqrt{y}}\times\left(1+\frac{155}{4}\chi^2\right)-\chi\Phi\Phi'(1+35\chi^2)\right]. \tag{71}$$

Using Eqs. (50)–(52), from Eq. (71) for $\chi \ll 1$ we have

$$\frac{W_{11/2}}{W_{11/2}(0)}=1+\frac{315}{8}\chi^2-\frac{525}{128}\chi^4+\frac{8575}{1024}\chi^6+\sin\frac{2}{3\chi}\times\left(-\frac{315}{16}\chi^4-\frac{26425}{512}\chi^6\right)+\cos\frac{2}{3\chi}\left(\frac{1785}{64}\times\chi^5+\frac{29904875}{18432}\chi^7\right). \tag{72}$$

The correction of order χ^2 is consistent with the result obtained in Ref. 15. The oscillatory contributions indicate a similar field dependence for the probabilities of forbidden and allowed β-transitions. In both cases the oscillations show up only in terms $\sim\chi^4$. A comparison of the expansions of W_ν for different values of $\nu=3/2$, 5/2, 9/2, and 11/2 demonstrates clearly an enhancement in the numerical coefficient of χ^2 with increasing ν. Here, however, the contribution of the oscillatory corrections decreases. Figures 7 and 8 show plots of the probabilities of pair production by photons with different polarizations, as well as of the probabilities of allowed and forbidden β-processes, as functions of the crossed field parameter χ which also confirm our conclusions regarding the monotonic and oscillatory behavior of the probabilities for these processes.

6. CONCLUSION

This study of different processes in external electromagnetic fields with rather complicated configurations including a constant magnetic field and a circularly polarized electromagnetic wave provides a basis for a number of conclusions regarding the behavior of the probabilities of these processes over wide ranges of variation in the field parameters, almost without restrictions on the field strengths involved in the superposition.

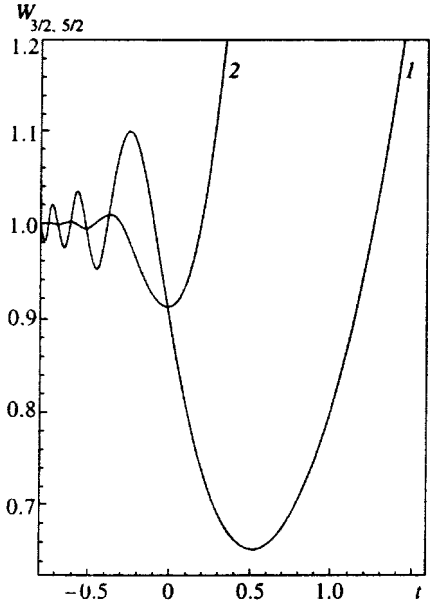


FIG. 7. The normalized probabilities $W_{3/2}$ (curve *1*) and $W_{5/2}$ (curve *2*) as functions of the plane wave field parameter $t=(2/3)\log(2\chi)$.

1. The probabilities of these processes, which are different in nature, nevertheless have a similar field dependence in the time integral representations for the nonrelativistic limit. The distinguishing features in this case essentially reduce to different power law dependences for the functions of time, $t^{-\nu}$, in the integrands, where in the absence of a field we have $\nu=3/2$ for pair production by two photons with identical polarizations, $\nu=5/2$ for pair production by two γ-photons with different polarizations, $\nu=9/2$ for allowed β-

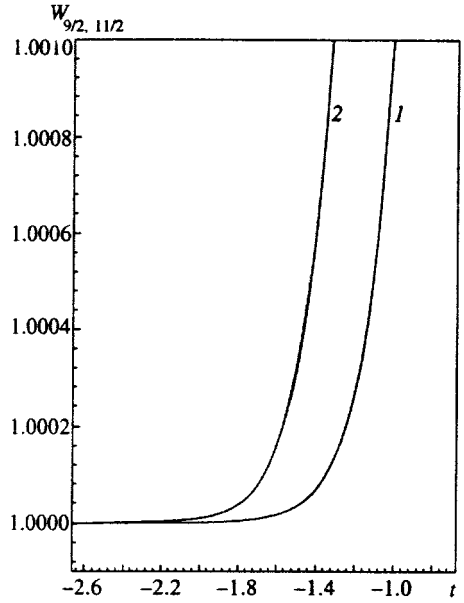


FIG. 8. The normalized probabilities of β-decay for allowed, $W_{9/2}$ (curve *1*), and forbidden, $W_{11/2}$ (curve *2*), β-transitions as functions of the plane wave field parameter $t=(2/3)\log(2\chi)$.

decay, and $\nu=11/2$ for the unique β-transitions in the first forbiddenness class. These analogies make it possible to study all these processes in an external electromagnetic field with a complicated configuration via a single approach employing integral Mellin transforms.[5,16]

2. In the case of sufficiently weak magnetic fields $\mu \ll I$ the processes are mainly shaped by the plane wave field and for frequencies $\omega \ll Im$ the leading term is a field dependence in a form characteristic for a constant crossed field. It should be noted that this case has been analyzed in detail before[13] for processes with $\nu=3/2$ and $\nu=9/2$. However, in an electromagnetic field consisting of the superposition of a constant magnetic field and a wave field, when the condition $\omega_H \gg \omega$ is satisfied, even when each of the parameters in the inequality is not large as before, the character of the field dependence can change significantly. In particular, for the process $\nu=5/2$ it can be seen that if interference processes are of order χ^2 in a constant crossed field, then the combination field leads to a reduction in the order of the correction in χ by unity, i.e., it already shows up in terms $\sim\chi$. A similar reduction in the order of appearance of the oscillatory corrections is also observed in the case of the process $\nu=9/2$. In this case the reduction in the order of the dependence on χ is from χ^4 to χ^3. It is known that for $\nu=3/2$ in a plane wave field $\chi \ll 1$, the oscillatory corrections (interference effects) exceed the perturbation theory corrections and already show up in terms of order χ. In the case of superposed fields subject to the condition $\omega_H/\omega \geqslant \xi \gg 1$, the analogous contributions no longer depend on $\chi \ll 1$. Under these conditions a new parameter arises which in this approximation is independent of the frequency characteristics of the field but is determined by the ratio of the strengths of the fields constituting the configuration. Note that in the case of the unique transitions of the first forbiddenness class, interference effects in show up in the total probability for the process in a constant plane wave field, as in the case of allowed β-decay, in terms $\sim\chi^4$. We might expect that for transitions in the higher forbiddenness classes this trend in the behavior of the corrections to the total probability for the process owing to the effect of the field on the product electron would be retained. In this case the degree of singularity at zero in the time integral representations ($\nu=13/2, 15/2\ldots$), would increase, which leads to an increase in the monotonic contributions and a relative reduction in the role of interference effects. Then, as noted previously,[15] the corrections to the transition probabilities for the higher forbiddenness classes owing to removal of the prohibition in an external field can only become smaller.

3. Representing the total probabilities for these processes in the form of expansions in terms of the magnetic field parameter μ/I shows that in this case for $\nu=3/2, 5/2$ there are root singularities at the points $I=2\mu n$, where $n=1,2\ldots$, which have been pointed out before.[20,21] For these values of the parameter the product electron and positron have a zero component of the momentum along the magnetic field. It is easy to see that the divergence in the total probability for certain values of the parameter is a consequence of neglecting the interaction of the e^+ and e^- and shows up when their velocities are smallest. Studies of the behavior of the probabilities for pair production near these points require special consideration, which with a more rigorous statement of the problem should, of course, eliminate these singularities. With increasing magnetic field for $\mu=I/2$ the oscillatory behavior in the probabilities and cross sections for pair production ceases and for $\mu>I/2$ the cross section rises linearly for $\nu=3/2$, while when the photons are oppositely polarized ($\nu=5/2$) their intersection cross section goes identically to zero.

4. The asymptotic expansions of the probabilities for the processes with $\nu=3/2$, 5/2, and 9/2 for $\mu \ll I$ and $\chi \ll 1$ have been written in the form of series in powers of χ with coefficients depending on the parameter μ/I. In this case, along with the combination terms, which are regular for all three processes in the monotonic part of the expansion, the interference contributions in the pair production processes contain terms proportional to inverse powers of $\chi \ll 1$: $\sim\chi^{-3}$ for $\nu=3/2$ and $\sim\chi^{-2}$ for $\nu=5/2$. This sort of behavior also indicates that in a combination field for certain values of the parameters the role of interference effects may be enhanced compared to the case of a purely plane wave field.

5. The numerical calculations of the probabilities of the processes with $\nu=3/2$, 5/2, 9/2, and 11/2 shown in the figures clearly illustrate the reduced contribution of interference effects with increasing ν, which characterizes one or another process and simultaneously leads to an enhancement of the monotonic dependences.

In conclusion, the author thanks O. F. Dorofeev and V. R. Khalilov for their interest in this work and for valuable comments. The author also thanks the Competitive Center for Fundamental Natural Science of the State Committee of the Russian Federation for Higher Education at St. Petersburg State University for support (Grant No. 95-0-5.3-58).

[1] V. B. Berestetskiĭ, E. M. Lifshits, and L. P. Pitaevskiĭ, *Quantum Electrodynamics*, Pergamon, Oxford (1982).
[2] I. M. Ternov, V. N. Rodionov, and O. F. Dorofeev, ÉChAYa **20**, 51 (1989).
[3] V. I. Ritus, Trudy FIAN **111**, 84 (1979).
[4] A. I. Nikishov and V. I. Ritus, Trudy FIAN **168**, 232 (1983).
[5] I. M. Ternov, V. R. Khalilov, and V. N. Rodionov, *The Interaction of Charged Particles with Strong Electromagnetic Fields* [in Russian], Izdatel'stvo Moskovskogo un-ta., Moscow (1982).
[6] V. I. Ritus, Zh. Éksp. Teor. Fiz. **56**, 986 (1969) [Sov. Phys. JETP **29**, 532 (1969)].
[7] V. N. Rodionov, in *Electrons in Strong Electromagnetic Fields: An Advanced Classical and Quantum Treatment*, V. R. Khalilov, ed., Gordon and Breach, Amsterdam (1996), Ch. 9, p. 275.
[8] I. M. Ternov, V. N. Rodionov, and O. F. Dorofeev, Zh. Éksp. Teor. Fiz. **84**, 1225 (1983) [Sov. Phys. JETP **57**, 710 (1983)].
[9] I. M. Ternov, V. N. Rodionov, O. F. Dorofeev, V. G. Zhulego, A. E. Lobanov, and V. C. Peres-Fernandez, Vestnik MGU. Fizika. Astronomiya **24**(4), 79 (1983).
[10] V. N. Rodionov, S. G. Starcheus, M. A. Tasev, and I. M. Ternov, Zh. Éksp. Teor. Fiz. **94**(1), 56 (1988) [Sov. Phys. JETP **67**, 30 (1988)].
[11] M. B. Voloshchin, Yad. Fiz. **38**, 814 (1983).
[12] A. I. Nikishov and V. I. Ritus, Zh. Éksp. Teor. Fiz. **85**, 24 (1983) [Sov. Phys. JETP **58**, 14 (1983)].
[13] A. I. Nikishov and V. I. Ritus, Zh. Éksp. Teor. Fiz. **85**, 1544 (1983) [Sov. Phys. JETP **58**, 898 (1983)].
[14] E. Kh. Akhmedov, Zh. Éksp. Teor. Fiz. **85**, 1521 (1983) [Sov. Phys. JETP **58**, 883 (1983)].
[15] E. Kh. Akhmedov, Zh. Éksp. Teor. Fiz. **87**, 1541 (1984) [Sov. Phys. JETP **60**, 884 (1984).

[16] V. N. Rodionov, Zh. Éksp. Teor. Fiz. **111**, 3 (1997) [JETP **84**, 1 (1997)].
[17] W. Becker, W. H. Louisel, J. D. McCullen, and M. O. Scully, Phys. Rev. Lett. **47**, 1262 (1981).
[18] H. Reiss, Phys. Rev. C **27**, 1199 (1983).
[19] A. E. Lobanov, V. N. Rodionov, and V. R. Khalilov, Yad. Fiz. **32**, 174 (1980) [Sov. J. Nucl. Phys. **32**, 90 (1980)].
[20] A. E. Lobanov and A. R. Muratov, Zh. Éksp. Teor. Fiz. **87**, 1140 (1984) [Sov. Phys. JETP **60**, 651 (1984)].
[21] B. A. Lysov, O. S. Pavlova, and A. F. Zhuravlev, Vestnik MGU. Fizika. Astronomiya **12**(5), 557 (1971).
[22] Yee Jack Ng and Wu-yang Tsai, Phys. Rev. D **16**, 286 (1977).
[23] O. F. Dorofeev and V. N. Rodionov, Vestnik MGU. Fizika. Astronomiya **23**(3), 90 (1982).
[24] A. Erdélyi, *Higher Transcendental Functions*, Vol. 1, McGraw-Hill, N.Y. (1953).

Translated by D. H. McNeill

Electric currents of excitations in the one-dimensional attractive Hubbard model

S. I. Matveenko

Landau Institute for Theoretical Physics, Russian Academy of Sciences, 117940 Moscow, Russia
(Submitted 20 May 1997)
Zh. Éksp. Teor. Fiz. **113**, 204–212 (January 1998)

We have calculated electric currents of various types of excitations in the Hubbard model. Both spin and charge excitations carry the electric current. The electric charge is a continuous function of the band filling and the single-site repulsion potential. © *1998 American Institute of Physics.* [S1063-7761(98)01401-2]

1. INTRODUCTION

Electron–electron interactions have a great effect on properties of one-dimensional systems. This results in critical behavior at $T=0$ with algebraic decay of correlation functions. The particle momentum distribution differs from a Fermi-liquid step function. As distinct from a Fermi liquid, such systems are called Luttinger liquids. In the framework of the one-dimensional ($1D$) Hubbard model with attractive interaction we investigate the electrical conductivity properties of the excitations. Interest in this model arises because the Hubbard model with attraction is the simplest one with dominant superconductivity fluctuations.

In contrast to Fermi systems, where the electric charge of the quasiparticles can have integer values e or $2e$, we will show that the value of the electron charge of one-particle states can depend continuously on the system parameters. The charge of the one-particle excitations is defined by $q=j/v$, where $j=-\partial\epsilon/\partial A$ is the electric current, defined as the derivative of the energy with respect to the magnetic vector potential, and $v=\partial\epsilon/\partial p$ is the velocity of the excitation. We will find that the spin and particle–hole excitations near the Fermi surface carry the electric current. This effect is absent for models with an exactly linear electron spectrum, so that in the weak coupling regime the current is proportional to the Fermi-velocity dispersion.

Some time ago unusual electrical properties of excitations were found in $1d$ electron–phonon Peierls systems. As a result of the electron–phonon interactions a transition from a metallic to an insulating state with creation of a charge density wave takes place. It was found that excitations in the Peierls system (solitons, polarons)[1,2] may have fractional charges depending continuously on the band filling and the electron–phonon coupling constant. But it is obvious that after integrating the Peierls model over the phonon degrees of freedom we will have an electron model with some effective electron–electron interaction (attraction). For example, we get in the quantum limit (the ion mass tends to zero) an effective ''g-ology'' Hamiltonian with a backscattering term due to the electron–phonon interaction. Therefore excitations in the two systems may have common properties. We have found indeed that both model excitations carry noninteger electric charges.

The plan of the paper is as follows. In Sec. 2 we quote known results that are needed. In Sec. 3 we calculate the currents and charges of excitations for the model with $U<0$ and generalize some results to the $U>0$ case.

2. PROPERTIES OF THE MODEL

2.1 Attractive interaction

The Hamiltonian for the Hubbard ring in a magnetic flux Φ is

$$H=-\sum_{j,\sigma}\{c^{\dagger}_{j,\sigma}c_{j+1,\sigma}\exp(i\nu)+\text{H.c.}\}+4U\sum_{j}n_{j,\uparrow}n_{j,\downarrow}-\frac{h}{2}\sum_{j}(n_{j,\uparrow}-n_{j,\downarrow})-\mu\sum_{j,\sigma}n_{j,\sigma}, \tag{1}$$

where N_a is the number of sites, $c^{\dagger}_{n,\sigma}, c_{n,\sigma}$ are the creation and annihilation operators for electrons with spins $\sigma=\uparrow,\downarrow$, $U<0$ is the onsite attraction amplitude of particles with opposite spins, h is the spin magnetic field, μ is the chemical potential,

$$\nu=2\pi\Phi/N_a\Phi_0, \quad \Phi=AN_a,$$

$\Phi_0=hc/e$ is the magnetic unit flux, and A is the vector potential of the orbital magnetic field.

The electric current by definition is

$$j=-\left.\frac{\partial H}{\partial A}\right|_{A=0}. \tag{2}$$

The ground state and excitations are described by sets of quasimomenta k_j, and rapidities λ_α, which are solutions of the Bethe ansatz (BA) equations[3]

$$N_a k_j-N_a\nu-\sum_{\beta=1}^{M}\theta(2\sin k_j-2\lambda_\beta)=2\pi I_j, \quad j=1,\ldots,N, \tag{3}$$

$$N_a P_0(\lambda_\alpha)-2N_a\nu-\sum_{j=1}^{X}\theta(2\lambda_\alpha-2\sin k_j)-\sum_{\beta=1}^{M}\theta(\lambda_\alpha-\lambda_\beta)=2\pi J_\alpha, \quad \alpha=1,\ldots,M, \tag{4}$$

where

$$\theta(x)=2\arctan(x/2u); \quad u=|U|.$$

$N=2M+X$ is the number of particles, M is the number of pairs, X is the number of unpaired particles, and $P_0(\lambda_\alpha)=2\,\mathrm{Re}\arcsin(\lambda_\alpha-iu)$ is the bare momentum of a pair. We treat states with a number of singlet bound pairs and a num-

 1063-7761/98/010115-05$15.00

ber of electrons with uncompensated up-spins. In contrast to the case $U>0$, where all wavenumbers are real in the ground state, all particles are paired and the wavenumbers are complex. If the external magnetic field exceeds some critical value h_c, unpaired electrons are formed. Equations (3) with real numbers k_j describe electrons with uncompensated up-spins. Singlet bound pairs are characterized by a pair of complex wavenumbers $k_\alpha^{\pm}$ and a rapidity λ_α connected through the relation

$$\sin k_\alpha^{\pm}=\lambda_\alpha \pm iu, \quad \alpha=1,...,M.$$

Equations (4) are obtained from the equations of Lieb and Wu[4] by eliminating the complex wavenumbers $k_\alpha^{\pm}$.

In the ground state the numbers I_j, J_α are distributed symmetrically about zero. They satisfy

$$I_j=\frac{1}{2}(1-N+2M)+j-1, \quad J_\alpha=\frac{1}{2}(1-M)+\alpha-1.$$

In the case of moderate fields $h<h_c$, all spins are paired $(X=0)$.

The system energy is

$$W=\sum_j [e_0(k_j)-\mu]+\sum_\alpha [E_0(\lambda_\alpha)-2\mu], \tag{5}$$

where

$$E_0(\lambda_\alpha)=-4\,\mathrm{Re}\sqrt{1-(\lambda_\alpha-iu)^2},$$

$$e_0(k_j)=-2\cos k_j-\frac{h}{2}$$

are the bare energies of pairs and unpaired electrons.

The momentum is

$$P=\sum_j (k_j-\nu)+\sum_\alpha [P_0(\lambda_\alpha)-2\nu].$$

The density functions of the k and λ distributions are usually introduced as

$$\rho(k)=\frac{1}{2\pi}\frac{dp(k)}{dk}, \quad \sigma(\lambda)=\frac{1}{2\pi}\frac{dp(\lambda)}{d\lambda},$$

where

$$p(k)=k-\frac{1}{N_a}\sum_{\beta=1}^{M}\theta(2\sin k-2\lambda_\beta),$$

$$p(\lambda)=P_0(\lambda_\alpha)-\frac{1}{N_a}\sum_{j=1}^{X}\theta(2\lambda_\alpha-2\sin k_j)$$

$$-\frac{1}{N_a}\sum_{\beta=1}^{M}\theta(\lambda_\alpha-\lambda_\beta).$$

In the thermodynamic limit Eqs. (3) and (4) can be written in the matrix form[3]

$$\rho(k,\lambda)=\rho_0(k,\lambda)+K\otimes\rho(k',\lambda'), \tag{6}$$

where

$$\rho(k,\lambda)=(\rho(k),\sigma(\lambda))^T,$$

$$\rho_0(k,\lambda)=\left(\frac{1}{2\pi},\frac{1}{2\pi}\frac{dP_0(\lambda)}{d\lambda}\right)^T,$$

$$K(k,\lambda|k',\lambda')=\begin{pmatrix}0 & -2\cos kK(2(\sin k-\lambda')) \\ -2K(2(\lambda-\sin k')) & -K(\lambda-\lambda')\end{pmatrix},$$

$$K(x)=\frac{1}{2\pi}\frac{d\theta(x)}{dx}=\frac{1}{2\pi}\frac{4u}{4u^2+x^2}.$$

The product $\otimes$ indicates the usual matrix product and integration over the common variables, from $-Q$ to Q over k and from $-\Lambda$ to Λ over λ, respectively; the superscript ''T'' means matrix transposition.

Equation (5) acquires the form

$$W=\epsilon_0^T\otimes\rho=\rho_0^T\otimes\epsilon,$$

where

$$\epsilon=\epsilon_0+K^T\otimes\epsilon, \tag{7}$$

$\epsilon=(e(k),\epsilon(\lambda))^T$ are the excitation energies of unpaired electrons and of the pairs, $\epsilon_0=(e_0(k)-\mu,E_0(\lambda)-2\mu)^T$.

It is known that for magnetic fields less than the critical value $h<h_c$ the spectrum of paired excitations is gapless $(\epsilon(\pm\Lambda)=0)$, while unpaired electron states have a gap $e(k=0)\neq 0$.

The matrix of dressed charges is defined by[3,5]

$$\xi(k,\lambda)=\begin{pmatrix}\xi_1^1(k) & \xi_1^2(k) \\ \xi_2^1(\lambda) & \xi_2^2(\lambda)\end{pmatrix}, \tag{8}$$

where

$$\xi(k,\lambda)=\begin{pmatrix}1 & 0 \\ 0 & 1\end{pmatrix}+K^T\otimes\xi. \tag{9}$$

For the case $h<h_c$ only the function $\xi_2^2(\lambda)=Z(\lambda)$ is relevant, and it satisfies the equation[6]

$$Z+\int_{-\Lambda}^{+\Lambda}K(\lambda-\lambda')Z(\lambda')d\lambda'=1,$$

$$\frac{1}{2}\leqslant Z(\lambda)\leqslant 1, \quad \frac{\partial Z}{\partial\Lambda}<0, \quad \frac{\partial Z(\lambda)}{\partial u}>0. \tag{10}$$

The solution of this equation is known in some limits:

$$\rho\to 1, \quad 2Z^2(\Lambda)=1-\frac{1}{2\ln[C/(1-\rho)]},$$

$$C=\sqrt{\frac{8}{\pi e}}\,I_0\left(\frac{\pi}{2u}\right), \quad \rho=1, \quad 2Z^2(\Lambda)=1,$$

$$\rho\to 0, \quad \frac{\rho}{u}\to 0, \quad 2Z^2(\Lambda)=\frac{1}{2}\left(1+\frac{\rho}{2}\sqrt{1+\frac{1}{u^2}}\right).$$

For magnetic fields $h>h_{c1}$ the gap in the spectrum of unpaired excitations closes $(e(0)=0)$ and for $h=h_{c2}$ the system undergoes a transition into the saturated ferromagnetic ground state. In the region $h_{c2}>h>h_{c1}$ the dressed charge matrix has been found in Ref. 7 as

$$\xi(Q,\Lambda)=\begin{pmatrix} 1+\dfrac{\kappa k_0}{2} & 0 \\ -\dfrac{1}{2}-\kappa k_0 & \left(1+\dfrac{u}{2\pi\lambda_0}\right)\dfrac{1}{\sqrt{2}} \end{pmatrix},$$

where

$$\kappa=\frac{\ln 2}{2u},\quad k_0=\sqrt{\frac{h-h_c}{\eta}},$$

$$\eta=1-2\int_0^\infty \frac{tdt\ J_1(t)}{1+\exp(2ut)}>0,\quad \lambda_0=\frac{2u}{\pi}\ln\frac{C}{1-\rho}.$$

2.2. Repulsive interaction

Similar equations are valid for the model with $U>0$. The matrix K differs from (6) only in the signs of the nondiagonal terms. A density function $\rho(k)$ describes the distribution of particle quasimomenta and the function $\sigma(\lambda)$ describes the rapidities of spin-down particles. The energy and the momentum are

$$W=-\sum\ 2\cos k_j,\quad p=\sum\ (k_j-\nu).$$

The dressed charge matrix is[5]

$$\xi=\begin{pmatrix} \xi(k) & \xi(k)/2 \\ 0 & \sqrt{2}/2 \end{pmatrix}, \tag{11}$$

where $\xi(k)$ satisfies the equation

$$\xi(k)=1+\int_{-Q}^{Q}dk'\ \cos k'\widetilde{K}(\sin k-\sin k')\xi(k'), \tag{12}$$

$$\widetilde{K}=\frac{1}{2\pi}\int_0^\infty \frac{e^{-\omega u}}{\cosh(\omega u)}\cos(\omega\lambda)d\lambda.$$

The solution $1\leqslant\xi(Q)\leqslant\sqrt{2}$ is known in some limiting cases:[5]

$\sin Q/u\ll 1,\quad \xi(Q)=1+\sin Q\ \ln 2/(\pi u)$;

$\rho=1,\quad Q=\pi,\quad \xi(Q)=1,\quad 0<u<\infty$;

$\rho\to 1,\quad \xi(Q)=1+\ln 2(1-\rho)f(u)/u,$

$f(u)=1-\int_0^\infty dx\ J_0(x)\exp(-ux)/\cosh xu$;

$\rho\ll 1,u,\quad \xi(Q)=1+\rho\ \ln 2/u$;

$u\gg 1,\quad \xi(Q)=1+\sin\ \pi\rho\ \ln 2/(\pi u)$;

$u\ll\sin Q,\quad \xi(Q)=\sqrt{2}[1-u/(2\pi\ \sin Q)]$.

3. ELECTRIC CURRENTS

3.1. Attractive interaction

3.1.1. Gapless paired excitations ($h<h_{c1}$). In the ground state all particles are bound into singlet pairs. The lowest excitations are gapless excitations of bound pairs. For a sufficiently strong magnetic field ($h>h_c$) unpaired electron excitations are gapless too.

The particle–hole excitations of pairs are described by a set of numbers $J_\alpha=J_\alpha^0+\Theta(\alpha-\alpha_0)$ (for hole excitations), where $\{J_\alpha^0\}$ and $\{\lambda_\alpha^0\}$ are the ground-state sets. Here $\Theta(x)$ is the usual step-function.

From (3) we find an equation for the function $\overline{\sigma}(\lambda_\alpha)=N_a\sigma(\lambda_\alpha)\delta\lambda_\alpha$:

$$\widetilde{\sigma}(\lambda)+\int_{-\Lambda}^{\Lambda}K(\lambda-\lambda')\widetilde{\sigma}(\lambda')d\lambda'=\Theta(\lambda-\lambda_0)+\frac{2\Phi}{\Phi_0}, \tag{13}$$

where $\delta\lambda=\lambda_\alpha-\lambda_\alpha^0$ is the shift of a number λ_α due to the excitation and $\sigma(\lambda_0)$ is the ground-state solution.

Taking the derivative with respect to ν we have $\partial\widetilde{\sigma}/\partial\nu=N_a\xi_2(\lambda)/\pi$. The energy is

$$W=\sum\ E_0(\lambda_\alpha^0)+\sum\ E_0'(\lambda_\alpha)\delta\lambda_\alpha+\frac{1}{2}\sum\ E_0''(\lambda_\alpha^0)(\delta\lambda)^2$$

$$=\int E_0'\widetilde{\sigma}(\lambda)d\lambda+\frac{1}{2}\int E_0''\frac{\widetilde{\sigma}^2}{N_a\sigma(\lambda)}\,d\lambda. \tag{14}$$

Then the current $j=-\partial W/\partial A$ equals

$$j=\frac{2}{\Phi_0}\int\frac{E_0''\widetilde{\sigma}(\lambda)\xi_2(\lambda)}{\sigma(\lambda)}\,d\lambda. \tag{15}$$

The momentum is $p=\int P_0'(\lambda)\widetilde{\sigma}(\lambda)d\lambda$, where $\widetilde{\sigma}$ is the solution of Eq. (13) for $\Phi=0$.

For excitations with small momenta $p\propto(\Lambda-\lambda_0)$ we can easily see that the current $j\propto pE_0''(\Lambda)$ is proportional to the dispersion of the Fermi velocity. In the limit $u\to 0$ we find

$$\widetilde{\sigma}=\Theta(\lambda-\lambda_0)/2,\quad \sigma(\lambda)=1/2,\quad Z=1/2,$$

$$j=(\pi/\Phi_0)E_0(p_\Lambda)p.$$

As we showed in the case of repulsion,[8–10] the current would be absent in the linear-spectrum approximation.

3.1.2. States with added particles. We now calculate the currents of states obtained by adding an electron pair or unpaired electron. The simplest way to calculate the current is to include the magnetic vector-potential in the $1/N_a$ corrections of the energy, found in[11]:

$$\delta W=\frac{2\pi}{N_a}\sum_{n=1,2}v_n\{(Z^{-1}\Delta\mathbf{N})_n^2+(Z^T\mathbf{D})_n^2+I_n^+ +I_n^-\}. \tag{16}$$

where $v_1=\partial\epsilon(Q)/\partial p(Q)$ and $v_2=\partial\epsilon(\Lambda)/\partial p(\Lambda)$ are the Fermi velocities of unpaired and pair excitations, respectively, and $Z_{ij}=\xi_j^i(Q,\Lambda)$. Note that for $h<h_c$ unpaired excitations have a gap, so for this case $v_1=0$. In the case $h>h_c$ both singlet and pair excitations are gapless, that is, $v_1,\ v_2\neq 0$. Here $\Delta\mathbf{N}^T=(\Delta N_1,\ \Delta N_2)$, $\Delta N_1,\ \Delta N_2$ are the numbers of added unpaired electrons and bound pairs, respectively; $\mathbf{D}^T=(D_1,D_2)$, $2D_1, 2D_2$ are differences in the numbers of positive and negative I_j, J_α numbers, respectively; $I_1^\pm=\Sigma I_j^\pm$, $I_2^\pm=\Sigma J_\alpha^\pm$ are sums of quantum numbers of particle–hole states near the right (+) and left (−) Fermi points of the k and λ seas. To include the vector potential we substitute

$$D_1 \to D_1 + \frac{N_a \nu}{2\pi}, \quad D_2 \to D_2 + \frac{N_a \nu}{\pi}.$$

By varying (16) over the vector potential A we obtain for the electric current

$$j = 2 \sum_{n=1,2} v_n (Z_{1n} D_1 + Z_{2n} D_2)(Z_{1n} + 2Z_{2n}). \tag{17}$$

For pair excitations we have $2D_2 = 1$, $D_1 = 0$,

$$j = 2 v_2 Z_{22}^2(\Lambda). \tag{18}$$

Therefore, the charge is equal to

$$q = 2Z_{22}^2(\Lambda). \tag{19}$$

As follows from (10), the charge may acquire any value in the interval $1/2 \leqslant q \leqslant 2$. In the case $\rho \to 1$ we have from (10)

$$q = 1 - \frac{1}{2\ln[C/(1-\rho)]}, \quad \rho = 1, \quad q = 1, \quad 0 < u < \infty.$$

In the limit $\rho \to 0$, $\rho/u \to 0$

$$q = \frac{1}{2}\left(1 + \frac{\rho}{2}\sqrt{1 + \frac{1^2}{u}}\right).$$

In the limit $u \ll 1$, $\Lambda \gg u$ $K(x) \approx \delta(x)$, $Z \to 1/\sqrt{2}$ and $q \to 1$. In the limit $u \to \infty$, $\Lambda \ll u$ we have $Z \approx 1$, $q \approx 2$. In strong magnetic fields $h > h_c$ we obtain

$$q = \left(1 + \frac{u}{2\pi\lambda_0}\right)^2.$$

For unpaired added particles the current and charge are found by substituting $2D_1 = 1$, $D_2 = 0$ in Eq. (16):

$$j = v_1 q, \quad q = Z_{11}(Q)[Z_{11}(Q) + 2Z_{21}(Q)] = 1 - 4\int K(2(\sin k - \lambda))\xi_{22}(\lambda)\, d\lambda.$$

In the case $u \to 0$ we have $q \to 0$, that is, the excitations do not carry an electric current. By using the Wiener–Hopf technique we find in the limit $\Lambda/u \gg 1$, $u \ll 1$ that

$$q \approx \sqrt{\frac{8}{\pi e}} \exp\left(-\frac{\Lambda\pi}{2u}\right).$$

In the opposite limit $u \to \infty$, $\Lambda \ll u$ we obtain $q \approx 1 - 4\Lambda/(\pi u)$.

Note that in this linear-spectrum approximation we find that particle–hole excitations do not carry current ($j = 0$), in accordance with previous results.

3.2. Repulsive interaction

3.2.1. Hole and particle states. A similar expressions can be derived for the case $U > 0$. For spin single–triplet excitations or particle–hole states we derived[8]

$$\widetilde{\rho}(k) = \frac{N_a \nu}{2\pi} + f(k) + \int_{-Q}^{Q} \widetilde{\rho}(k') \cos k' \overline{K} \times (\sin k - \sin k')\, dk', \tag{20}$$

where $f(k) = \Sigma (1/\pi)\tan^{-1}\{\exp[2\pi(\sin k - \lambda_i)/u]\}$ for spin excitations and $f(k) = \theta(k - k_i)$ for hole states, $\overline{\rho(k_j)} = N_a \rho(k_j)\delta k_j$, $\rho(k)$ is a known function for the ground state, and δk_j is the shift of a number k_j due to the excitation. For the energy, momentum and current we find

$$W = \sum \epsilon_0(k_j) = \sum \epsilon_0(k_j^0) + \sum \epsilon_0'(k_j)\delta k_j + \frac{1}{2}\sum \epsilon_0''(k_j)(\delta k_j)^2 \to \frac{1}{2}\int \frac{\epsilon_0'' \widetilde{\rho}^{\,2}}{N_a \rho(k)}\, dk,$$

where $\epsilon_0 = -2\cos k_j$, and $\delta k = \overline{\rho}(k)/N_a\rho(k)$, and

$$p = \int \widetilde{\rho}(k)\, dk, \quad j = \frac{1}{\Phi_0}\int_{-Q}^{Q} dk\, \frac{\epsilon_0'' \widetilde{\rho}\, \xi(k)}{\rho(k)}. \tag{21}$$

where $\xi(k)$ is the solution of Eq. (12). In accordance with the results,[8–10] we have $j \propto p\epsilon_0''(p_F)$. In the limit $u \to 0$ we again find $j = 2p\cos(\pi\rho/2)$. Equation (21) supplements the results of Refs. 8 and 9.

3.2.2. States with added particles. A similar treatment can be carried out for the repulsive model $u > 0$. In this case the subscripts 1.2 in Eq. (16) correspond to charge and spin degrees of freedom, respectively, so v_1 and v_2 are the charge and spin excitation velocities. In contrast to the $u < 0$ case both types of excitations are gapless. To include the orbital magnetic field we substitute

$$D_1 \to D_1 + \frac{N_a \nu}{2\pi}.$$

By varying with respect to the vector potential we obtain for the electric current

$$j = 2v_1(Z_{11}D_1 + Z_{21}D_2)Z_{11} = v_1(2D_1 + D_2)\xi^2(Q). \tag{22}$$

Here $2D_1$ ($2D_2$) denotes the difference in the numbers of positive and negative I_j (J_α), respectively. Changing the number of spin-up and spin-down electrons by $\Delta N_\uparrow$ and $\Delta N_\downarrow$, i.e., changing the total number of particles by $\Delta N_1 = \Delta N_\uparrow + \Delta N_\downarrow$ and the number of spin-down particles by $\Delta N_2 = \Delta N_1$, is equivalent to adding (removing) ΔN_1 extra I_j and ΔN_2 extra J_α values. The values I_J, J_α may be integer or half-integer, depending on the parity of the numbers N_1, N_2. Therefore the numbers D_1, D_2 depend on ΔN_1, ΔN_2 nontrivially, so that

$$D_1 = \frac{\Delta N_1 + \Delta N_2}{2} \pmod 1, \quad D_2 = \frac{\Delta N_1}{2} \pmod 1.$$

Adding a spin-up particle to the system $\Delta N_1 = 1$, $\Delta N_2 = 0$ we obtain $D_1 = \pm 1/2$, $D_2 = \mp 1/2$. Similarly, adding a spin-down particle corresponds to $\Delta N_1 = \Delta N_2 = 1$, $D_1 = 0$, $D_2 = \pm 1/2$. In both cases we find for the electric current

$$j = q v_1, \quad q = \xi^2(Q)/2. \tag{23}$$

Substituting solutions of Eq. (12) into (23) we find $1/2 \leqslant q \leqslant 1$ and

$\sin Q/u \ll 1$, $q = 1/2 + \sin Q \ln 2/(\pi u)$;
$\rho = 1$, $Q = \pi$, $q = 1/2$, $0 < u < \infty$;
$\rho \to 1$, $q = 1/2 + \ln 2(1-\rho)/u f(u)$;
$\rho \ll 1, u$; $q = 1/2 + \rho \ln 2/u$;
$u \gg 1$, $q = 1/2 + \sin \pi\rho \ln 2/(\pi u)$;
$u \ll \sin Q$, $q = 1 - u/(4\pi \sin Q)$.

These results complement our earlier results.[8–10]

4. CONCLUSIONS

We have considered the electron currents of excitations in the one-dimensional Hubbard model. We found that for particle–hole excitations with a small momentum the current is proportional to the momentum and to the dispersion of the Fermi velocity. Therefore, this current is absent in the linear spectrum approximation. We have calculated the currents and charges of states with added particles (unpaired or paired bound electrons in the case $u<0$). We found that the charge is noninteger and depends continuously on the band filling and the onsite potential u. Note that a spin–charge decoupling has no place in the case of the Hubbard model in the magnetic field. The contributions of spin- and charge-density waves cannot be described by two independent effective Hamiltonians. Charge- and spin-density waves interact. The physical values (the spectrum of conformal operator dimensions, the electric charges in question, etc.) are determined by the 2×2 dressed charge matrix rather than two scalar coupling constants. As a result both spin and charge excitations cary the electron current.

I am very grateful to S. A. Brazovskii for a stimulated discussions. This work was supported by the Russia Fund for Fundamental Research under Grant No 960217791, and partially by the ''Statphysics'' program of the Ministry of Science.

[1] S. Brazovskii and N. Kirova, Sov. Sci. Rev. 5, 99 (1984).
[2] S. Brazoskii and S. Matveenko, Zh. Éksp. Teor. Fiz. **87**, 1398 (1984) [Sov. Phys. JETP **60**, 804 (1984)].
[3] F. Woynarovich, Z. Phys. B **85**, 269 (1991).
[4] E. H. Lieb and F. Y. Wu, Phys. Rev. Lett. **20**, 1445 (1968).
[5] H. Frahm and V. E. Korepin, Phys. Rev. B **42**, 10553 (1990); **43**, 5653 (1991).
[6] N. M. Bogolubov and V. E. Korepin, Int. J. Mod. Phys. B **3**, 427 (1989).
[7] N. M. Bogolobov and V. E. Korepin, Theor. Math. Phys. **832**, 331 (1988).
[8] S. I. Matveenko, Zh. Éksp. Teor. Fiz. **94**, 213 (1988) [Sov. Phys. JETP **67**, 1416 (1988)].
[9] S. Matveenko and S. Brazovskii, Zh. Éksp. Teor. Fiz. **78**, 892 (1994) [JETP **105**, 1653 (1994)].
[10] S. Brazovskii, S. Matveenko, and P. Nozieres, J. de Phys. I. France 4, 571 (1994).
[11] F. Woynarovich, J. Phys. A: Math. Gen. 22, 4243 (1989).

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

The magnetostriction of Invar alloys

A. Yu. Romanov and V. P. Silin

P. N. Lebedev Physical Institute, Russian Academy of Sciences, 117924 Moscow, Russia
(Submitted 5 June 1997)
Zh. Éksp. Teor. Fiz. **113**, 213–227 (January 1998)

The principles of the theory of a phase transition into a magnetically ordered state are formulated for Invar alloys and other similar inhomogeneous ferromagnets, for which the concept of a local Curie temperature distribution corresponding to the experimentally observed broadened temperature interval of the transition into the ferromagnetic state has existed for 10 years. A method is proposed for obtaining information about the local temperature distribution from experimental data on the change in the properties of magnets in response to a change in temperature. For iron–nickel–chromium alloys it is shown how to obtain the temperature dependence of the magnetostrictional susceptibility of the paraprocess from data on the magnetic contribution to the thermal expansion coefficient. This confirms the important role of the local Curie temperature distribution, and it also indicates a need for new analysis of experimental data on temperature-broadened magnetic ordering phase transitions. © *1998 American Institute of Physics.* [S1063-7761(98)01501-7]

1. INTRODUCTION

The unflagging interest in Invar alloys, which were discovered one hundered years ago,[1] is due, for one thing, to the virtually unlimited technical applications determined by the unique properties of such alloys and, for another, to the status of physical investigations of Invar alloys, which is far from a complete understanding.[2] It is generally believed that the unique properties of Invar alloys are associated with magnetism, since they are manifested in the transition of the alloys into the ferromagnetic state. For example, the unique decrease of the volume thermal expansion coefficient of the iron–nickel alloy $Fe_{65}Ni_{35}$ appears, though in a comparatively wide region, near the Curie temperature of the phase transition into a magnetically ordered state.[3] According to the Landau theory of second-order phase transitions[4] an abrupt change in the thermal expansion coefficient and other properties with the establishment of magnetic ordering should occur at the Curie temperature. In reality, however, no sharply expressed abrupt changes in properties at a transition into the ferromagnetic state are observed in Invar alloys[2] (see also, for example, Ref. 5). Investigators have become so accustomed to this state of affairs that ordinarily no attempt is made to give a quantitative description of magnetic ordering phase transitions, broadened over a wide temperature interval, in Invar alloys. The present paper is devoted to filling this lacuna. We shall discuss here an important property of Invar alloys, magnetostriction.

Our analysis is based on the old and well-known view that Invar alloys are inhomogeneous with respect to their structure or composition[6] (see also Ref. 7–13). It is thought that such an inhomogeneity results in an inhomogeneity-dependent local Curie temperature distribution, which corresponds to a continuous transition into the ferromagnetic state with a broadened temperature interval, determined by this distribution, Although this view is widely held, it should be noted that there is a definite conflict in the statements made by the authors of various treatments of the quantitative effect of the inhomogeneity in Invar alloys. For example, Refs. 13 and 14 report a small difference (several degrees) between the temperatures of the peak in the specific heat, the peak in the thermal expansion coefficient, and the Curie point obtained from magnetic measurements. On the other hand, the establishment of the existence of tails in the magnetization (see, for example, Refs. 2 and 6) indicates a substantial (at least 10°) high-temperature shift of the onset temperature of a transition into the magnetically ordered ferromagnetic state, as compared with the lower value of the temperature ordinarily called the Curie temperature. The latter is determined, specifically, by the method of coefficients of the thermodynamic potential.[6] In the face of such a conflicting view of the properties of broadened phase transitions, it is necessary to make a systematic theoretical analysis based on the approach indicated in the monograph by K. P. Belov.[6]

The first step in this direction was taken in Ref. 15, where the problem of an approximate, in many ways qualitative, determination of the local Curie temperature distribution of some Invar alloys was solved on the basis of the rough scaling proposed there.

In the present paper we propose an elaboration of the Belov approach,[6] making possible a practical quantitative analysis of experimental data near the broadened phase transition of Invar alloys. This approach is applied to the analysis of magnetostriction. In Sec. 2 we present the theoretical assumptions that are required for our analysis and on the basis of which the temperature dependences of different phenomena occurring near a broadened magnetic-ordering transition can be related to one another. Section 3 is devoted to the rough-scaling approximation indicated in Ref. 15. The results of an analysis, based on the approach developed, of the experimental data from Ref. 16 on the magnetostriction of iron-nickel Invar alloys $Fe_{65}Ni_{35-x}Cr_x$ are presented in Sec. 4. We employ below the data of Ref. 16 for the thermal expansion coefficient and for the magnetostrictional susceptibility of the paraprocess (induced magnetostriction) only for alloys with chromium concentrations $x=0$, 2.5, and 5,

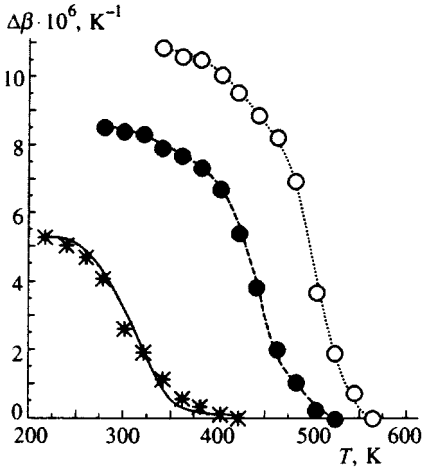


FIG. 1. Temperature dependence of the magnetic contribution to the thermal expansion coefficient for three alloys $Fe_{65}Ni_{35-x}Cr_x$. The dots represent the experimental results of Ref. 16: ○ ($x=0$), ● ($x=2.5$), and ∗ ($x=5$). The curves represent continuous approximations of the experimental points: solid trace—$x=5$, dashed trace—$x=2.5$, dotted trace—$x=0$.

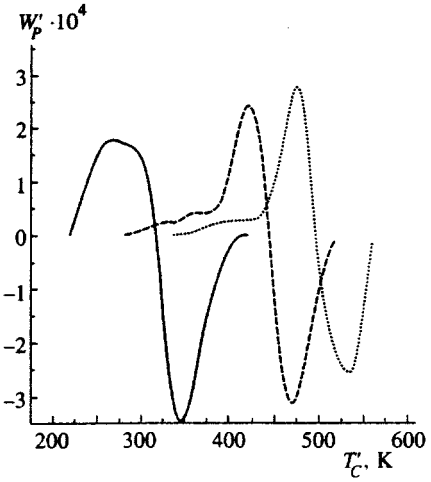


FIG. 3. Functions $\partial W(T_C')/\partial P$ for the three alloys $Fe_{65}Ni_{35-x}Cr_x$. The dotted trace corresponds to $x=0$, the dashed trace to $x=2.5$, and the solid trace to $x=5$.

since as x increases it becomes difficult to distinguish the magnetic contribution to the thermal expansion coefficient and the results are inaccurate. The corresponding discussion is given in Sec. 4. In Fig. 1, the dotted, dashed, and solid curves are for the alloys $Fe_{65}Ni_{35-x}Cr_x$ with $x=0$, 2.5, and 5, respectively, and they characterize the magnetic contribution to the thermal expansion coefficient. The dots, whose sizes correspond to the experimental errors, represent the data from Fig. 1 in Ref. 16. These experimental points are represented by open circles ($x=0$), filled circles ($x=2.5$), and asterisks ($x=5$). The curves interpolate the experimental temperature dependence. As an illustration, the result of our analysis is presented in Fig. 2, where experimental data for magnetostrictional susceptibility of the paraprocess are presented as dots taken from Fig. 3 of Ref. 16. The curves in Fig. 2 were constructed for the same susceptibility on the basis of the experimental data of Fig. 1 and the approach

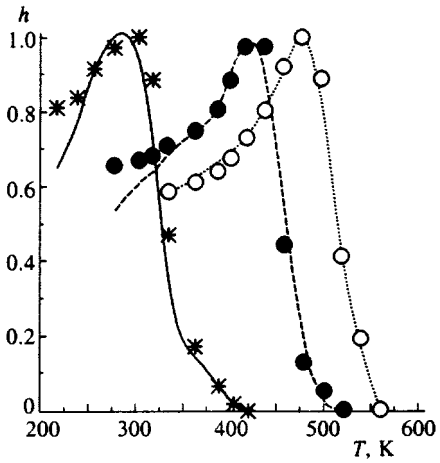


FIG. 2. Temperature dependences of the induced magnetostriction, normalized to 1. The dots represent the experimental results of Ref. 16: ○ ($x=0$), ● ($x=2.5$), and ∗ ($x=5$). The curves are computational results: solid trace—$x=5$, dashed trace—$x=2.5$, and dotted trace—$x=0$.

proposed in the present paper for analyzing the experimental data on the temperature dependence of the properties of Invar alloys. A general discussion of the results obtained in this paper is given in concluding section, Sec. 5.

It should be underscored that the material presented below represents the required first step in understanding the inhomogeneity-determined properties of Invar alloys. In so doing, we do not deny that magnetic fluctuations exist; we believe that to understand the temperature dependences in detail allowance must be made for magnetic fluctuations. However, our main thrust is to determine the role of the distribution of nonuniform Curie temperatures, primarily because even though assertions concerning this have been made for many years in the literature, no real attempts have been made to give a quantitative interpretation of the properties of Invar alloys and other ferromagnets on this basis.

2. THEORETICAL MODEL

Our approach goes back to the ideas presented in Ref. 6, where it was proposed that a thermodynamic potential of the form

$$\Phi_M(P,T,[M])=\frac{1}{V}\int d\mathbf{r}\left\{\frac{\alpha}{2}(T-T_C(\mathbf{r}))M^2+\frac{b}{4}M^4\right\}. \tag{2.1}$$

be used to describe the properties of inhomogeneous magnets. Here the local Curie temperature $T_C(\mathbf{r})$ is a function of the coordinates. This dependence is the reason for the broadened magnetic-ordering transition. In addition to this, $T_C(\mathbf{r})$ and the coefficients α and β in the thermodynamic potential are functions of the pressure P. Equation (2.1) is an extension of the Ginzburg–Landau thermodynamic potential to the case of an inhomogeneous magnet. Here gradient terms are neglected. We do this only because the effects discussed below are insensitive to their presence. In the future, we intend to return to a more complete description of inhomogeneous ferromagnets in order to study effects that cannot be understood without taking the gradient terms into account.

Since Eq. (2.1) is a functional of $M(\mathbf{r})$ and in the model (2.1) the spatial inhomogeneity is due only to the spatial dependence of the local Curie temperature $T_C(\mathbf{r})$, the thermodynamic potential (2.1) can be written in the form[6,17]

$$\Phi_M(P,T,[M])=\int dT'_C W(T'_C) \times\left\{\frac{\alpha}{2}(T-T'_C)M^2+\frac{b}{4}M^4\right\}. \quad (2.2)$$

Now, in contrast to Eq. (2.1), the local magnetization density as a function of T'_C, and $W(T'_C)$ is determined by the relation[15]

$$W(T'_C)=\frac{1}{V}\int d\mathbf{r}\ \delta(T'_C-T_C(\mathbf{r})). \quad (2.3)$$

With this definition the function (2.3) plays the role of a local Curie temperature distribution function, normalized to 1. The average magnetization can be represented as

$$\langle M\rangle=\frac{1}{V}\int d\mathbf{r}'M(\mathbf{r}')=\int dT'_C W(T'_C)M(T'_C).$$

The approach based on the thermodynamic potential (2.2) with the distribution function $W(T'_C)$ prescribed *a priori* was proposed in Ref. 6 for understanding broadened magnetic-ordering phase transitions. The qualitative properties of the phenomena accompanying a transition into the ferromagnetic state were studied, assuming a prescribed distribution function of the local temperature, in Refs. 12 and 17–19.

Information about the distribution function $W(T'_C)$ was obtained in Ref. 15 using rough scaling, based on the model (2.1) for analyzing experimental data. In the present report we likewise will not make any detailed assumptions about $W(T'_C)$, but will take the next step following Ref. 15. Specifically, we shall examine the relations between the experimental laws that make it possible to obtain the model (2.1).

We shall present below somewhat more consequences of Eq. (2.1) than required to discuss magnetostriction, whose properties will be interpreted with the aid of the model (2.1), in iron–nickel–chromium Invar alloys. Equations that will aid in understanding the basis for the rough scaling, introduced in Ref. 15, of the temperature dependence of Invar alloys will also be derived.

We note that with the modern understanding of ferromagnetism of metals it should be obvious that Eq. (2.1) takes into account to some degree the thermal magnetic fluctuations, which determine, specifically, the linear variation as a function of temperature beyond a narrow range of temperatures.

Proceeding now to the equations required for what follows, we determine the average magnetic field intensity H by varying the functional (2.2) with respect to the local magnetization density $M(T'_C)$:

$$H=\alpha(T-T'_C)M+bM^3. \quad (2.4)$$

This makes it possible to use below the thermodynamic potential

$$\Phi_H(P,T,H)=\Phi_M(P,T,[M])-H\langle M\rangle. \quad (2.5)$$

The latter formula gives for the entropy

$$S_H=-\frac{\alpha}{2}\int dT'_C W(T'_C)M^2. \quad (2.6)$$

In the limit of zero magnetic field, $H=0$, Eq. (2.4) gives for the spontaneous local magnetization

$$M^2=\frac{\alpha}{b}(T'_C-T) \quad \text{for } T<T'_C. \quad (2.7)$$

so that Eq. (2.6) in this limit gives

$$S_{H=0}=-\frac{\alpha^2}{2b}\int_T^{T_C^0} dT'_C W(T'_C)(T'_C-T), \quad (2.8)$$

where T_C^0 is the maximum local Curie temperature. The following relation between the specific heat and the local Curie temperature distribution function is an immediate consequence:

$$C_{H=0}(T)=T\frac{\alpha^2}{2b}\int_T^{T_C^0} W(T'_C)dT'_C. \quad (2.9)$$

In Ref. 15 it was proposed that this relation be used to define $W(T)$ on the basis of the experimental dependence of the magnetic contribution to the specific heat, corresponding to Eq. (2.9). Specifically, according to Eq. (2.9), we have

$$W(T)=-\frac{2b}{\alpha^2}\frac{d[C_{H=0}/T]}{dT}. \quad (2.10)$$

Proceeding now to the equations required for describing magnetostriction, we shall make some simplifying assumptions. First and foremost, we note that it is reasonable to use the expression (2.1) itself only for comparatively small values of the magnetization, which we shall assume is the case everywhere in constructing our description of the magnetic properties in the transitional region for the establishment of ferromagnetism, corresponding to the temperature range where $W(T)$ is actually different from zero. Further, remembering that the magnetization is comparatively small, when differentiating the thermodynamic potential with respect to pressure we shall neglect the derivatives of the coefficients α and b in the thermodynamc potential. This corresponds to the standard approach for studying the properties of homogeneous magnets near the Curie temperature that employs an assumption about the largest effect due to the pressure dependence of the Curie temperature. In our model the latter corresponds to the pressure dependence of the local Curie temperature distribution function $W(T'_C)$. In this approximation Eq. (2.5) makes it possible to describe magnetostriction by means of the relation

$$\frac{\Delta V}{V}=\omega_M=\int dT'_C\frac{\partial W(T'_C)}{\partial P} \times\left\{\frac{\alpha}{2}(T-T'_C)M^2+\frac{b}{4}M^4-HM\right\}. \quad (2.11)$$

Hence, taking account of Eq. (2.4), we obtain for the magnetostrictional thermal expansion coefficient:

$$\beta_H=\left(\frac{\partial\omega}{\partial T}\right)_{P,H}=\int dT_C'\,\frac{\partial W(T_C')}{\partial P}\,\frac{\alpha}{2}\,M^2. \tag{2.12}$$

In the limit of zero magnetic field Eq. (2.12) becomes

$$\beta_{H=0}=\frac{\alpha^2}{2b}\int_T^{T_C^0}dT_c'\,\frac{\partial W(T_C')}{\partial P}\,(T_C'-T). \tag{2.13}$$

This formula makes it possible to use the experimentally determined temperature dependence $\beta_{H=0}(T)$ to determine the following quantities:

$$\int_T^{T_C^0}dT_C'\,\frac{\partial W(T_C')}{\partial P}=-\frac{2b}{\alpha^2}\,\frac{d\beta_{H=0}(T)}{dT}, \tag{2.14}$$

$$\frac{\partial W(T)}{\partial P}=\frac{2b}{\alpha^2}\,\frac{d^2\beta_{H=0}(T)}{dT^2}. \tag{2.15}$$

The magnetostriction susceptibility of the paraprocess (induced magnetostriction) can be found by differentiating Eq. (2.11) with respect to the magnetic field

$$h=\left(\frac{\partial\omega_M}{\partial H}\right)_{P,T}=-\int dT_C'\,\frac{\partial W(T_C')}{\partial P}\,M. \tag{2.16}$$

In the limit of zero magnetic field this formula becomes

$$h=-\sqrt{\frac{\alpha}{b}}\int_T^{T_C^0}dT_C'\,\frac{\partial W(T_C')}{\partial P}\,\sqrt{T_C'-T}. \tag{2.17}$$

If the temperature dependence of the thermal expansion coefficient is known, then according to Eq. (2.15) the expression (2.17) can be put in the form

$$h=-\frac{2}{\alpha}\,\sqrt{\frac{b}{\alpha}}\int_T^{T_C^0}dT_C'\,\frac{d^2\beta_{H=0}(T_C')}{dT_C'^2}\,\sqrt{T_C'-T}. \tag{2.18}$$

The possibilities of the model that we used for interpreting experiments can be understood by comparing the temperature dependence, calculated in this manner, of the magnetostriction susceptibility of the paraprocess with the experimentally measured dependence.

Finally, we obtain using Eq. (2.5) the magnetic contribution to the compressibility

$$\kappa_H=-\int dT_C'\,\frac{\partial^2W(T_C')}{\partial P^2}\times\left\{\frac{\alpha}{2}\,(T-T_C')M^2+\frac{b}{4}\,M^4-HM\right\}. \tag{2.19}$$

In the limit of a zero magnetic field, taking into account Eq. (2.7), we have

$$\kappa_{H=0}=\frac{\alpha^2}{4b}\int_T^{T_C^0}dT_C'\,\frac{\partial^2W(T_C')}{\partial P^2}\,(T-T_C')^2. \tag{2.20}$$

Hence it follows that if the temperature dependence $\kappa_{H=0}(T)$ is determined experimentally, then we have from Eq. (20)

$$\frac{d\kappa_{H=0}(T)}{dT}=\frac{\alpha^2}{2b}\int_T^{T_C^0}dT_C'\,\frac{\partial^2W(T_C')}{\partial P^2}\,(T-T_C'). \tag{2.21}$$

$$\frac{d^2\kappa_{H=0}(T)}{dT^2}=\frac{\alpha^2}{2b}\int_T^{T_C^0}dT_C'\,\frac{\partial^2W(T_C')}{\partial P^2}, \tag{2.22}$$

$$\frac{d^3\kappa_{H=0}(T)}{dT^3}=-\frac{\alpha^2}{2b}\,\frac{\partial^2W(T)}{\partial P^2}. \tag{2.23}$$

Since the compressional susceptibility of the paraprocess, according to Eq. (2.19), has the form

$$\left(\frac{\partial\kappa_H}{\partial H}\right)_{P,T}=\int dT_C'\,\frac{\partial^2W(T_C')}{\partial P^2}\,M, \tag{2.24}$$

we obtain in the limit $H=0$

$$\left(\frac{\partial\kappa_H}{\partial H}\right)_{H=0}=\sqrt{\frac{\alpha}{b}}\int_T^{T_C^0}dT_C'\,\frac{\partial^2W(T_C')}{\partial P^2}\,\sqrt{T_C'-T}. \tag{2.25}$$

Using Eq. (2.23) we find

$$\left(\frac{\partial\kappa_H}{\partial H}\right)_{H=0}=-\frac{2}{\alpha}\,\sqrt{\frac{b}{\alpha}}\int_T^{T_C^0}dT_C'\,\frac{d^3\kappa_{H=0}(T_C')}{dT_C'^3}\,\sqrt{T_C'-T}. \tag{2.26}$$

Experiments performed in recent years on the pressure dependence of the elastic moduli open up the possibility of determining the third pressure derivative of the local Curie temperature distribution function. Indeed, we have

$$\left(\frac{\partial\kappa_H}{\partial P}\right)_{H,T}=-\int dT_C'\,\frac{\partial^3W(T_C')}{\partial P^3}\left\{\frac{\alpha}{2}\,(T-T_C')M^2+\frac{b}{4}\,M^4-HM\right\}. \tag{2.27}$$

Hence we obtain, similarly to Eq. (2.23),

$$\frac{d^3}{dT^3}\left(\frac{\partial\kappa_H}{\partial P}\right)_{H=0}=-\frac{\alpha^2}{2b}\,\frac{\partial^3W(T)}{\partial P^3}. \tag{2.28}$$

The equations presented in this section show how information about the local Curie temperature distribution function and its pressure derivatives can be obtained on the basis of experimental data. Three curves for $\partial W(T_C')/\partial P$, which were obtained according to Eq. (2.15) by differentiating the curves in Fig. 1, are presented in Fig. 3 for the three alloys $Fe_{65}Ni_{35-x}Cr_x$ for $x=0$, 2.5, and 5.

3. ON ROUGH SCALING

In Ref. 15 we proposed on the basis of experimental data a rough scaling of the temperature dependence of the properties of Invar alloys in the transitional temperature range where, in the terminology of Ref. 6, magnetization tails are present, i.e. ferromagnetism is established gradually in the entire volume of the magnet. This scaling gave surprising agreement with experimental data for permalloy, the iron–nickel alloy $Fe_{65}Ni_{35}$. In this section we shall substantiate the possibility of rough scaling, proposed in Ref. 15, of the broadened phase transition of Invar alloys.

To begin with, we note that if the local Curie temperature distribution function has the form

$$W(T_C')=\delta(T_C'-T_C), \tag{3.1}$$

then our analysis reduces to the standard Ginzburg–Landau theory of homogeneous magnets. Here T_C is the ordinary Curie temperature, information about which is contained in many works, including data on its pressure dependence $T_C(P)$. In reality, for Invar alloys the ordinary Curie temperature is probably not adequately defined. On the other hand, the arguments presented in the preceding section make it possible to assume that the temperature T_m of the maximum of the local Curie temperature probability distribution is at least close to what is designated as the Curie temperature in experiments. In this connection, we state the assumption that

$$W(T_C')=w(T_C'-T_m(P)), \tag{3.2}$$

where the strongest pressure dependence is due to the function $T_m(P)$. Then

$$\frac{dW(T)}{dP}=-\frac{dw(T-T_m(P))}{dT}\frac{dT_m}{dP}, \tag{3.3}$$

$$\frac{d^2W(T)}{dP^2}=\frac{d^2w(T-T_m(P))}{dT^2}\left(\frac{dT_m}{dP}\right)^2 - \frac{dw(T-T_m(P))}{dT}\frac{d^2T_m}{dP^2}. \tag{3.4}$$

Using the equation (3.3), we now put Eq. (2.13) into the form

$$\beta_{H=0}=-\frac{\alpha^2}{2b}\frac{dT_m}{dP}\int_T^{T_C^0}dT_C'(T_C'-T)\frac{dw(T_C'-T_m(P))}{dT_C'} =\frac{\alpha^2}{2b}\frac{dT_m}{dP}\int_T^{T_C^0}dT_C'W(T_C'). \tag{3.5}$$

If in accordance with Ref. 15 we now use the function

$$f(T)=\int_T^{T_C^0}dT_C'W(T_C'), \tag{3.6}$$

then the equations (3.5) and (2.9) can be rewritten in the rough-scaling form proposed in Ref. 15:

$$\frac{C_{H=0}(T)}{T}=\frac{\alpha^2}{2b}f(T),\quad \beta_{H=0}(T)=\frac{\alpha^2}{2b}\frac{dT_m}{dP}f(T). \tag{3.7}$$

Finally, if the expression (3.4) is substituted into the right-hand side of Eq. (2.20), then we obtain for the compressibility

$$\kappa_{H=0}=\frac{\alpha^2}{2b}\left(\frac{dT_m}{dP}\right)^2\int_T^{T_C^0}dT_C'W(T_C') + \frac{\alpha^2}{4b}\frac{d^2T_m}{dP^2}\int_T^{T_C^0}dT_C'(T'-T)W(T_C'). \tag{3.8}$$

Neglecting the second term in this equation gives a third equation of rough scaling:[15]

$$\kappa_{H=0}(T)=\frac{\alpha^2}{2b}\left(\frac{dT_m}{dP}\right)^2 f(T). \tag{3.9}$$

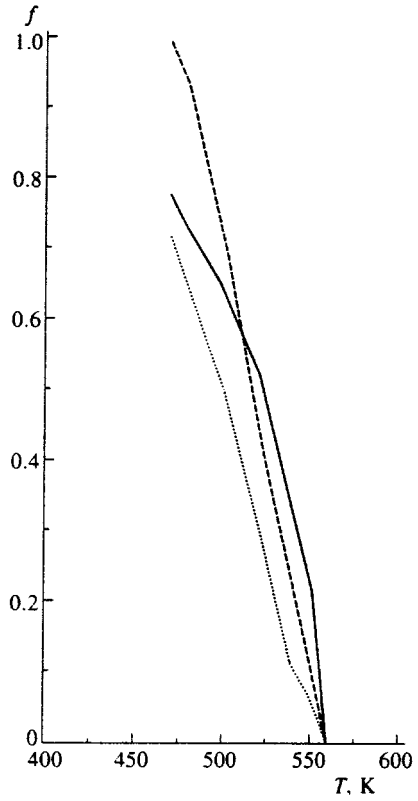


FIG. 4. Functions $f(T)$ obtained from the experimental data: dashed curve—specific heat measurements,[20] dotted curve—measurements of the thermal expansion coefficient,[21] solid curve—compressibility measurements.[22]

The satisfactory agreement of the rough-scaling formulas (3.7) and (3.9) with experiment indicates the relative smallness of the second term in Eq. (3.8). To illustrate the effectiveness of rough scaling we display in Fig. 4 the experimental temperature dependences $C_{H=0}(T)/T$ (dashed curve[20]), $\beta_{H=0}(T)$ (dotted curve[21]), and $\kappa_{H=0}(T)$ (solid curve[22]), all normalized to 1 and characterizing the function $f(T)$ corresponding to the alloy $Fe_{65}Ni_{35}$. A discussion of the construction of these curves is given in the next section.

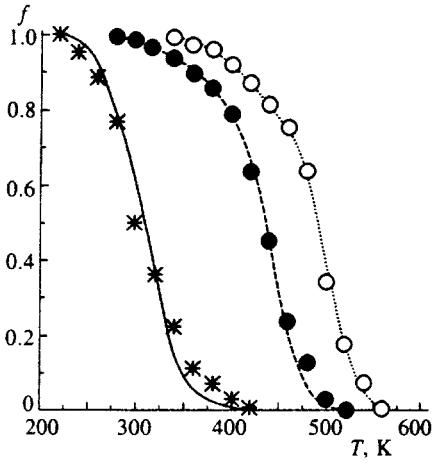


FIG. 5. Functions $f(T)$ obtained from Fig. 1. The symbols have the same meaning as in Fig. 1.

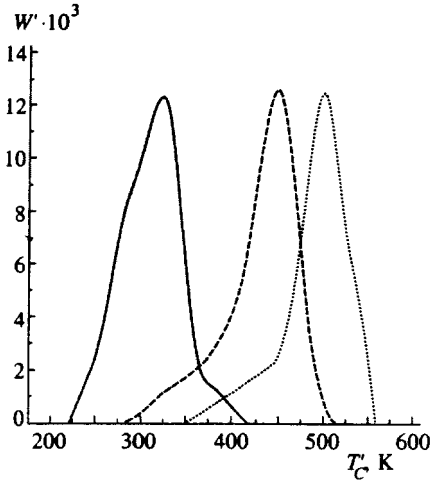


FIG. 6. Distribution functions $W(T'_C)$ of the local Curie temperatures for three alloys, as determined from the data in Fig. 5 for $f(T)$. The dotted curve corresponds to $x=0$, the dashed curve to $x=2.5$, and the solid curve to $x=5$.

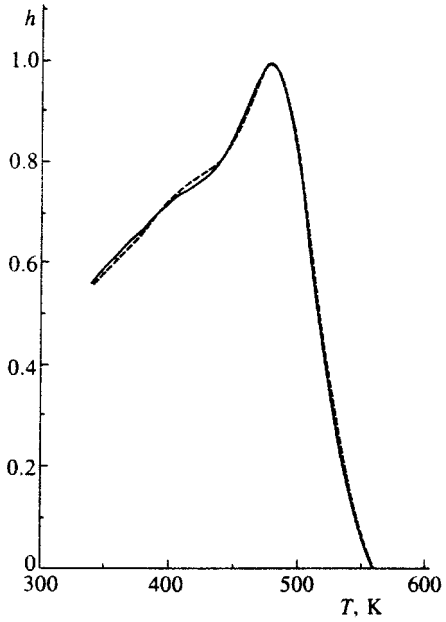


FIG. 7. Temperature dependences, normalized to 1, of the magnetostriction calculated from Eqs. (3.10)—solid curve and (2.18)—dashed curve.

Figure 5 displays three functions $f(T)$ constructed in accordance with Eq. (3.7) according to the data in Fig. 1 for the iron–nickel–chromium alloy $Fe_{65}Ni_{35-x}Cr_x$ for three chromium concentrations $x=0$, 2.5, and 5. Differentiating the curves in Fig. 5 according to Eq. (3.6) makes it possible to use the rough scaling approach to obtain the local Curie temperature distribution function of Invar alloys. The results are illustrated in Fig. 6, where the temperatures T_m of the maximum of the local Curie temperature distribution are 500 K ($x=0$), 440 K ($x=2.5$), and 320 K ($x=5$). These values are very close to the Curie temperatures indicated on the plots of the temperature dependence of the thermal expansion coefficient in Ref. 16: 490 K ($x=0$), 430 K ($x=2.5$), and 300 K ($x=5$), respectively.

The approximation (3.2) makes it possible to write Eq. (2.17) in the form

$$h(T)=-\frac{1}{2}\sqrt{\frac{\alpha}{b}}\frac{dT_m}{dP}\int_T^{T_C^0}\frac{dT'_C W(T'_C)}{\sqrt{T'_C-T}}. \tag{3.10}$$

We now have a new possibility for describing the magnetostriction susceptibility of the paraprocess with the aid of this formula and the local Curie temperature distribution function found from experiment on the basis of Eq. (3.7). Figure 7 presents for the alloy $Fe_{65}Ni_{35}$ the temperature dependence $h(T)$ obtained in this manner (solid curve) and according to Eq. (2.18) (dashed curve). The difference is very small.

4. COMPARISON OF THE EXPERIMENTAL DATA

In using the approach presented for analyzing experiments it is necessary to establish the maximum local Curie temperature $T'_C=T_C^0$ at which a magnetically ordered state of a ferromagnetic alloy starts to form. The corresponding arguments are clearest in the approach presented in Sec. 3. For this, we employ the relation (3.3) and write down the equation

$$\omega_M=-\frac{\alpha}{2}\frac{dT_m}{dP}\int dT'_C W(T'_C)M^2. \tag{4.1}$$

Let us compare this expression to the well-known Moriya–Usami formula[23]

$$\omega=(C_0/K)M_L^2, \tag{4.2}$$

where C_0 is the magnetoelastic constant, K is the bulk modulus, and M_L^2 is the squared magnetization averaged over thermal fluctuations. In the standard theory of homogeneous magnets the magnetoelastic constant is proportional to the pressure derivative of the Curie temperature. Our analysis yields

$$\frac{C_0}{K}=-\frac{\alpha}{2}\frac{dT_m}{dP}. \tag{4.3}$$

The combination

$$\int dT'_C W(T'_C)M^2=M_0^2 \tag{4.4}$$

arising in Eq. (4.1) is the squared magnetization averaged over the local Curie temperature distribution. It is natural to assume that the total mean squared magnetization consists of the magnetization averaged over both the thermal fluctuations and the local Curie temperature distribution. For this reason, we employ a generalization of Eqs. (4.1) and (4.2) in the form (4.2), but we shall assume that

$$M_L^2=M_0^2+2\,\delta m_t^2+\delta m_l^2. \tag{4.5}$$

Here, in contrast to the standard Moriya–Usami formula, M_0^2 is determined by Eq. (4.4), while ordinarily such an expression is taken to mean the squared average magnetization. Aside from this, in accordance with our approach the mean squared transverse δm_t^2 and longitudinal δm_l^2 fluctuations should correspond not only to thermal (Bose) averaging but

also averaging over the local Curie temperature distribution. Leaving a detailed theory of such fluctuations for subsequent publications, we shall use here some general assumptions. Just as in the standard theory with one Curie temperature, in the $H=0$ limit the quantity M_0^2 equals zero for $T>(T_C')_{\max}=T_C^0$. At the same time δm_t^2 and δm_l^2 are different from zero in the paramagnetic region, which is why the temperature dependence $\omega(T)$ deviates from the ordinarily discussed Grüneisen extrapolation of $\omega_{\text{par}}(T)$ from the paramagnetic temperature range located comparatively far from the region of the phase transition. This signifies that the onset temperature of magnetic ordering cannot be identified with the temperature at which $\omega(T)$ differs from $\omega_{\text{par}}(T)$. Moreover, $\omega(T)-\omega_{\text{par}}(T)$ cannot be identified with the magnetic contribution ω_M determined by our formulas from Sec. 2, since it characterizes only the part of the magnetostriction that is not due to thermal fluctuations.

In the quantitative analysis of experiments we employ the thermal expansion coefficient. The temperature dependence of δm_t^2 and δm_l^2 gives the thermal-fluctuation contribution to the thermal expansion coefficient. On the other hand, their derivatives with respect to the magnetic field are equal to zero in the paramagnetic region in the limit $H=0$. This circumstance makes it possible to assume that the temperature $T_C^0=(T_C')_{\max}$ at which local spontaneous magnetization first appears equals the experimentally measured temperature at which the magnetostrictional susceptibility of the paraprocess becomes different from zero. The values of such temperatures for the iron–nickel–chromium alloys of interest to us were determined in Ref. 16.

Determining in this manner the maximum local Curie temperature we note that, for example, according to Fig. 1 of Ref. 16 the difference between the experimental value $\beta(T_C^0)$ and the Grüneisen paramagnetic extrapolation $\beta_{\text{par}}=(\partial\omega_{\text{par}}(T)/\partial T)$ for the alloy $Fe_{65}Ni_{35}$ equals $\delta\beta(T_C^0)=2.2\cdot 10^{-6}$ K^{-1}. This is 17% of the largest difference $\delta\beta(T)=\beta(T)-\beta_{\text{par}}$ in the region of the broadened transition, equal to $\delta\beta_{\max}=13.1\cdot 10^{-6}$ K^{-1}. In this connection, we extract from the data in Fig. 1 of Ref. 16 the magnetic contribution $\beta_M(T)$ of interest to us as follows. First, we obtain $\delta\beta(T)$ as the difference of the experimentally measured value of the thermal expansion coefficient $\beta(T)$ and the Grüneisen extrapolation β_{par}. Second, we subtract $\delta\beta(T_C^0)$ from $\delta\beta(T)$. These are the values of $\Delta\beta(T)=\delta\beta(T)-\delta\beta(T_C^0)$ that are presented in Fig. 1 and used in our analysis when comparing the results of the theory described above with the magnetostriction susceptibility data for the paraprocess. A discussion of such a comparison is given in the concluding section.

The curve $\beta_{H=0}(T)$ presented in Fig. 4 corresponds to the magnetic contribution $\Delta\beta(T)$ obtained by the procedure described here. A similar procedure of subtracting the fluctuation contribution at $T=T_C^0$ is used for the magnetic contribution, presented in Fig. 4, to the specific heat and compressibility. This distinguishes Fig. 4 from Fig. 1 of Ref. 15, since our new determination of the onset of the region of a broadened transition narrows the temperature range where spatial variation of the local Curie temperatures appears.

5. DISCUSSION

The analysis performed above on the basis of the simple theoretical model (2.1) and (2.2) made it possible to give a comparatively simple method for determining the properties of the local Curie temperature distribution function (2.3) from experimental data on the specific heat, magnetostriction, and magnetoelasticity. The relations obtained make it possible to interrelate the temperature dependences of different properties of spatially inhomogeneous ferromagnets. It was shown for the example of magnetostriction how the temperature dependence of the magnetostriction susceptibility of the paraprocess can be determined from experimental data on the thermal expansion coefficient. Our model is helpful for identifying the role of the spatial inhomogeneity of ferromagnets such as Invar alloys, as well as other magnetic materials in which magnetic-ordering phase transitions are characterized by the presence of magnetization tails.[6]

The possible approaches presented in Secs. 2 and 3 and the surprising agreement between the results obtained using them, as demonstrated in Fig. 7, are the first steps in the construction of models of the properties of the local Curie temperature distribution function. Rough scaling[15] is the simplest method for analyzing experimental data. For this reason, the establishment of the meaning of such scaling, as proposed in Sec. 3, is important.

In conclusion, we underscore that despite its demonstrated success the approach proposed for analyzing experiments requires elaboration regarding the role of magnetic thermal fluctuations in inhomogeneous ferromagnets as well as other properties, requiring a more general analysis than Eq. (2.1), taking account of the spatial derivatives of the local magnetization.

Finally, we note that the situation described in the present paper is characteristic not only of ferromagnets. For example, according to Ref. 24, a general property of all ferroelectric crystals is the broadening of the phase transition with increasing defect density. It is believed, in complete analogy to Ref. 6, that there exists a distribution of local transition temperatures whose values depend on the distribution of defects over different local regions in the crystal. The approach described above for explaining the properties of broadened phase transitions can be transferred not only to ferroelectrics but also to other materials.

This work was performed as part of Russian Fund for Fundamental Research Project 96-02-17318-a.

[1] Ch. E. Guillaume, Compt. Rendue Acad. Sci., Paris **125**, 235 (1897).
[2] E. F. Wasserman, in *Ferromagnetic Materials*, edited by K. H. J. Buschov and E. P. Wohlfarth, Elsevier, Amsterdam, 1990, Vol. 6, p. 240.
[3] P. Chevenard, Compt. Rendue Acad. Sci., Paris **122**, 594 (1921).
[4] L. D. Landau and E. M. Lifshitz, *Statistical Physics*, Pergamon Press, N.Y., part 1 (1980) [Russian original, Nauka, Moscow, 1976].
[5] M. Matsui and S. Shikazumi, J. Phys. Soc. Jpn. **45**, 458 (1978).
[6] K. P. Belov, *Magnetic Transitions*, Consultants Bureau, N.Y., 1962 [Russian original, Gostekhizdat, Moscow, 1959].
[7] E. P. Wohlfarth, J. Phys. **32**, 1 (1971).
[8] G. Hausch and H. Warlimont, Acta Metall. **21**, 401 (1973).
[9] S. Shtrikman and E. P. Wohlfarth, Physica **60**, 427 (1972).
[10] D. M. Edwards, J. Mathon, and E. P. Wohlfarth, J. Phys. F **5**, 1619 (1975).
[11] D. Wagner and E. P. Wohlfarth, J. Phys. F **9**, 717 (1979).

[12] Y. Suzuki, H. Miyajima, G. Kido *et al.*, J. Phys. Soc. Jpn. **50**, 817 (1981).
[13] O. Yamada and E. du Tremolet de Lacheisserie, J. Phys. Soc. Jpn. **53**, 729 (1984).
[14] B. Rellinghaus, J. Kästner, T. Schneider, and E. F. Wasermann, Phys. Rev. B **51**, 2983 (1995).
[15] A. Yu. Romanov, V. P. Silin, and D. Wagner, Phys. Lett. A **226**, 310 (1997).
[16] A. V. Deryabin, V. K. Kazantsev, and B. N. Shvetsov, J. Magn. Magn. Mater. **51**, 98 (1985).
[17] V. P. Silin, D. Wagner, and V. M. Zverev, Phys. Lett. A **199**, 395 (1995).
[18] D. Wagner, A. Yu. Romanov, and V. P. Silin, Zh. Éksp. Teor. Fiz. **109**, 1753 (1996) [JETP **82**, 945 (1996)].
[19] A. Yu. Romanov and V. P. Silin, Fiz. Met. Metalloved. **83**, 111 (1997).
[20] W. Bendick, H. H. Ettwig, and W. Pepperhoff, J. Phys. F **8**, 2525 (1978).
[21] E. F. Wasserman, Invited Talk, TMS Meeting, Cincinnati, USA (1996).
[22] M. Shiga, K. Makita, K. Uematsu *et al.*, J. Phys.: Condens. Matter **2**, 1239 (1990).
[23] T. Moriya and K. Usami, Solid State Comm. **34**, 95 (1980).
[24] B. A. Strukov, Sorosovskiĭ obrazovatel'nyĭ zhurnal, No. 12, 95 (1996).

Translated by M. E. Alferieff

Kinetics of formation and growth of antiphase domains during second-order phase transitions

L. I. Stefanovich and É. P. Fel'dman

Donetsk Physicotechnical Institute, Ukrainian National Academy of Sciences, 340114 Donetsk, Ukraine
(Submitted 22 October 1996)
Zh. Éksp. Teor. Fiz. **113**, 228–239 (January 1998)

The kinetics of the formation and growth of antiphase domains during second-order phase transitions is investigated theoretically in the Ginzburg–Landau model using a statistical approach. It is shown that depending on the initial conditions both uniform and polydomain-type ordering can be realized in thermodynamic equilibrium. It is found that for small initial inhomogeneities the ordering process can be nonmonotonic in its initial stages. It is established that for special initial conditions long-lived ordered structures of a special type arise in the intermediate stages of the ordering process. The characteristic formation time of the domain structure is estimated. © *1998 American Institute of Physics.* [S1063-7761(98)01601-1]

1. INTRODUCTION

In the study of systems which are far from thermodynamic equilibrium, it is most important to determine the paths of the transition to equilibrium.

In Refs. 1 and 2, the evolution of strongly nonequilibrium systems is studied through analysis of kinetic equations derived from a microscopic treatment. These equations are used either to study phase separation in alloys[1] or phase separation and ordering occurring simultaneously.[2]

In Refs. 3 and 4 we developed a statistical approach to the analysis of the dynamics of nonequilibrium systems. This approach made it possible to elucidate the flow and sequence of stages of phase separation in alloys and glasses. Our objective in the present paper is to apply the statistical approach to determine the paths of the transition from disorder to order in systems undergoing a second-order phase transition.

One or several quantities, called order parameters, are ordinarily introduced as a quantitative characteristic of the change in state of a body as it passes through the phase-transition point T_c. The physical meaning of the order parameter is not specified in a phenomenological theory. The system under study can be an alloy undergoing ordering, a ferroelectric, an antiferromagnet, etc. If cooled rapidly enought from a temperature above T_c to a temperature below T_c such systems will evolve under isothermal conditions from a disordered to an ordered state. As a rule, the ordered state cannot be characterized uniquely—it can be either homogeneous ordering, when the order parameter η is identical throughout the entire system (single crystal or crystallite, if the sample is polycrystalline), or inhomogeneous ordering, when the system separates into domains in each of which η assumes some value from a set of equivalent values.

The thermodynamics of ordering has been studied in detail. We shall attempt to follow the kinetics of restructuring of an initially disordered system with two equivalent states of equilibrium. For definiteness, we shall employ the terminology referring to ordering of a binary alloy AB with two equivalent sublattices. The alloy can be characterized by a single scalar parameter of long-range order. This alloy possesses two states of equilibrium which differ by the sign of the order parameter. Such a model can be used to describe the kinetics of ordering of the simplest structures encountered in a number of ordered alloys, for example, in alloys such as β-brass. Ordering of atoms in systems of this type is characterized by how completely the sites of different types (forming sublattices) are filled with atoms of different types.

In the present paper, we attempt to determine in detail how ordering (for $T<T_c$) of an initially disordered system develops. Will homogeneous ordering occur over the entire sample (crystallite) or will the sample decompose into domains separated by antiphase boundaries? A preliminary report on this subject is given in Ref. 5.

We emphasize that the initial state of a nonequilibrium system plays an important role in determining how the order parameter evolves.

Of necessity, the initial state of the system must be prescribed statistically, since, first, inhomogeneities of the order parameter which are randomly distributed in space are continuously formed in the sample when it is cooled rapidly and, second, thermal fluctuations of the order parameter are always present. The spatial scale of the corresponding inhomogeneities is much larger than the lattice parameter. These inhomogeneities will evolve completely differently depending on the initial state of the nonequilibrium system, specifically, either along the homogeneous ordering path or through formation of a well developed polydomain structure followed by growth of the domains. Kinetically retarded intermediate polydomain structures can also appear for certain initial conditions. We shall assume that despite the closeness of the temperature to the critical temperature $((T_c-T)/T_c \ll 1)$ there is still a possibility of going beyond the fluctuation region and therefore ordering of the alloy can be described within the Landau theory.

2. FORMULATION OF THE PROBLEM

Let us consider a substitution-type binary alloy consisting of N_A atoms of type A and N_B atoms of type B, where the crystal lattice, containing N sites, can be divided into two equivalent sublattices. In this case, the relative concentra-

tions of sites of each type are both equal to 1/2. Then the degree η of long-range order can be determined by the expression

$$\eta = 2p_{\mathrm{A}}^{(1)} - 1, \tag{1}$$

where $p_{\mathrm{A}}^{(1)} = N_{\mathrm{A}}^{(1)}/N^{(1)}$ is the probability that sites of the first type are occupied by atoms A. The (long-range) order parameter η which we have introduced is proportional to the deviation of the probability $p_{\mathrm{A}}^{(1)}$ from its value 1/2 in the disordered alloy. For this reason, in the disordered alloy we have $\eta = 0$, and in ordered alloys this quantity will assume all higher values the closer the state of the crystal is to an ideal ordered state. In the limit of complete ordering the order parameter η, as follows from Eq. (1), assumes the values ± 1. It is completely obvious that in accordance with the symmetry of the initial model of the alloy undergoing ordering there are two equivalent possibilities, corresponding at $T = 0$ to the values of the order parameter ± 1, for predominant filling of sublattices with atoms A and B. However, the equilibrium values of the order parameter cannot be established in the entire macroscopic volume of the crystal (or even a separate crystallite) over macroscopic times (on the order of the time over which two neighboring atoms change places). Over these times relaxation of the short-range order over the entire volume of the crystal mainly occurs. The kinetics of this stage is quite difficult to describe, and we shall not dwell on it in detail. From the macroscopic point of view, however, prescribing the initial state consists of indicating the spatial distribution of the long-range order parameter, in other words, describing the field of the order parameter at some moment in time, taken as the initial moment. Ordinarily, after relaxation of short-range order, the alloy consists of intertwined fragments. Neighboring fragments differing by the sign of the order parameter are called antiphase domains, and the transitional regions between domains are called, correspondingly, antiphase boundaries. We note, incidentally, that grain boundaries in polycrystals are also antiphase domains, since neighboring grains differ only by spatial orientation, and the energy of a crystallite is degenerate (with infinite multiplicity) with respect to the orientational parameters (Euler angles).

To construct a theoretical description of the subsequent stages of ordering of an alloy, we shall represent the nonequilibrium correction to the thermodynamic potential in the form of the Ginzburg–Landau functional

$$\Phi = \int \varphi dV = \int \left[f(\eta) + \frac{1}{2}\,\delta(\nabla\eta)^2 \right] dV. \tag{2}$$

where $f(\eta)$ is the free-energy density in homogeneously ordered regions of the solid solution (alloy); $\delta(\nabla\eta)^2$ is the first nonvanishing term in the expansion of f in a series in the gradients of the order parameter and describes the contribution of antiphase boundaries to the free energy of the system. Here the quantity δ is of order Ur_0^2, where U is the mixing energy of the solid solution[6] and r_0 is the characteristic interaction radius of the atoms of the solid solution.

In the alloys studied, the order–disorder transition is a second-order phase transition[6] and near the ordering temperature T_c it can be described in the following phenomenological model:

$$f(\eta) = T_c\left(-\frac{\alpha}{2}\,\eta^2 + \frac{1}{4}\,\eta^4 \right), \tag{3}$$

where

$$\alpha \equiv (T_c - T)/T_c \tag{4}$$

is a dimensionless parameter characterizing the closeness of the temperature T to which the sample is cooled to the ordering temperature T_c. The model (3) explicitly takes into account the degeneracy of the energy with respect to the sign of the order parameter. We shall describe the dynamics of such a system under isothermal conditions at temperature T by means of the Landau–Khalatnikov equation[7] for the nonconserved order parameter,

$$\frac{\partial\eta(\mathbf{r},t)}{\partial t} = -\gamma\,\frac{\delta\Phi\{\eta(\mathbf{r},t)\}}{\delta\eta(\mathbf{r},t)}, \tag{5}$$

where the right-hand side contains the functional derivative of the functional (2) with respect to the order parameter, the coefficient of proportionality being the kinetic coefficient γ, which is proportional to the probability of two neighboring atoms in the alloy changing places.

Substituting the expression (2) into Eq. (5) and taking account of Eq. (3) gives

$$\frac{\partial\eta}{\partial t} = \gamma T_c r_0^2 \Delta\eta + \gamma T_c \alpha\eta - \gamma T_c \eta^3, \tag{6}$$

where we have used the fact that U can be replaced by T_c on account of the redefinition of r_0. Equation (6) is a nonlinear differential equation of the diffusion type. Note that the kinetic coefficient γ can be represented in the form $(t_r T_c)^{-1}$, where t_r is the characteristic time of an elementary restructuring event (for example, a displacement of a atom or neighboring atoms changing places). According to Ref. 6, this time is given by

$$t_r \sim \Omega_D^{-1}\,\exp(w/T), \tag{7}$$

where Ω_D is the Debye frequency (ordinarily $\sim 10^{12}\ \mathrm{s}^{-1}$), w is the activation energy, and T is the temperature of the alloy.

If r_0 and t_r are chosen as the characteristic scales of length and time, respectively, then Eq. (6) can be rewritten in the form

$$\partial\eta/\partial t = \Delta\eta + \alpha\eta - \eta^3. \tag{8}$$

Here Δ is the Laplacian. Note that in the simplest case, when the order η is spatially uniform and we have $\alpha > 0$, two equivalent stable solutions of the equation exist in the limit of long times ($t \to \infty$):

$$\eta(\mathbf{r},t) = \eta_e = \pm\sqrt{\alpha}. \tag{9}$$

An equation of the form (8) was obtained earlier by Allen and Cahn[8] (see also Ref. 9) and used to study the motion of solitary antiphase boundaries.

The initial condition for Eq. (8) consists of prescribing the function $\eta(\mathbf{r},t)$ at $t = 0$. Since, as underscored earlier,

this initial function is random, the order parameter is a random function of the coordinates at $t \neq 0$ also, i.e., Eq. (8) will describe the space–time evolution of the random order-parameter field $\eta(\mathbf{r},t)$.

3. BASIC COMPUTATIONAL RELATIONS

It is not necessary to know the detailed distribution of the order parameter η over the entire macroscopic sample in order to construct a description of the kinetics of ordering of alloys. For this reason, our problem will be to find the main physically significant characteristics of this function. They include the mean value $\langle \eta(r,t)\rangle \equiv \overline{\eta}(t)$ of the order parameter and the correlation function

$$\langle [\eta(\mathbf{r},t)-\overline{\eta}][\eta(\mathbf{r}',t)-\overline{\eta}]\rangle \equiv \langle \xi(\mathbf{r},t)\xi(\mathbf{r}',t)\rangle \equiv K(\mathbf{s},t), \tag{10}$$

where $\mathbf{s}=\mathbf{r}-\mathbf{r}'$, i.e., we employ the assumption that the order-parameter field is statistically homogeneous.

The equations for $\eta(t)$ and $K(\mathbf{s},t)$ can be derived from the basic equation (8) both by averaging Eq. (8) itself and by averaging after premultiplying both sides of the equation by $\eta(r,t)$.

To obtain a closed system of equations for the functions $\overline{\eta}(t)$ and $K(\mathbf{s},t)$, i.e., one not containing higher-order moments, we assume that

$$\langle \xi^2(\mathbf{r},t)\xi(\mathbf{r}',t)\rangle = 0 \tag{11}$$

for all $\mathbf{r}$ and $\mathbf{r}'$, and we use for the fourth-order correlation function $\langle \xi^3(\mathbf{r},t)\xi(\mathbf{r}',t)\rangle$ a distribution of the form

$$\langle \xi^3(\mathbf{r},t)\xi(\mathbf{r}',t)\rangle = \langle \xi^2(t)\rangle \langle \xi(\mathbf{r},t)\xi(\mathbf{r}',t)\rangle \equiv K(0,t)K(\mathbf{s},t). \tag{12}$$

One justification for this procedure is that there is only one spatial scale in this problem. For this reason, it is natural to assume that the functional dependence of the correlation function $\langle \xi^3(\mathbf{r},t)\xi(\mathbf{r}',t)\rangle$ on $|\mathbf{r}'-\mathbf{r}|$ is identical to the functional dependence of the correlation function $\langle \xi(\mathbf{r})\xi(\mathbf{r}')\rangle \equiv K(|\mathbf{r}-\mathbf{r}'|)$ on the distance between the points $\mathbf{r}$ and $\mathbf{r}'$. We can refer also to the model two-point distribution function proposed in Ref. 10 for the random order-parameter field

$$P(\xi_1(\mathbf{r}),\xi_2(\mathbf{r}')) = \rho(\xi_1)(\xi_2)\left[1+\frac{K(\mathbf{r}-\mathbf{r}')}{\langle \xi^2\rangle^2}\,\xi_1\xi_2\right],$$

where $\rho(\xi)$ is a one-point distribution function. It can be verified by direct integration that for this model the correlation function $\langle \xi_1^3\xi_2\rangle$ is proportional to $\langle \xi_1\xi_2\rangle$. It is known that for a Gaussian random field ξ the distribution (12) is exact. However, the coefficient of proportionality between $\langle \xi^4\rangle$ and $\langle \xi^2\rangle^2$ is not 1, as in Eq. (12), but 3. This choice of the coefficient arises because in the present problem, especially at long times, the one-point distribution function is very far from Gaussian. From physical considerations, based on the equivalence of states with order parameters having the same magnitude but different signs, it follows that the one-point distribution function at sufficiently long times ($t \gg \alpha^{-1}$) is very close to a curve with two sharp maxima at $\xi = \pm \xi_e$. For such distribution functions, a direct calculation shows that the coefficient in question is close to 1.

As a consequence of Eq. (8) and the assumptions (11) and (12), we obtain ultimately the system of equations

$$\frac{d\overline{\eta}}{dt} = \frac{1}{2}\,[\alpha\overline{\eta} - 3K(0,t)\overline{\eta} - \overline{\eta}^3], \tag{13}$$

$$\frac{\partial K(\mathbf{s},t)}{\partial t} = \Delta K(\mathbf{s},t) + [\alpha - K(0,t) - 3\overline{\eta}^2]K(\mathbf{s},t). \tag{14}$$

This is a system of nonlinear differential equations (one of which is a partial differential equation) and cannot be solved analytically. But, on account of its degeneracy it can be reduced to a system of ordinary differential equations by Fourier transforming Eq. (14) with respect to the variable $\mathbf{s}$:

$$\frac{d\overline{\eta}}{dt} = \frac{1}{2}\,[\alpha\overline{\eta} - 3K(0,t)\overline{\eta} - \overline{\eta}^3], \tag{15}$$

$$\frac{d\widetilde{K}((\mathbf{q},t)}{dt} = [-q^2 + \alpha - K(0,t) - 3\overline{\eta}^2]\widetilde{K}(\mathbf{q},t), \tag{16}$$

where $\widetilde{K}(\mathbf{q},t)$ is the Fourier transform of the correlation function $K(\mathbf{s},t)$.

The system of equations (15) and (16) for the functions $\overline{\eta} = \overline{\eta}(t)$ and $\widetilde{K} = \widetilde{K}(\mathbf{q},t)$ in turn can be reduced by simple manipulations to a system of nonlinear ordinary differential equations for the mean value of the order parameter $\overline{\eta} = \overline{\eta}(t)$ and its variance $D = D(t)$:

$$\frac{d\overline{\eta}}{dt} = \frac{1}{2}\,(\alpha - 3D - \overline{\eta}^2)\,\overline{\eta}, \tag{17}$$

$$\frac{dD}{dt} = [\alpha_{\text{eff}}(t) - D - 3\overline{\eta}^2]D, \tag{18}$$

where we have introduced $D = D(t) \equiv K(0,t)$ and

$$\alpha_{\text{eff}}(t) \equiv \alpha - 1/r_c^2(t). \tag{19}$$

Here we have employed the natural definition of the correlation length,

$$\frac{1}{r_c^2(t)} \equiv \frac{\int q^2 \widetilde{K}(\mathbf{q},t)\,d\mathbf{q}}{\int \widetilde{K}(\mathbf{q},t)\,d\mathbf{q}}. \tag{20}$$

Equation (16) is constructed so that the correlation length at an arbitrary time is determined by the Fourier transform of the correlation function at $t = 0$:

$$\frac{1}{r_c^2(t)} \equiv \frac{\int q^2 \widetilde{K}(\mathbf{q},0)\exp(-q^2 t)\,d\mathbf{q}}{\int \widetilde{K}(q,0)\exp(-q^2 t)\,d\mathbf{q}}. \tag{21}$$

Like the function $K(\mathbf{s},0)$, the function $\widetilde{K}(\mathbf{q},0)$ must undoubtedly be prescribed. According to Eq. (19), the function $\alpha_{\text{eff}}(t)$ is determined completely by the character of the evolution of the correlation length $r_c(t)$ of the system undergoing ordering (in our case the correlation length is associated with the characteristic spatial scale of the ordered regions).

An asymptotic analysis of the expression (21) for nonpathological initial correlation functions gives a reliable interpolation formula

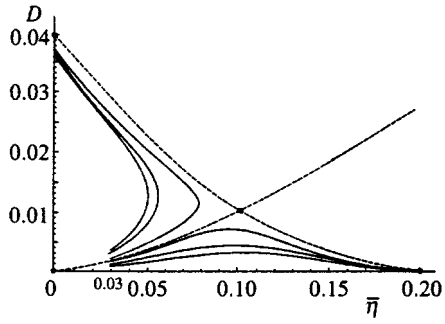


FIG. 1. Phase portrait of an alloy undergoing ordering: broken curves—separatrices; filled dots—singular points of the asymptotic system of equations (24). The results of numerical integration of the complete system of equations (23) for $\alpha=0.04$ and $r_c(0)=10$ and the following initial conditions are also presented here: $\eta_0=0.03$; $D_0=0.0007$, 0.001, 0.0015, 0.002, 0.003, 0.0035.

$$r_c(t)=\sqrt{r_c^2(0)+2t/3}, \tag{22}$$

where $r_c(0)$ is the initial correlation length of the system ($t=0$). The last formula confirms the well-known[11] result that the domains grow in size according to a diffusion law as $\sqrt{t}$ (if, of course, $t\gg r_c^2(0)$). Thus, by virtue of Eqs. (19) and (22), the system of equations (17) and (18) for $\overline{\eta}(t)$ and $D(t)$ assumes the form

$$\frac{d\overline{\eta}}{dt}=\frac{1}{2}\,\overline{\eta}[\alpha-3D(t)-\overline{\eta}^2],$$

$$\frac{dD(t)}{dt}=\left[-\left(\frac{2}{3}\,t+r_c^2(0)\right)^{-1}+\alpha-D(t)-3\overline{\eta}^2\right]D(t). \tag{23}$$

4. LONG-TIME ASYMPTOTIC BEHAVIOR OF THE SYSTEM

It is of fundamental interest to investigate the system of equations (23) near the ordering temperature T_c, when $\alpha\ll 1$ holds and the initial correlation length is not too large so that we have $r_c^2(0)\ll 1/\alpha\ll d^2$, where d is the characteristic size of a crystallite (recall that we are measuring time in units of t_r and distance in units of the interatomic interaction length r_0). Then, asymptotically, at times $t\gg 1/\alpha$ the system of equations (23) goes over to the system of equations with constant coefficients

$$\frac{d\overline{\eta}}{dt}=\frac{1}{2}\,\overline{\eta}[\alpha-3D(t)-\overline{\eta}^2],$$

$$\frac{dD(t)}{dt}=[\alpha-D(t)-3\overline{\eta}^2]D(t) \tag{24}$$

with the initial conditions

$$\overline{\eta}(0)=\eta_0,\quad D(0)=D_0. \tag{25}$$

The solution of the system (24) with initial conditions (25) makes it possible to obtain information about the last stage of the ordering process. But even this simplified system cannot be solved analytically. However, it can be analyzed qualitatively on the basis of the concept of the phase portrait of the system[12] (in this case, in the variables $\overline{\eta}$ and D; see Fig. 1).

For $\alpha<0$, i.e., at a temperature above the critical temperature, there is only one singular point. Its coordinates are $\overline{\eta}=0$, $D=0$ and it is a stable point. This signifies, as one would expect, that the system passes into a disordered state irrespective of the initial conditions.

For $\alpha>0$, i.e., $T<T_c$, there exists an entire system of singular points. However, only the points located in the upper half-plane of the $(\overline{\eta},D)$ phase portrait are physically meaningful. Moreover, we shall analyze only four singular points of the system (24) located in the first quadrant (since the second quadrant is added by symmetry).

The first singular point ($\overline{\eta}=0$, $D=0$), corresponding to a homogeneous disordered state, is an unstable point (Fig. 1).

The second singular point ($\overline{\eta}=\eta_e=\sqrt{\alpha}, D=0$) corresponds to a homogeneous ordered state and is a stable point (Fig. 1).

The third singular point ($\overline{\eta}=0$, $D=\alpha$) corresponds to nonuniform ordering and is likewise a stable point (Fig. 1). This point corresponds to the possibility of realizing a multidomain structure. The condition $\overline{\eta}=0$ means that domains of different signs are equally likely to exist. Since the variance of the order parameter equals $\alpha=\eta_e^2$, the volume of the transitional regions (antiphase boundaries) is negligibly small compared with the total volume of the domains within each of which the order parameter equals $+\eta_e$ or $-\eta_e$.

Finally, there is a fourth singular point, a saddle point, with the coordinates $\overline{\eta}=\sqrt{\alpha}/2$, $D=\alpha/4$ (Fig. 1). A pair of separatrices with slope angles having tangents (near the singular point) equal to $\sqrt{\alpha}$ or $-\sqrt{\alpha}$, respectively, passes through the saddlepoint (Fig. 1). The separatrix emanating from the origin of the coordinates and passing through this saddlepoint separates the phase portrait into two parts. The left-hand upper part is a region of attraction of the inhomogeneous (polydomain) state and the right-hand lower part is a homogeneous single-domain state. Depending on the initial conditions (η_0,D_0), the phase trajectories of the system will lie in one of the sectors indicated above. This is illustrated in the phase portrait (Fig. 1), where, besides the analytically determined singular points and separatrices of the asymptotic system of equations (24), a series of results obtained by integrating the complete system of equations (23) numerically is presented (in the form of phase trajectories).

If the mean value of the order parameter $\eta_0\neq 0$ in the initial state and exceeds the average size $\sqrt{D_0}$ of the inhomogeneities, then the system will evolve directly into a single-domain state. The sign of the order parameter in the state of thermodynamic equilibrium will be the same as the sign of $\overline{\eta}$ in the initial state. A deviation of the system in one direction or the other with respect to the sign of the order parameter in the initial state can be caused, generally speaking, by different factors, both random and deterministic. Specifically, this can be the action of a field coupled to the order parameter, the presence of definite temperature gradients, the influence of grain boundaries, stacking faults, and so on.

However, if in the initial state the inhomogeneities are quite developed and the mean value of the order parameter equals zero or is small ($\sqrt{D_0}>\eta_0$), then a well developed polydomain structure forms in the system over a time

$t \sim \alpha^{-1}$. The characteristic size of the domains, like the characteristic thickness of the transitional layer (antiphase domain wall) between them, will become $\sim \alpha^{-1/2}$ by this time, and on the whole over the sample the volume fractions of domains with different signs will be the same. Subsequently, according to Eq. (22), the domains will grow in size according to the diffusion law $\propto \sqrt{t}$, while the thicknesses of the antiphase boundaries will remain unchanged at the level $\alpha^{-1/2}$. In the absence of long-range action the multidomain state is not thermodynamically stable. This state can be regarded as being long-lived, and its characteristic lifetime satisfies $\alpha^{-1} \ll t \ll d^2$. In the situation under study, the system does not pass directly into a thermodynamically stable single-domain state but rather through a stage of domain formation and growth. Domains continue to grow until their sizes reach magnitudes on the order of the dimensions of the crystallite, when by virtue of the boundary conditions the system will give preference to a domain with a definite sign. We note that the process of formation of a domain structure is completed by the time $t \sim d^2$.

5. ANALYSIS OF THE COMPLETE EVOLUTION EQUATIONS

As a result of an investigation of the asymptotic system (24) instead of the system (23), it was found that the initial and intermediate stages of ordering fall outside the framework of our analysis. To investigate the entire process of evolution of a system undergoing ordering, it is necessary to solve the complete system of equations (23), where one of the coefficients is explicitly time-dependent. This does not change the results of the qualitative analysis of the asymptotic behavior of the system. Only the quantitative estimates of the relaxation time and characteristic spatial scales change. The analytical investigation of the system (23) becomes problematic at short and intermediate times.

To trace all stages of evolution of a system undergoing ordering, we solved the complete system of equations (23) numerically. Numerical analysis was performed for different temperatures T and for different initial sizes of the ordered regions (i.e., the parameters α and $r_c(0)$ were varied). Moreover, the initial conditions of the problem (η_0 and D_0) were varied. It was found that for sufficiently small α and values of $r_c(0)$ close to 1 the ordering process can be nonmonotonic—at first the variance $D(t)$ of the order parameter decreases, i.e., disordering occurs, and only after this does normal formation and growth of domains occur (Fig. 2).

The second interesting feature found explicitly as a result of numerical integration of the complete system of equations (23) is that quasistationary intermediate structures of a special type with unequal volume fractions of the corresponding ordered phases can form during the ordering process. A similar phenomenon is described in Ref. 13, where numerical analysis of two-dimensional models showed a virtual phase at intermediate stages of separation of intermetallides.

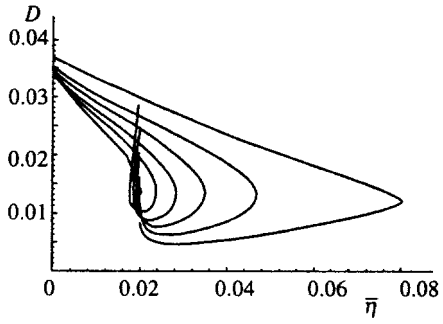


FIG. 2. Phase portrait of an alloy undergoing ordering for $\alpha = 0.04$, $r_c(0) = 10$ and the initial conditions $\eta_0 = 0.02$, $D_0 = 0.008$, 0.012, 0.016, 0.02, 0.024, 0.028.

In our case the quasistationary states corresponding to intermediate polydomain structures with unequal volumes of ordered phases occur in a very narrow range of initial conditions and are manifested in the form of horizontal sections (plateaus or steps) on the evolution curves for the mean value $\bar{\eta}(t)$ of the order parameter (Fig. 3a) and the variance $D(t)$ (Fig. 3b).

As the results of numerical analysis show, the metastable intermediate structures (with the appropriate initial conditions (η_0, D_0)) are quite stable formations. Before passing into a thermodynamically equilibrium homogeneous state or very long-lived polydomain state a system undergoing ordering spends an appreciable time in such a kinetically retarded state.

Parametric plots show the ordering process especially clearly. We present in Fig. 4 a phase portrait of the system (23) for the same values of the parameters and initial conditions as in Fig. 3. It is clearly seen that there are two groups of phase trajectories, corresponding to homogeneous and inhomogeneous orderings. Two of these trajectories, which correspond to $D_0 = 0.0018234$ and 0.0018235, are very close to asymptotic separatrices and pass near the saddle singular point. On the segment from the initial values (η_0, D_0) up to the saddle point they practically merge with one another (because of the poor resolution on the numerical plot, Fig. 4);

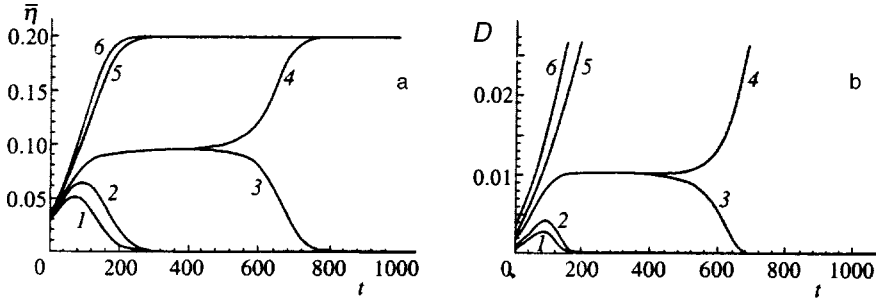


FIG. 3. Results of numerical integration of the system (23) for $\alpha = 0.04$, $r_c(0) = 10$ and the following initial conditions: $\eta_0 = 0.03$, $D_0 = 0.007$ (*1*), 0.001 (*2*), 0.0018234 (*3*), 0.0018235 (*4*), 0.0025 (*5*), 0.0035 (*6*); a—evolution of the mean value $\bar{\eta}(t)$ of the order parameter; b—evolution of the variance $D(t)$.

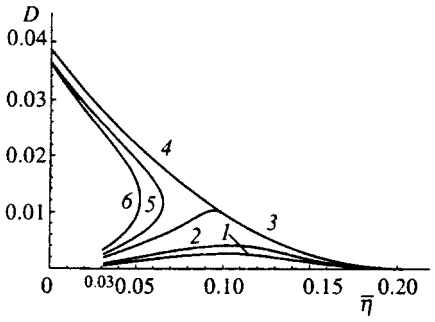


FIG. 4. Phase portrait of the system (23) for the same values of the parameters and initial conditions as in Fig. 3.

after the singular point they move apart toward singular points corresponding to homogeneous and inhomogeneous orderings. In reality, these phase trajectories lie on different sides of the separatrix emanating from the coordinate origin (0, 0) and entering the saddle point. The process whereby the phase trajectories pass near the saddle lasts for a quite long time. We interpret this as a kinetic retardation of the ordering process.

Analysis of the system of equations (23) makes it possible to estimate the characteristic formation time τ of the domain structure. If the time t_r of an elementary restructuring event in the system is determined by the expression (7), then for the time indicated above we have

$$\tau = \frac{t_r}{\alpha} = \frac{1}{\Omega_D} \frac{T_c}{T_c - T} \exp\left(\frac{w}{T}\right) \tag{26}$$

i.e., τ is a nonmonotonic function of temperature. This nonmonotonicity has been observed, specifically, in experimental investigations of the temperature dependence of the relaxation time in the initial stage of ordering of the alloy $AuCu_3$.[14]

The transition of the alloy into an ordered state is associated with redistribution of atoms over the lattice sites (the exponential factor in Eq. (26)). These processes are characterized by definite activation energies w, and they are extremely slow at low temperatures. As the isothermal holding temperature increases, the ordering time at first decreases rapidly. As this temperature approaches the ordering temperature T_c, for thermodynamic reasons another factor becomes important in connection with the fact that the ordering process slows down as the difference of the free energies of the ordered and disordered phases corresponding to a given temperature decreases. At temperatures close to the ordering temperature the pre-exponential factor in Eq. (26) becomes substantial, and for this reason ordering occurs slowly, i.e., the ordering time once again reaches large values. If the temperature is assumed to be much less than the height of the potential barrier ($T \ll w$), then w can be estimated from the following equation according to the position of the temperature minimum T_m of the ordering time:

$$w = T_c^2/(T_c - T). \tag{27}$$

It should be noted that T_m has been determined experimentally for a number of alloys (see, for example, Ref. 6).

6. CONCLUSIONS

The questions investigated in this paper pertain to the kinetics of phase transformations in systems undergoing ordering, where a second-order phase transition occurs. We were not interested in the establishment of short-range order. The systems undergoing ordering were studied at mesoscopic and macroscopic times and it was assumed that all local rearrangement processes had already occurred. The subsequent relaxation of the system is associated with the appearance of long-range order, when the appearance of superstructures becomes overwhelming likely. We attempted to take account of spatial inhomogeneity, which always appears even at the stage when short-range order arises, as well as the random character of the inhomogeneity.

To this end, for the theoretical description of the establishment of long-range order we employed a statistical approach which we developed earlier.[3,4] This made it possible to investigate qualitatively in a continuous approximation all stages of relaxation of long-range order in an alloy undergoing ordering.

[1] V. G. Vaks, S. V. Beiden, and V. Yu. Dobretsov, JETP Lett. **61**, 68 (1995).
[2] J. W. Cahn, A. Novick-Cohen, J. Stat. Phys. **76**, 877 (1994).
[3] É. P. Fel'dman and L. I. Stefanovich, Zh. Éksp. Teor. Fiz. **96**, 1513 (1989) [Sov. Phys. JETP **69**, 858 (1989)].
[4] É. P. Fel'dman and L. I. Stefanovich, Zh. Éksp. Teor. Fiz. **98**, 1695 (1990) [Sov. Phys. JETP **71**, 951 (1990)].
[5] E. P. Fel'dman and L. I. Stefanovich, JETP Lett. **63**, 983 (1996).
[6] M. A. Krivoglaz and A. A. Smirnov, *Theory of Order-Disorder in Alloys*, American Elsevier, N.Y., 1965 [Russian original, Fizmatgiz, Moscow, 1958].
[7] L. D. Landau and I. M. Khalatnikov, Dokl. Akad. Nauk SSSR **96**, 469 (1954).
[8] S. M. Allen and J. W. Cahn, Acta Metall. **27**, 1085 (1979).
[9] B. A. Strukov and A. P. Levanyuk, *Physical Principles of Ferroelectric Phenomena in Crystals* [in Russian], Nauka, Moscow, 1995.
[10] J. S. Langer, M. Baron, and H. D. Miller, Phys. Rev. A **11**, 1417 (1975).
[11] I. M. Lifshitz, Zh. Éksp. Teor. Fiz. **42**, 1354 (1962) [Sov. Phys. JETP **15**, 939 (1962)].
[12] A. A. Andronov, E. A. Leontovich, I. I. Gordon *et al.*, *Qualitative Theory of Second-Order Dynamical Systems* [in Russian], Nauka, Moscow, 1966.
[13] Long-Quing Chen and A. G. Khachaturyan, Phys. Rev. B **46**, 5899 (1992).
[14] N. W. Lord, J. Chem. Phys. **21**, 692 (1953).

Translated by M. E. Alferieff

Duality in two-dimensional $Z(N)$-symmetric spin models on a finite lattice

A. I. Bugriĭ*) and V. N. Shadura

N. N. Bogolyubov Institute of Theoretical Physics, Ukrainian National Academy of Sciences, 252143 Kiev, Ukraine

(Submitted 10 February 1997)
Zh. Éksp. Teor. Fiz. **113**, 240–260 (January 1998)

We propose a method for deriving duality relations for two-dimensional inhomogeneous $Z(N)$-symmetric models on a finite square lattice wound around a torus. The method is used to obtain duality relations for the vector Potts model, the Berezinskiĭ–Villain $Z(N)$-model, the Ashkin–Teller model, and the 8-vertex model on a lattice obliquely wound around a torus, as well as an exact relation linking the partition functions of the latter two models. © *1998 American Institute of Physics.* [S1063-7761(98)01701-6]

1. INTRODUCTION

Studies of duality in models of statistical mechanics and field theory serve as an important method for a nonperturbative investigation of the phase diagrams of the models. The first to discover the duality transformation for the two-dimensional Ising model were Kramers and Wannier.[1] Kadanoff and Ceva[2] generalized the duality relation to the inhomogeneous Ising model, in which the coupling constants along the horizontal and vertical bonds, $K_x(r)$ and $K_y(r)$, are arbitrary functions of lattice site coordinates:

$$\left\{\prod_{\tilde r,\mu}\ \sinh 2\tilde K_\mu(\tilde r)\right\}^{-1/4}\tilde Z[\tilde K] = \left\{\prod_{r,\mu}\ \sinh 2K_\mu(r)\right\}^{-1/4} Z[K], \tag{1}$$

$$\sinh 2K_x(r)\sinh 2\tilde K_{-y}(\tilde r)=1,$$

$$\sinh 2K_y(r)\sinh 2\tilde K_{-x}(\tilde r)=1. \tag{2}$$

Here the subscript μ takes the values x and y, a tilde indicates that the quantity refers to the dual lattice, and $\tilde K_{-\mu}(\tilde r)=\tilde K_\mu(\tilde r-\hat\mu)$ (here $\hat\mu=\hat x,\hat y$ are unit vectors along the x and y axes).

As noted by Kramers and Wannier[1] and Kadanoff and Ceva,[2] for a homogeneous model this relation becomes exact in the thermodynamic limit, i.e., for the specific free energy. However, in the inhomogeneous case the very procedure of passing to the thermodynamic limit is ambiguous. In Ref. 2 the duality relation (1) was derived for spherical boundary conditions, which for lattice models are nonphysical. The complexity of deriving the duality relation for a lattice wound around a torus, say by comparing the high- and low-temperature expansions for the partition function, is due to the need to allow for closed graphs that encompass the torus several times.

In Refs. 3 and 4 we proposed an exact duality relation for an inhomogeneous two-dimensional Ising model on a square lattice R of finite dimensions $n\times m$ wound around a torus (in Fig. 1 the dimensions of the lattices R and $\tilde R$ are 3×3 and the spins at the opposite edges of the lattice must be assumed to be identical):

$$\left\{\prod_{\tilde r,\mu}\ \sinh 2\tilde K_\mu(\tilde r)\right\}^{-1/4}\tilde Z^{(\tilde p_x,\tilde p_y)}[\tilde K] = \frac{1}{2}\left\{\prod_{r,\mu}\ \sinh 2K_\mu(r)\right\}^{-1/4}\sum_{p_x,p_y=0}^{1} T^{\tilde p_x,\tilde p_y}_{p_x,p_y} Z^{(p_x,p_y)}[K]. \tag{3}$$

Here the $Z^{(p_x,p_y)}[K]$ are the partition functions of the Ising model with appropriate combinations of periodic (index $p_x,p_y=0$) and antiperiodic (index $p_x,p_y=1$) boundary conditions along the x and y axes:

$$Z^{(p_x,p_y)}[K]=\sum_{[\sigma]}\ \exp\left\{\sum_{r,\nu} K_\nu(r)\sigma(r)\nabla_\nu^{(p_\nu)}\sigma(r)\right\}, \tag{4}$$

$$\hat T=\begin{pmatrix} 1 & 1 & 1 & 1\\ 1 & 1 & -1 & -1\\ 1 & -1 & 1 & -1\\ 1 & -1 & -1 & 1\end{pmatrix}. \tag{5}$$

The Ising spin takes two values: $\sigma(r)=\pm 1$; the coordinates of the sites of the dual lattice $\tilde R$ coincide with the coordinates of the centers of the plaquettes of the original lattice R and vice versa (see Fig. 1): $\tilde r=r+(\hat x+\hat y)/2$ and $r=(x,y)$, where $x=1,\ldots,n$ and $y=1,\ldots,m$. The shift operators ∇_x and ∇_y act on $\sigma(r)$ in the following way:

$$\nabla_x\sigma(r)=\sigma(r+\hat x),\quad \nabla_y\sigma(r)=\sigma(r+\hat y).$$

They also satisfy the boundary conditions

$$\nabla_x^{(p_x)}\sigma(n,y)=(-)^{p_x}\sigma(1,y),$$

$$\nabla_y^{(p_y)}\sigma(x,m)=(-)^{p_y}\sigma(x,1). \tag{6}$$

In Ref. 3 we used the exact solution of the two-dimensional Ising model to derive the duality relation (3) for the homogeneous model, while for the inhomogeneous case this relation was proved to be valid by perturbative techniques and was checked by direct calculations involving small lattices. In Ref. 4 we obtained duality relations for an Ising model with a magnetic field applied at the lattice edges.

In this paper we formulate a systematic method for deriving the duality relations for $Z(N)$-symmetric spin models

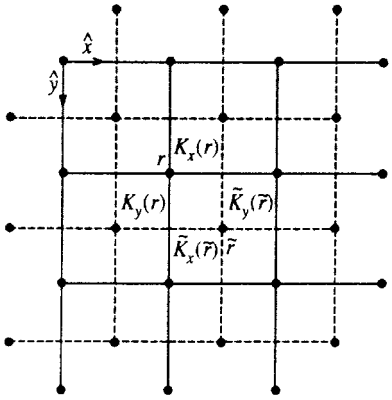


FIG. 1. Correspondence between the coordinates and coupling constants in mutually dual lattices R (solid lines) and $\tilde{R}$ (dashed lines).

on a square lattice wound around a torus. In principle, the method can easily be generalized to lattices of large dimensions compactified on a hypertorus and to lattice models with continuous global or gauge symmetry. By way of example, we derive the duality relations for the vector Potts model, the Berezinskiĭ–Villain $Z(N)$-model (a $Z(N)$-symmetric Gaussian model),[5,6] the Ashkin–Teller model, and the 8-vertex model on a lattice obliquely wound around a torus, as well as an exact relation linking the partition functions of the last two models. The duality relations for these models were obtained by Zamolodchikov[7] and Dotsenko[8] (see also Refs. 9 and 10) without taking the boundary conditions into account. Here we also show that Eq. (3) is a special case of the duality relation for the vector Potts model at $N=2$.

Before explaining the idea of the method, we introduce the concepts of a magnetic dislocation, ''topological charge,'' and the gauge transformation of configurations of coupling constants using the example of the vector Potts model, which possesses global discrete symmetry group Z_N. The Hamiltonian of this model can be written

$$-\beta H^{(p,q)}[K,\sigma]=\frac{1}{2}\sum_{r,\nu}\{K_\nu(r)\sigma^*(r)\nabla_\nu^{(p_\nu)}\sigma(r)+\text{c.c.}\}, \tag{7}$$

where the spin variable takes N values:

$$\sigma(r)=\exp\left[i\frac{2\pi}{N}l(r)\right], \quad l(r)=0,\ldots,N-1.$$

The indices $p_x=p$ and $p_y=q$ $(p,q=0,\ldots,N-1)$ in the Hamiltonian (7) denote cyclic boundary conditions:

$$\nabla_x^{(p)}\sigma(n,y)=\exp\left(i\frac{2\pi}{N}p\right)\sigma(1,y),$$

$$\nabla_y^{(q)}\sigma(x,m)=\exp\left(i\frac{2\pi}{N}q\right)\sigma(x,1), \tag{8}$$

and accordingly, for the variable $l(r)$ we have

$$l(n+1,y)=l(1,y)+p, \quad l(x,m+1)=l(x,1)+q. \tag{9}$$

Periodic boundary conditions correspond to $p=0$ and $q=0$.

If we use the explicit form (8) of the boundary conditions, it is convenient to represent the Hamiltonian $H^{(p,q)}[K,\sigma]$ in the form of the Hamiltonian $H_D^{(0,0)}[K,d,\sigma]$ with a magnetic dislocation $D^{(p,q)}$ corresponding to the boundary conditions (p,q) and with periodic boundary conditions for the spin variable $\sigma(r)$:

$$\begin{aligned}-\beta H^{(p,q)}[K,\sigma]&=-\beta H_D^{(0,0)}[K,d,\sigma]\\&=\frac{1}{2}\sum_{r,\nu}\left\{K_\nu(r)\exp\left[i\frac{2\pi}{N}d_\nu^{(p,q)}(r)\right]\right.\\&\quad\left.\times\sigma^*(r)\nabla_\nu^{(0)}\sigma(r)+\text{c.c.}\right\}\\&=\sum_{r,\nu}K_\nu(r)\cos\left\{\frac{2\pi}{N}[\Delta_\nu l(r)\right.\\&\quad\left.+d_\nu^{(p,q)}(r)]\right\}.\end{aligned} \tag{10}$$

Here $\Delta_\nu=\nabla_\nu^{(0)}-1$ is the difference derivative with periodic boundary conditions, and it is convenient to interpret the vector fields $K_\nu(r)$ and $d_\nu^{(p,q)}(r)$ defined at the lattice edges as the absolute value and phase of the coupling constant involving two neighboring spins. The magnetic dislocation $D^{(p,q)}$ is determined by the phase

$$d_\nu^{(p,q)}(r)=(d_x(r),d_y(r))=(p\,\delta_{B_X}(r),q\,\delta_{B_Y}(r)), \tag{11}$$

which is finite along the boundary cycles B_X and B_Y, which fix the spatial configuration of the dislocation on the torus:

$$\delta_{B_X}(r)=\sum_{r'\in B_X}\delta^2(r-r'), \quad \delta_{B_Y}(r)=\sum_{r'\in B_Y}\delta^2(r-r'), \tag{12}$$

where $\delta^2(r-r')$ is the Kronecker delta, and

$$B_X=\{(x,m),x=1,\ldots,n\},$$

$$B_Y=\{(n,y),y=1,\ldots,m\}.$$

The phase $d_\nu^{(p,q)}(r)$ can be interpreted as the density of the ''topological'' charge Q_ν of the magnetic dislocation. For instance, for the dislocation $D^{(p,q)}$ this charge is

$$Q_\nu\sum_r d_\nu^{(p,q)}(r)=(pn,qm). \tag{13}$$

We call the $D^{(p,q)}$ (with $p,q=0,\ldots,N-1$) the magnetic basis dislocations. Note that periodic boundary conditions along all torus cycles $(p=q=0)$ actually correspond to the case without magnetic dislocations. Nevertheless, for the sake of convenience we have introduced the notation $D^{(0,0)}$ for such boundary conditions.

The Hamiltonian (7) is invariant under local Z_N-gauge transformations[11]:

$$\sigma'(r)=\exp\left[i\frac{2\pi}{N}\phi(r)\right]\sigma(r),$$

$$K'_\mu(r)=\exp\left[i\frac{2\pi}{N}\phi(r)\right]K_\mu(r)\exp\left[i\frac{2\pi}{N}\phi(r+\hat{\mu})\right], \tag{14}$$

where $\phi(r)$ has periodic boundary conditions. This invariance leads to the following expression for the partition function Z_V of the vector Potts model:

$$Z_V^{(p,q)}[K]=\sum_{[\sigma]}\ \exp\{-\beta H^{(p,q)}[K,\sigma]\}$$

$$=\sum_{[\sigma]'}\ \exp\{-\beta H^{(p,q)}[K',\sigma']\}=Z_V^{(p,q)}[K'].$$

The gauge transformations of the variables $l(r)$ and $d_\mu(r)$ in the Hamiltonian (10) have the form

$$l'(r)=l(r)+\phi(r),\quad d'_\mu(r)=d_\mu^{(p,q)}(r)+\Delta_\mu\phi(r). \quad (15)$$

These transformations lead to a deformation of magnetic basis dislocations and to the emergence of new closed dislocations. Then $d'_\mu(r)$ describes the field of closed magnetic dislocations on a torus and is the density of the topological charge of such a field. Obviously, the total topological charge of such a field of magnetic dislocations is unaltered by gauge transformations. For instance, for the Hamiltonian $H_D^{(0,0)}[K,d,\sigma]$ with dislocation $D^{(p,q)}$ we have

$$Q'_\mu=\sum_r\ d'_\mu(r)=\sum_r\ d_\mu^{(p,q)}(r)+\sum_r\ \Delta_\mu\phi(r).$$

Here the periodic boundary conditions for the phase $\phi(r)$ nullify the second term, with the result that $Q'_\mu=Q_\mu$. This implies that the set of configurations of the coupling constants $[K,d]$ (containing closed dislocations) can be divided into gauge-inequivalent classes $\Omega^{(p,q)}$, with each class having its own topological charge $Q_\mu=(pn,qm)$. The elements of a class $\Omega^{(p,q)}$ are generated by the gauge transformations (14) from the magnetic basis dislocation $D^{(p,q)}$.

The idea behind the proposed method consists in the following. In deriving duality relations via Fourier transformation, delta functions arise. The condition that the argument of the delta function must vanish leads to a system of equations that determine the relationship between the original spin variables and the dual spin variables. Usually (see, e.g., Ref. 9) the boundary conditions are not taken into account, and only one solution of this system of equations is written. However, for the model on a torus there are many solutions, which can be divided into gauge-inequivalent classes $\widetilde{\Omega}^{(\tilde{p},\tilde{q})}$ of the configurations of coupling constants for the dual model, with each class having a well-defined value of the dual topological charge $\widetilde{Q}_\mu=(\tilde{p}n,\tilde{q}m)$. Hence, in establishing the duality transformation of the partition function we must sum over all the gauge-inequivalent classes on the dual lattice with coefficients dependent on the boundary conditions on the original lattice. For instance, for $N=2$ there are four gauge-inequivalent classes: (0,0), (1,0), (0,1), and (1,1). The four terms on the right-hand side of the duality relation (3) for the Ising model reflect this situation.

In Sec. 2 we derive the duality relation for the vector Potts model on a square lattice wound around a torus. We discuss in detail the cases of self-duality for this model: $N=2$, the Ising model; and $N=3$ and $N=4$, two Ising subsystems interacting via the edge spins. In Sec. 3 we obtain the duality relations for the Berezinskiĭ–Villain $Z(N)$-model, which is self-dual for any value of N. We discuss in detail the case where $N=4$, which corresponds to the Ashkin–Teller model with subsystems of spins $\sigma(r)$ and $s(r)$ interacting in a special manner at the boundaries of the lattices wound around a torus. In Sec. 4 we derive the duality relations for the inhomogeneous Ashkin–Teller model and the inhomogeneous 8-vertex model on a lattice obliquely wound around a torus, and the exact relationship between the partition functions of these models.

2. VECTOR POTTS MODEL

We begin by deriving the duality relation for the vector Potts model. To this end we use the standard method[7–9] of duality transformation in spin models. The partition function of this model can be written

$$Z_V^{(p,q)}[K,d]=\sum_{[l]}\ \exp\{-\beta H^{(p,q)}[K,l]\}$$

$$=\sum_{[l]}\ \exp\left\{\sum_{r,\mu}\ K_\mu(r)\cos\left[\frac{2\pi}{N}[\Delta_\mu l(r)+d_\mu^{(p,q)}\times(r)]\right]\right\} \quad (16)$$

$$=\sum_{[l]}\sum_{[t]}\ \exp(-\beta\widetilde{H}[t])\exp\left\{i\frac{2\pi}{N}\sum_{r,\mu}\ t_\mu(r)\times[\Delta_\mu l(r)+d_\mu^{(p,q)}(r)]\right\} \quad (17)$$

$$=\sum_{[t]}\ \exp\left\{-\beta\widetilde{H}[t]+i\frac{2\pi}{N}\sum_{r,\mu}\ t_\mu(r)\times d_\mu^{(p,q)}(r)\right\}\prod_r\ N\delta_N(\Delta_\mu t_\mu(r-\hat{\mu})), \quad (18)$$

where

$$\sum_{[l]}=\prod_r\left(\sum_{l(r)=0}^{N-1}\right),\quad \sum_{[t]}=\prod_{r,\mu}\left(\sum_{t_\mu(r)=0}^{N-1}\right).$$

In (17) we took the Fourier transform of the vector field $t_\mu(r)$, where at each site $t_\mu(r)=0,1,\ldots,N-1$, and $-\beta\widetilde{H}[t]$ is the Fourier transform of the Hamiltonian (10):

$$-\beta\widetilde{H}[t]=\sum_{k=0}^{M}\sum_{r,\mu}\ g_\mu^{(k)}(K)\cos^k\left[\frac{2\pi}{N}t_\mu(r)\right]. \quad (19)$$

Here $M=N/2$ if N is even and $M=(N-1)/2$ if N is odd. In (18), $\delta_N(s)$ is the Kronecker delta modulo N: it is equal to unity if $s=NL$ (here L in an integer), and zero otherwise.

To get rid of the delta functions in (18) we must solve the equation

$$\Delta_\mu t_\mu(r-\hat{\mu})=0\pmod N. \quad (20)$$

The nontrivial solutions of this equation on a torus can be written

$$t_\mu^{(\alpha)}(r)=\epsilon_{\mu\nu}\Delta_\nu\tilde{l}(\tilde{r}-\hat{\nu})+\epsilon_{\mu\nu}\tilde{d}_\nu^{(\alpha)}(\tilde{r}-\hat{\nu}), \quad (21)$$

where α labels the solutions, $\tilde{l}(\tilde{r})=0,1,\ldots,N-1$ is specified at the sites of the dual lattice, and $\tilde{d}_\nu^{(\alpha)}(\tilde{r})$ is the density of the topological charge (corresponding to a given αth solution) of the field of closed magnetic dislocations on the dual lattice,

$$\tilde{d}_\mu^{(\alpha)}(\tilde{r})=\sum_{i\in Z_\alpha} s_i^{(\alpha)} \sum_{r'\in\Gamma_i} \epsilon_{\mu\nu} a_\nu(r')\delta^2(r-r'),$$

$$s_i^{(\alpha)}=0,1,\ldots,N-1. \tag{22}$$

Here, by analogy with (11) and (12), the function $\tilde{d}_\mu^{(\alpha)}(\tilde{r})$ is defined on the edges of the dual lattice. For the sake of convenience we have written the dislocations on the dual lattice in terms of closed paths Γ_i on the original lattice. In (22), Z_α denotes the subset of paths corresponding to solution α out of the set Γ of all closed paths on the original lattice ($\Gamma_i\in\Gamma$). The vector $a_\mu(r)$ is equal to $e_\mu(r)$ if the detour direction of path Γ_i at site r (counterclockwise) is the same as that of the unit vector $e_\mu(r)=\hat{\mu}$ at that site; otherwise $a_\mu(r)=-e_\mu(r)$.

The expression (22) can easily be obtained if we note that the solution (21) formally satisfies Eq. (20) at the site $\tilde{r}$ provided that

$$\epsilon_{\mu\nu}\Delta_\mu \tilde{d}_\nu^{(\alpha)}(\tilde{r}-\hat{\mu}-\hat{\nu})=0,$$

which becomes an identity if one of the three following pairs of conditions is met:

$$\tilde{d}_y^{(\alpha)}(\tilde{r}-\hat{y})=\tilde{d}_y^{(\alpha)}(\tilde{r}-\hat{x}-\hat{y}),$$
$$\tilde{d}_x^{(\alpha)}(\tilde{r}-\hat{x})=\tilde{d}_x^{(\alpha)}(\tilde{r}-\hat{x}-\hat{y});$$
$$\tilde{d}_y^{(\alpha)}(\tilde{r}-\hat{y})=\tilde{d}_x^{(\alpha)}(\tilde{r}-\hat{x}),$$
$$\tilde{d}_y^{(\alpha)}(\tilde{r}-\hat{x}-\hat{y})=\tilde{d}_x^{(\alpha)}(\tilde{r}-\hat{x}-\hat{y});$$
$$\tilde{d}_y^{(\alpha)}(\tilde{r}-\hat{y})=-\tilde{d}_x^{(\alpha)}(\tilde{r}-\hat{x}-\hat{y}),$$
$$\tilde{d}_x^{(\alpha)}(\tilde{r}-\hat{x})=\tilde{d}_y^{(\alpha)}(\tilde{r}-\hat{x}-\hat{y}).$$

For these solutions to be valid simultaneously at a set of sites we must require that the sites lie on closed paths Γ_i on the torus, i.e., these solutions must be ''matched'' in such a way that they form closed magnetic dislocations.

We denote a configuration of the coupling constants on the dual lattice corresponding to the solution (22) by $[\tilde{d}^{(\alpha)}]$. Depending on the number labeling the solution, these configurations contain both closed dislocations not encompassing the cycles of the torus (their topological charge is zero) and dislocations encompassing the cycles. Dislocations of the first type can be eliminated via gauge transformations (15) on the dual lattice, while dislocations of the second type can be transformed into magnetic basis dislocations $\tilde{D}^{(\tilde{p},\tilde{q})}$ on the dual lattice. This means that all the configurations $[\tilde{d}^{(\alpha)}]$ can be divided into gauge-inequivalent classes $\tilde{\Omega}^{(\tilde{p},\tilde{q})}$ with topological charge

$$\tilde{Q}_\nu=\sum_{\tilde{r}} \tilde{d}_\nu^{(\tilde{p},\tilde{q})}(\tilde{r})=(\tilde{p}n,\tilde{q}m),$$

where $\tilde{p},\tilde{q}=0,1,\ldots,N-1$. Since the duality relations link the partition functions of the two models—functions that are gauge-invariant quantities—in removing the delta functions in (18) one should sum over the gauge-inequivalent solutions of Eq. (20):

$$t_\mu^{(\tilde{p},\tilde{q})}(r)=\epsilon_{\mu\nu}\Delta_\nu \tilde{l}(\tilde{r}-\hat{\nu})+\epsilon_{\mu\nu}\tilde{d}_\nu^{(\tilde{p},\tilde{q})}(\tilde{r}-\hat{\nu}), \tag{23}$$

where $\tilde{d}_\mu^{(\tilde{p},\tilde{q})}$ is defined on the dual lattice by relations similar to (11)–(13). Plugging these solutions into (18), we obtain

$$Z_V^{(p,q)}[K,d]=\frac{1}{N}\sum_{\tilde{p},\tilde{q}}\sum_{[\tilde{l}]} \exp(-\beta\tilde{H}[\Delta_\mu\tilde{l}+\tilde{d}_\mu^{(\tilde{p},\tilde{q})}])$$
$$\times\exp\left\{ i\frac{2\pi}{N}\sum_{r,\mu}\epsilon_{\mu\nu}d_\mu^{(p,q)}(r)[\Delta_\nu\tilde{l}(\tilde{r}-\hat{\nu})\right.$$
$$\left.+\tilde{d}_\nu^{(\tilde{p},\tilde{q})}(\tilde{r}-\hat{\nu})]\right\}.$$

Here the factor $1/N$ must be introduced because Eq. (23) leads to a situation in which the sum over the configurations $[l]$ is N times the sum over the configurations $[t]$. Noting that

$$\sum_{r,\mu}\epsilon_{\mu\nu}d_\mu^{(p,q)}(r)\Delta_\nu\tilde{l}(\tilde{r}-\hat{\nu})=0,$$

we can write (18) in the more compact form

$$Z_V^{(p,q)}[K,d]=\frac{1}{N}\sum_{\tilde{p},\tilde{q}}\exp\left\{ i\frac{2\pi}{N}\sum_{r,\mu}\epsilon_{\mu\nu}d_\mu^{(p,q)}(r)\right.$$
$$\left.\times\tilde{d}_\nu^{(\tilde{p},\tilde{q})}(\tilde{r}-\hat{\nu})\right\}\tilde{Z}_V^{(\tilde{p},\tilde{q})}[\tilde{K},\tilde{d}]$$
$$=\frac{1}{N}\sum_{\tilde{p},\tilde{q}}\exp\left\{ i\frac{2\pi}{N}(p\tilde{q}-q\tilde{p})\right\}$$
$$\times\tilde{Z}_V^{(\tilde{p},\tilde{q})}[\tilde{K},\tilde{d}], \tag{24}$$

where

$$\tilde{Z}_V^{(\tilde{p},\tilde{q})}[\tilde{K},\tilde{d}]=\sum_{[\tilde{l}]}\exp(-\beta\tilde{H}^{(\tilde{p},\tilde{q})}[\tilde{l},\tilde{d}])$$

is the partition function of the model on the dual lattice. The duality relation (24) shows that the vector Potts model is not self-dual for arbitrary values of N. Here we discuss the cases with $N=2,3,4$, for which this model is self-dual.[8]

The Hamiltonian (7) at $N=2$ coincides with the Hamiltonian (4) of the Ising model. In this case Eq. (19) yields

$$-\beta\tilde{H}_2^{(\tilde{p},\tilde{q})}[\tilde{l}]=\sum_{r,\nu}[g_\nu^{(0)}(\tilde{K})+g_\nu^{(1)}(\tilde{K})\cos(\pi\Delta_\nu^{(\tilde{p}_\nu)}\tilde{l}(r))].$$

To find the coefficients $g_\mu^{(i)}(\tilde{K})$ we take the inverse Fourier transform:

$$\exp\left\{\sum_{k=0}^{M} g_\mu^{(k)}(K)\cos^k\left(\frac{2\pi}{N}\tilde{t}_\mu\right)\right\}$$

$$= \frac{1}{N}\sum_{n=0}^{N-1} \exp\left[K_\nu \cos\left(\frac{2\pi}{N}n\right) - i\frac{2\pi}{N}n\tilde{t}_\nu\right], \quad \mu \neq \nu. \tag{25}$$

At $N=2$ $(M=1)$ this relation readily yields

$$\exp\{2g_\mu^{(0)}(\tilde{r})\} = \tfrac{1}{2}\sinh 2K_\nu(r),$$

$$\exp\{-2g_\mu^{(1)}(\tilde{r})\} = \tanh K_\nu(r) = \exp[-2\tilde{K}_{-\mu}(\tilde{r})],$$

where the second equation coincides with (2). Using these expressions, we can write the duality relation (24) as

$$\left\{\prod_{r,\nu} \sinh 2K_\nu(r)\right\}^{-1/4} Z^{(p,q)}[K] = \frac{1}{2}\left\{\prod_{\tilde{r},\nu} \sinh 2\tilde{K}_\nu(\tilde{r})\right\}^{-1/4} \times \sum_{\tilde{p},\tilde{q}=0}^{1} \exp\{i\pi(p\tilde{q}-q\tilde{p})\}\tilde{Z}^{(\tilde{p},\tilde{q})}[\tilde{K}], \tag{26}$$

where

$$\tilde{Z}^{(\tilde{p},\tilde{q})}[\tilde{K}] = \sum_{[l]} \exp\sum_{\tilde{r},\nu} \tilde{K}_\nu(\tilde{r})\cos[\pi\Delta_\nu^{(\tilde{p}_\nu)}\tilde{l}(r)] = \sum_{[\sigma]} \exp\sum_{\tilde{r},\nu} \tilde{K}_\nu(\tilde{r})\tilde{\sigma}(\tilde{r})\nabla_\nu^{(\tilde{p}_\nu)}\tilde{\sigma}(\tilde{r}),$$

and $\tilde{\sigma}(\tilde{r}) = \pm 1$. Clearly, (26) coincides with the duality relation (3), since the exponential on the right-hand side of Eq. (26) is nothing more than the matrix $\hat{T}$ in (3):

$$T_{\tilde{p}_x,\tilde{p}_y}^{p_x,p_y} = T_{\tilde{p},\tilde{q}}^{p,q} = \exp\{i\pi(p\tilde{q}-q\tilde{p})\}. \tag{27}$$

The Hamiltonian (7) at $N=3$ corresponds to the three-position vector Potts model. In this case (16) yields

$$\exp(g_\mu^{(0)}(\tilde{r})) = \frac{1}{3}\left\{\exp(K_\nu(r)) + 2\times\exp\left(-\frac{K_\nu(r)}{2}\right)\right\}^{1/3}\left\{\exp(K_\nu(r)) - \exp\left(-\frac{K_\nu(r)}{2}\right)\right\}^{2/3}\exp\left(-\frac{3}{2}g_\mu^{(1)}(\tilde{r})\right)$$

$$= \frac{2\sinh[3K_\nu(r)/4]}{3\cosh[3K_\nu(r)/4] - \sinh[3K_\nu(r)/4]} = \exp\left(-\frac{3}{2}\tilde{K}_\mu(\tilde{r})\right),$$

and the duality relation (24) assumes the form

$$Z_3^{(p,q)}[K] = \frac{1}{3}\prod_{\tilde{r},\mu} \exp(g_\mu^{(0)}(\tilde{r})) \times \sum_{\tilde{p},\tilde{q}=0}^{2} \exp\left\{i\frac{2\pi}{3}(p\tilde{q}-q\tilde{p})\right\}\tilde{Z}_3^{(\tilde{p},\tilde{q})}[\tilde{K}].$$

At $N=4$ the partition function of the model (7) with cyclic boundary conditions (9) can be written as the product of the partition functions of two Ising models interacting via boundary spins. Indeed, if we replace the spin variable via the relation

$$\sigma(r) = \exp\left(i\frac{\pi}{2}l(r)\right) = \frac{1}{2}\{\tau_1(r)+\tau_2(r)+i[\tau_1(r)-\tau_2(r)]\}, \tag{28}$$

which each value $l=0,1,2,3$ uniquely assigns $\tau_1(r)$ and $\tau_2(r)$ $(\tau_i(r)=\pm 1)$, the Hamiltonian (7) can be written in the form

$$-\beta H_4^{(p,q)} = \sum_{r,\nu} K_\nu(r)\cos\left\{\frac{\pi}{2}\Delta_\nu^{(p_\nu)}l(r)\right\} = \frac{1}{2}\sum_{r,\nu,i} K_\nu(r)\tau_i(r)\overline{\nabla}_\nu^{(p,q)}\tau_i(r). \tag{29}$$

Here $p,q=0,1,2,3$, and in view of the boundary conditions (8) for $\sigma(r)$ and Eq. (28), the operator $\overline{\nabla}_\nu^{(p,q)}$ satisfies the boundary conditions

$$\overline{\nabla}_x^{(p,q)}\tau_i(n,y) = \tau_i(n+1,y) = \left[\cos\frac{\pi p}{2} + (-)^j \sin\frac{\pi q}{2}\right]\tau_j(1,y), \tag{30}$$

$$\overline{\nabla}_y^{(p,q)}\tau_i(x,m) = \tau_i(x,m+1) = \left[\cos\frac{\pi q}{2} + (-)^j \sin\frac{\pi p}{2}\right]\tau_j(x,1), \tag{31}$$

which generate the interaction of spins $\tau_1(r)$ and $\tau_2(r)$ at the boundary $(i\neq j)$.

Using (25), we can calculate the coefficients in the Hamiltonian (29):

$$\exp(-g_\mu^{(0)}(\tilde{r})) = \frac{1}{2}\sinh K_\nu(r),$$

$$\exp(-g_\mu^{(1)}(\tilde{r})) = \tanh\frac{K_\nu(r)}{2} = \exp(-\tilde{K}_\mu(\tilde{r})),$$

$$g_\mu^{(2)}(\tilde{r}) = 0.$$

From (24) we obtain the duality relation for two Ising models on a torus that interact via boundary spins:

$$\left\{\prod_{r,\mu} \sinh K_\mu(r)\right\}^{-1/2} Z_{2I}^{(p,q)}[K] = \frac{1}{4}\left\{\prod_{\tilde{r},\mu} \sinh\tilde{K}_\mu(\tilde{r})\right\}^{-1/2} \times \sum_{\tilde{p},\tilde{q}=0}^{3} \exp\left\{i\frac{\pi}{2}(p\tilde{q}-q\tilde{p})\right\}\tilde{Z}_{2I}^{(\tilde{p},\tilde{q})}[\tilde{K}],$$

where for this model

$$\sinh K_x(r)\sinh\tilde{K}_{-y}(\tilde{r}) = 1,$$

$$\sinh K_y(r)\sinh\tilde{K}_{-x}(\tilde{r}) = 1.$$

3. BEREZINSKIĭ–VILLAIN $Z(N)$-MODEL

We now derive the duality relations for the Berezinskiĭ–Villain $Z(N)$-model[5,6] (a $Z(N)$-symmetric Gaussian model). The partition function of this model can be written[7,8]

$$Z_{BV}^{(p,q)}[K]=\sum_{[l]}\ \exp(-\beta H_G^{(p,q)}[K,l])$$
$$=\sum_{[l]}\sum_{[k]}\prod_{r,\mu}\ \exp\left\{-\frac{K_\mu(r)}{2}\left[\frac{2\pi}{N}\Delta_\mu l(r)-2\pi k_\mu(r)\right]^2\right\}, \tag{32}$$

where

$$\sum_{[l]}\cdots=\prod_r\left(\sum_{l(r)=0}^{N-1}\cdots\right),$$
$$\sum_{[k]}\cdots=\prod_{r,\mu}\left(\sum_{k_\mu(r)=-\infty}^{\infty}\right).$$

Here the quantity $l(r)=0,\ldots,N-1$ is specified at the sites of a square lattice, the superscript (p,q) specifies the boundary conditions (9), and the sum over k_μ ensures the periodicity of the Hamiltonian under shifts $l\to l(r)+NL(r)$, where L is an integer. By analogy with the vector Potts model, the partition function (32) can be written in terms of the basis dislocations $D^{(p,q)}$:

$$Z_{BV}^{(p,q)}[K,d]=\sum_{[l]}\sum_{[k]}\prod_{r,\mu}\ \exp\left\{-\frac{K_\mu(r)}{2}\left[\frac{2\pi}{N}[\Delta_\mu l(r)+d_\mu^{(p,q)}(r)]-2\pi k_\mu(r)\right]^2\right\}. \tag{33}$$

where $l(r)$ satisfies the periodic boundary conditions, and the topological charge density $d_\mu(r)$ is defined in (11)–(13).

To derive the duality relation we transform (33) as follows:

$$Z_{BV}^{(p,q)}[K,d]=\sum_{[l]}\sum_{[k]}\prod_{r,\mu}\exp\left\{-\frac{K_\mu(r)}{2}\times\left[\frac{2\pi}{N}[\Delta_\mu l(r)+d_\mu(r)]-2\pi k_\mu(r)\right]^2\right\}$$
$$=\left(\prod_{r,\mu}N\right)^{1/2}\sum_{[s]}\sum_{[k]}\int D\theta\prod_{r,\mu}\ \exp\left\{-\frac{K_\mu(r)}{2}\times\left[\left(\Delta_\mu\theta(r)+\frac{2\pi}{N}d_\mu(r)\right)-2\pi k_\mu(r)\right]^2+i\frac{N}{2}s(r)\theta(r)\right\} \tag{34}$$
$$=\left(\prod_{r,\mu}\ \frac{N}{2\pi K_\mu(r)}\right)^{1/2}\sum_{[s]}\sum_{[t]}\int D\theta\times\prod_{r,\mu}\ \exp\left\{-\frac{t_\mu^2(r)}{2K_\mu(r)}+it_\mu(r)\left(\Delta_\mu\theta(r)+\frac{2\pi}{N}d_\mu(r)\right)+i\frac{N}{2}s(r)\theta(r)\right\} \tag{35}$$
$$=\left(\prod_{r,\mu}\ \frac{N}{2\pi K_\mu(r)}\right)^{1/2}\times\sum_{[s]}\sum_{[t]}\prod_{r,\mu}\ \exp\left\{-\frac{t_\mu^2(r)}{2K_\mu(r)}+i\frac{2\pi}{N}t_\mu(r)d_\mu(r)\right\}\times\prod_r\ \delta\left(\sum_\mu\Delta_\mu t_\mu(r-\hat\mu)-Ns(r)\right), \tag{36}$$

where

$$\sum_{[t]}\cdots=\prod_{r,\mu}\left(\sum_{t_\mu(r)=-\infty}^{\infty}\cdots\right),\quad\int D\theta=\prod_r\int_0^{2\pi}\frac{d\theta(r)}{2\pi}.$$

In deriving (34), (35), and (36) we used, respectively, the summation formula

$$\frac{2\pi}{N}\sum_{l=0}^{N-1}\delta\left(\theta-\frac{2\pi}{N}l\right)=\sum_{s=-\infty}^{\infty}e^{iNs\theta},\quad 0\leqslant\theta\leqslant 2\pi,$$

the equality

$$\sum_{k=-\infty}^{\infty}\ \exp\left[-\frac{1}{2}K(f-2\pi k)^2\right]=\frac{1}{\sqrt{2\pi K}}\sum_{t=-\infty}^{\infty}\ \exp\left(-\frac{t^2}{2K}+itf\right),$$

and the definition of the Kroneckers delta

$$\int_0^{2\pi}\frac{d\theta(r)}{2\pi}e^{i\theta l}=\delta(l).$$

To remove the delta functions in (36) we must solve the equation

$$\sum_\mu\Delta_\mu t_\mu(r-\hat\mu)=Ns(r).$$

The analysis of the solutions of this equation is similar to the analysis of Eqs. (20)–(23), and leads to the following expressions for the gauge-inequivalent solutions:

$$t_\mu^{(\tilde p,\tilde q)}(r)=\epsilon_{\mu\nu}\Delta_\nu\tilde l(\tilde r-\hat\nu)+\epsilon_{\mu\nu}\tilde d_\nu^{(\tilde p,\tilde q)}(\tilde r-\hat\nu)-N\epsilon_{\mu\nu}\tilde k_\nu(\tilde r-\hat\nu),$$
$$s(r)=\epsilon_{\mu\nu}\Delta_\mu\tilde k_\nu(\tilde r-\hat\nu-\hat\mu).$$

These solutions are the magnetic basis dislocations $\tilde D^{(\tilde p,\tilde q)}$ on the dual lattice in the corresponding gauge-inequivalent $\tilde\Omega^{(\tilde p,\tilde q)}$ with topological charge $\tilde Q_\mu=(\tilde pn,\tilde qm)$ $(\tilde p,\tilde q=0,1,\ldots,N-1)$. To remove the delta functions in (36), we must then sum over all such solutions. As a result we arrive at the duality relation for the Berezinskiĭ–Villain $Z(N)$-model:

$$\left(\prod_{r,\mu}\frac{2\pi K_\mu(r)}{N}\right)^{1/4} Z_{BV}^{(p,q)}[K,d]$$
$$=\frac{1}{N}\sum_{\tilde p,\tilde q}\exp\left\{i\frac{2\pi}{N}(p\tilde q-q\tilde p)\right\}$$
$$\times\left(\prod_{\tilde r,\mu}\frac{2\pi\tilde K_\mu(\tilde r)}{N}\right)^{1/4}\tilde Z_{BV}^{(\tilde p,\tilde q)}[\tilde K,\tilde d], \tag{37}$$

where

$$K_\mu(r)\tilde K_{-\nu}(\tilde r)=\left(\frac{N}{2\pi}\right)^2,\quad \mu\neq\nu.$$

As shown in Ref. 7, at $N=2$ the $Z(N)$-symmetric Gaussian model corresponds to the Ising model, which agrees with our result (37), which in turn agrees with (3) in this case, while at $N=4$ it corresponds to the Ashkin–Teller model.[12]

We now discuss the latter case in greater detail. The Hamiltonian of the inhomogeneous Ashkin–Teller model can be written

$$-\beta H_{AT}^{(p_\nu,q_\nu)}=\sum_{r,\nu}[P_\nu(r)+I_\nu(r)\tau_1(r)\nabla_\nu^{(p_\nu)}\tau_1(r)$$
$$+J_\nu(r)\tau_2(r)\nabla_\nu^{(q_\nu)}\tau_2(r)$$
$$+L_\nu(r)\tau_1(r)\nabla_\nu^{(p_\nu)}\tau_1(r)\tau_2(r)\nabla_\nu^{(q_\nu)}\tau_2(r)]. \tag{38}$$

Here at each site r of a square $n\times m$ lattice R there are two spin variables, $\sigma(r)$ and $s(r)$, which take the values ±1; the subscript $\nu=x,y$ labels the x and y coordinate axes, and the indices $p_\nu=0,1$ and $q_\nu=0,1$ ($\nu=x,y$) describe the boundary conditions (6) on the spin variables $\sigma(r)$ and $s(\tau)$, respectively; I_ν, J_ν, and L_ν are the coupling constants along the corresponding axes.

The statistical weight of the Ashkin–Teller model is related to the statistical weight of the $Z(4)$-symmetric Gaussian model[12]:

$$\exp[P_\nu(r)+I_\nu(r)\tau_1(r)\nabla_\nu\tau_1(r)+J_\nu(r)\tau_2(r)\nabla_\nu\tau_2(r)$$
$$+L_\nu(r)\tau_1(r)\nabla_\nu\tau_1(r)\tau_2(r)\nabla_\nu\tau_2(r)]$$
$$=\sum_{k_\mu(r)=-\infty}^{\infty}\exp\left\{-2\pi^2K_\nu(r)\times\left[\frac14(\Delta_\nu l(r)+d_\nu(r))-k_\nu(r)\right]^2\right\},$$

where $l(r)$ is related to the spin variables $\tau_1(r)$ and $\tau_2(r)$ by (28). This leads to the following parametrization of the coupling constants of the Ashkin–Teller model in terms of the constants K_ν:

$$\exp(4P_\nu)=\omega_0\omega_1\omega_2\omega_3,\quad \exp(4L_\nu)=\frac{\omega_0\omega_3}{\omega_1\omega_2},$$

$$\exp(4I_\nu)=\frac{\omega_0\omega_1}{\omega_2\omega_3},\quad \exp(4J_\nu)=\frac{\omega_0\omega_2}{\omega_1\omega_3},$$

where

$$\omega_0=\sum_{m=-\infty}^{\infty}\exp(-2\pi^2K_\nu m^2),$$

$$\omega_1=\omega_2=\sum_{m=-\infty}^{\infty}\exp\left[-2\pi^2K_\nu\left(m-\frac14\right)^2\right],$$

$$\omega_3=\sum_{m=-\infty}^{\infty}\exp\left[-2\pi^2K_\nu\left(m-\frac12\right)^2\right].$$

This implies that $I_\nu=J_\nu$. Using this relation and (37), we can easily obtain the self-duality relation for the Ashkin–Teller model:

$$Z_{AT}^{(p,q)}[P,J,L]=\frac14\sum_{\tilde p,\tilde q=0}^{3}\exp\left\{i\frac{\pi}{2}(p\tilde q-q\tilde p)\right\}$$
$$\times\tilde Z_{AT}^{(\tilde p,\tilde q)}[\tilde P,\tilde J,\tilde L], \tag{39}$$

where

$$Z_{AT}^{(p,q)}=\sum_{[\tau_1],[\tau_2]}\exp\{-\beta H_{AT}^{(p,q)}[\tau_1,\tau_2]\}.$$

Here the Hamiltonian $H_{AT}^{(p,q)}$ has the same structure as the Hamiltonian (38), but the shift operators $\nabla_\nu^{(p_\nu)}$ and $\nabla_\nu^{(q_\nu)}$ are replaced by $\overline{\nabla}_{x,y}^{(p,q)}$ ($p,q=0,\ldots,3$) with the boundary conditions (30) and (31), which leads to the interaction of spin variables $\tau_1(r)$ and $\tau_2(r)$ at the boundaries. As noted earlier, these boundary conditions emerge because of the boundary conditions (9) on the variable $l(r)$ and the relation (28).

In (39) the dual coupling constants are related to the original coupling constants:

$$\exp(-2J_\mu(r)-2L_\mu(r))=\frac{\tanh\tilde J_{-\nu}(\tilde r)[1+\tanh\tilde L_{-\nu}(\tilde r)]}{1+\tanh^2\tilde J_{-\nu}(\tilde r)\tanh\tilde L_{-\nu}(\tilde r)},$$

$$\exp\{-4J_\mu(r)\}=\frac{\tanh^2\tilde J_{-\nu}(\tilde r)+\tanh\tilde L_{-\nu}(\tilde r)}{1+\tanh^2\tilde J_{-\nu}(\tilde r)\tanh\tilde L_{-\nu}(\tilde r)},$$

$$\frac12\exp(P_\mu+2J_\mu+L_\mu)=\exp(\tilde P_{-\nu})[\cosh^2\tilde J_{-\nu}\cosh\tilde L_{-\nu}$$
$$+\sinh^2\tilde J_{-\nu}\sinh\tilde L_{-\nu}].$$

These equalities constitute a special case of the relations to be obtained in Sec. 4, where we derive more general self-duality formulas for the Ashkin–Teller model with the Hamiltonian (38).

4. ASHKIN–TELLER MODEL AND 8-VERTEX MODEL

Consider the partition function

$$Z_{AT}^{(p_\nu,q_\nu)}[P,I,J,L]=\sum_{[\sigma],[s]}\exp(-\beta H_{AT}^{(p_\nu,q_\nu)}) \tag{40}$$

of the inhomogeneous Ashkin–Teller model on an $n\times m$ finite lattice R wound around a torus with the Hamiltonian (38) and the boundary conditions (6) on the spin variables $\sigma(r)$ and $s(r)$.

For a duality transformation of the partition function (40) we can use the duality relation (3) for the Ising model. To this end we factor the four-spin term in (40) via the simple identity

$$\exp[L_\nu(r)\sigma(r)\sigma(r+\hat{\nu})s(r)s(r+\hat{\nu})]$$

$$=\sum_{\tau_\nu(r)}\alpha_\nu\exp\{\gamma_\nu(r)\tau_\nu(r)$$

$$\times[\sigma(r)\sigma(r+\hat{\nu})+s(r)s(r+\hat{\nu})]\},$$

where the spin variable $\tau_\nu(r)=\pm 1$ is on the edges of the lattice $(r,r+\nu)$, and

$$\exp[2L_\nu(r)]=\cosh 2\gamma_\nu(r),\quad \exp[-L_\nu(r)]=2\alpha_\nu(r). \tag{41}$$

After factorization we arrive at the following representation of the partition function (40), of a linear combination of products of two partition functions of inhomogeneous Ising models:

$$Z_{AT}^{(p_\nu,q_\nu)}=\sum_{[\tau_\nu]}\prod_{r,\nu}\{\alpha_\nu(r)\exp(P_\nu(r))\}$$

$$\times Z_\sigma^{(p_x,p_y)}[I'_\mu]Z_s^{(q_x,q_y)}[J'_\nu], \tag{42}$$

where

$$Z_\sigma^{(p_x,p_y)}[I'_\mu]=\sum_{[\sigma]}\prod_{r,\nu}\exp[I'_\nu(r)\sigma(r)\nabla_\nu^{(p_\nu)}\sigma(r)],$$

$$Z_s^{(q_x,q_y)}[J'_\nu]=\sum_{[s]}\prod_{r,\nu}\exp[J'_\nu(r)s(r)\nabla_\nu^{(q_\nu)}s(r)],$$

and the coupling constants are

$$I'_\nu(r)=I_\nu(r)+\gamma_\nu(r)\tau_\nu(r),$$

$$J'_\nu(r)=J_\nu(r)+\gamma_\nu(r)\tau_\nu(r). \tag{43}$$

Since the duality relation (3) is valid for an arbitrary configuration of the coupling constants $[K]$, we can use it for the duality transformation of the partition functions $Z_s^{(q_x,q_y)}$ and $Z_\sigma^{(p_x,p_y)}$. Here, performing the duality transformation only for $Z_s^{(q_x,q_y)}$, for example, we find the relation between the partition functions of the Ashkin–Teller model and the 8-vertex model, while the simultaneous transformation of $Z_s^{(q_x,q_y)}$ and $Z_\sigma^{(p_x,p_y)}$ leads to the partition function of the Ashkin–Teller dual model.

We start with the first case. Using relation (3) for the duality transformation of $Z_s^{(q_x,q_y)}$, we obtain

$$Z_{AT}^{(p_\nu,q_\nu)}=\frac{1}{2}\prod_{r,\nu}\{\alpha_\nu(r)\exp(P_\nu(r))\}\sum_{[\tau_\nu]}Z_\sigma^{(p_x,p_y)}[I']$$

$$\times\sum_{\tilde{q}_x,\tilde{q}_y=0}^{1}\exp[i\pi(q_x\tilde{q}_y-q_y\tilde{q}_x)]$$

$$\times\sum_{[\tau]}\prod_{\tilde{r},\nu}\{\sinh 2\tilde{J}'_{-\nu}(\tilde{r})\}^{1/2}$$

$$\times\exp[\tilde{J}'_{-\nu}(\tilde{r})\tau(\tilde{r})\nabla_{-\nu}^{(q_\nu)}\tau(\tilde{r})], \tag{44}$$

where by $\tau(\tilde{r})$ we denote the spin variable dual to $s(r)$, and

$$\tanh\tilde{J}'_{-\nu}(\tilde{r})=\exp[-2J'_\mu(r)],\quad \mu\neq\nu, \tag{45}$$

with $\tilde{r}$ designating the sites of the dual lattice $\tilde{R}$ linked to the coordinates of the sites of the original lattice via the relationship $\tilde{r}=r+(\hat{x}+\hat{y})/2$, and $\tilde{J}'_{-\nu}(\tilde{r})=\tilde{J}'_\nu(\widetilde{r-\hat{\nu}})$. Thus, on the right-hand side of (44) we have the partition function in which the spin variables are given on two lattices, R and $\tilde{R}$. Using (45) and the identity

$$\exp(K\sigma_i\sigma_j)=\cosh K+\sinh K\sigma_i\sigma_j, \tag{46}$$

we can write the sum over $[\tau_\nu]$ in (44) in the following form:

$$\prod_{r,\nu}\sum_{\tau_\nu=\pm 1}[\exp(J'_\nu)\cosh I'_\nu$$

$$+\exp(J'_\nu)\sinh I'_\nu\sigma(r)\nabla_\nu^{(p_\nu)}\sigma(r)$$

$$+\exp(-J'_\nu)\cosh I'_\nu\tau(\tilde{r})\nabla_{-\mu}^{(\tilde{q}_{-\mu})}\tau(\tilde{r})$$

$$+\exp(-J'_\nu)\sinh I'_\nu\sigma(r)\nabla_\nu^{(p_\nu)}\sigma(r)\tau(\tilde{r})$$

$$\times\nabla_{-\mu}^{(\tilde{q}_\mu)}\tau(\tilde{r})],$$

$\mu\neq\nu$.

After summing over τ_ν we obtain

$$Z_{AT}^{(p_\nu,q_\nu)}=\frac{1}{2}\prod_{r,\nu}\frac{\exp[P_\nu(r)-L_\nu(r)]}{\sqrt{2}}$$

$$\times\sum_{\tilde{q}_x,\tilde{q}_y=0}^{1}\exp[i\pi(q_x\tilde{q}_y-q_y\tilde{q}_x)]$$

$$\times\sum_{[\tau,\sigma]}\prod_{r,\nu}\exp(J_\nu+L_\nu)\cosh(I_\nu+L_\nu)$$

$$\times\Bigg[1+\tanh(I_\nu+L_\nu)\sigma(r)\nabla_\nu^{(p_\nu)}\sigma(r)$$

$$+\exp(-2J_\nu)\frac{\cosh(I_\nu-L_\nu)}{\cosh(I_\nu+L_\nu)}\tau(\tilde{r})\nabla_{-\mu}^{(\tilde{q}_\mu)}\tau(\tilde{r})$$

$$+\exp(-2J_\nu)\frac{\sinh(I_\nu-L_\nu)}{\cosh(I_\nu+L_\nu)}\sigma(r)$$

$$\times\nabla_\nu^{(p_\nu)}\sigma(r)\tau(\tilde{r})\nabla_{-\mu}^{(\tilde{q}_\mu)}\tau(\tilde{r})\Bigg]. \tag{47}$$

The partition function of the 8-vertex model can be represented in the form of the partition function of a spin model consisting of two Ising sublattices R and $\tilde{R}$ coupled by four-spin interaction.[10] In general form, it is convenient to write the Hamiltonian of such a model as

$$-\beta H_{\mathrm{IV}}^{(p_\nu,q_\nu)}=\sum_{r,\nu}[P_\nu^v(r)+I_\nu^v(r)\sigma(r)\nabla_\nu^{(p_\nu)}\sigma(r)$$

$$+J_{-\mu}^v(\tilde{r})\tau(\tilde{r})\nabla_{-\mu}^{(q_\mu)}\tau(\tilde{r})$$

$$+L_\nu^v(r)\sigma(r)\nabla_\nu^{(p_\nu)}\sigma(r)\tau(\tilde{r})\nabla_{-\mu}^{(q_\mu)}\tau(\tilde{r})],$$

$\mu\neq\nu$, (48)

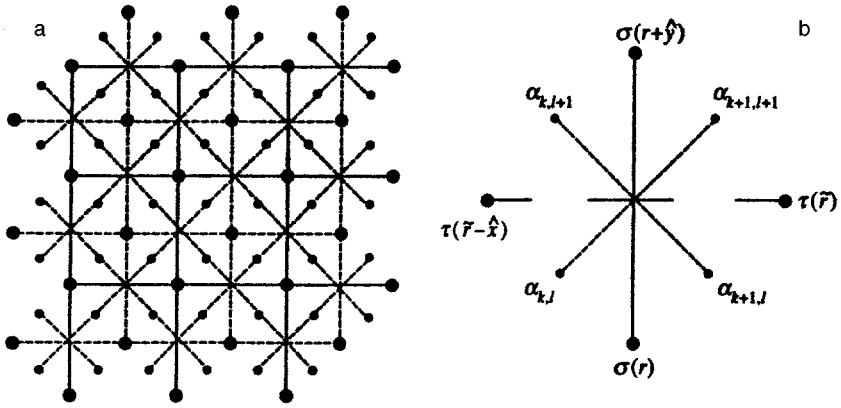


FIG. 2. a) The two-sublattice Ising model and the 8-vertex model with the lattice obliquely wound around a torus. b) The relationship between the spin variables $\sigma(r)$, $\tau(\tilde{r})$, and $\alpha_{k,l}$.

where again the spins $\sigma(r)=\pm 1$ belong to the original lattice R, and the spins $\tau(\tilde{r})=\pm 1$ are at the sites of the lattice R dual to the original lattice (see Fig. 2a). The partition function of such a model is the same as the partition function (47) for certain relations between the coupling constants of these models. Indeed, using (46), we find that

$$Z_{\mathrm{IV}}^{(p_\nu,q_\nu)}=\sum_{[\tau],[\sigma]}\exp(-\beta H_{\mathrm{IV}}^{(p_\nu,q_\nu)})$$

$$=\sum_{[\tau],[\sigma]}\prod_{r,\nu}\overline{P}_\nu[1+\overline{I}_\nu\sigma(r)\nabla_\nu^{(p_\nu)}\sigma(r)$$

$$+\overline{J}_\nu\tau(\tilde{r})\nabla_{-\mu}^{(q_\mu)}\tau(\tilde{r})+\overline{L}_\nu\sigma(r)\nabla_\nu^{(p_\nu)}\sigma(r)$$

$$\times\tau(\tilde{r})\nabla_\mu^{(q_\mu)}\tau(\tilde{r})], \tag{49}$$

and the coupling constants are determined below. Comparing (47) and (49), we obtain equations that relate the coupling constants of the Ashkin–Teller model to those of the spin model with the Hamiltonian (48),

$$\frac{\exp[P_\nu(r)+J_\nu(r)]}{\sqrt{2}}$$

$$=\exp[P_\mu^\upsilon(r)][\cosh I_\nu^\upsilon(r)\cosh J_{-\mu}^\upsilon(\tilde{r})\cosh L_\nu^\upsilon(r)$$

$$+\sinh I_\nu^\upsilon(r)\sinh J_{-\mu}^\upsilon(\tilde{r})\sinh L_\nu^\upsilon(r)]=\overline{P}_\nu, \tag{50}$$

$$\tanh(I_\nu+L_\nu)=\frac{\tanh I_\nu^\upsilon(r)+\tanh J_{-\mu}^\upsilon(\tilde{r})\tanh L_\nu^\upsilon(r)}{1+\tanh I_\nu^\upsilon(r)\tanh J_{-\mu}^\upsilon(\tilde{r})\tanh L_\nu^\upsilon(r)}=\overline{I}_\nu, \tag{51}$$

$$\exp(-2J_\nu)\,\frac{\cosh(I_\nu-J_\nu)}{\cosh(I_\nu+J_\nu)}$$

$$=\frac{\tanh J_{-\mu}^\upsilon(\tilde{r})+\tanh I_\nu^\upsilon(r)\tanh L_\nu^\upsilon(r)}{1+\tanh I_\nu^\upsilon(r)\tanh J_{-\mu}^\upsilon(\tilde{r})\tanh L_\nu^\upsilon(r)}=\overline{J}_\nu, \tag{52}$$

$$\exp(-2J_\nu)\,\frac{\sinh(I_\nu-J_\nu)}{\cosh(I_\nu+J_\nu)}$$

$$=\frac{\tanh L_\nu^\upsilon(r)+\tanh I_\nu^\upsilon(r)\tanh J_{-\mu}^\upsilon(\tilde{r})}{1+\tanh I_\nu^\upsilon(r)\tanh J_{-\mu}^\upsilon(\tilde{r})\tanh L_\nu^\upsilon(r)}=\overline{L}_\nu, \tag{53}$$

and one that relates the partition functions:

$$Z_{AT}^{(p_\nu,q_\nu)}=\frac{1}{2}\sum_{\tilde{q}_x,\tilde{q}_y=0}^{1}\exp[i\pi(q_x\tilde{q}_y-q_y\tilde{q}_x)]Z_{2I}^{(p_x,p_y,\tilde{q}_x,\tilde{q}_y)}. \tag{54}$$

It is easy to show that the partition function of the model with Hamiltonian (48) is related to the partition function of the inhomogeneous two-sublattice 8-vertex model:

$$Z_{\mathrm{IV}}^{(p_\nu,q_\nu)}=2Z_{8V}^{(p_x+q_x,p_y+q_y)}. \tag{55}$$

Indeed, as Fig. 2 shows (the edges of the lattice for the 8-vertex model are depicted by short dashed lines), this model is defined on a square lattice obliquely wound around a torus with toroidal boundary conditions (6) for the model (48) (in Fig. 2a the sublattices R and $\tilde{R}$ are 3×3, and spins situated on opposite edges of the sublattices must be assumed to be identical when the lattice is wound around the torus). The spin variables in the 8-vertex model are usually represented by arrows at the edges of the lattice, and take two values. For the sake of convenience, we denote them by Ising spin variables α that take values ± 1, located at sites in the middle of the corresponding edges. Then Fig. 2a clearly shows that in order to allow for the toroidal boundary conditions in the 8-vertex model on a lattice obliquely wound around the torus, it is convenient to represent the model as a square lattice of dimensions $2n\times 2m$ (the lattice being wound directly around the torus), with a lattice constant that is half the lattice constant of the original lattice R, on which the spins $\sigma(r)$ are specified. At the sites of such a lattice there are spin variables $\alpha_{k,l}=\pm 1$ satisfying the boundary conditions

$$\nabla_{x/2}^{(l_x)}\alpha_{2n,l}=(-1)^{l_x}\alpha_{1,l},\quad \nabla_{y/2}^{(l_y)}\alpha_{k,2m}=(-1)^{l_y}\alpha_{k,1},$$

where $l_x,l_y=0,1$ denote the boundary conditions, and the pairs (k,l), $k=1,\ldots,2n$, $l=1,\ldots,2m$ label the sites of the lattice. The relationship between the spin variables $\sigma(r)$, $\tau(\tilde{r})$, and $\alpha_{k,l}$ is shown in Fig. 2b and is specified by

$$\alpha_{k,l}=\sigma(r)\,\tau(\tilde{r}-\hat{x}),\quad \alpha_{k+1,l+1}=\sigma(r+\hat{y})\,\tau(\tilde{r}),$$

$$\alpha_{k,l+1}=\sigma(r+\hat{y})\,\tau(\tilde{r}-\hat{x}),\quad \alpha_{k+1,l}=\sigma(r)\,\tau(\tilde{r}). \tag{56}$$

We can therefore write the partition function of the 8-vertex model on such a lattice as

$$Z_{8V}^{(l_x,l_y)}=\sum_{[\alpha]}\ \exp(-\beta H_{8V}^{(l_x,l_y)})$$

$$=\sum_{[\alpha]}\prod_{i,j=1}^{n,m}\omega_1(\alpha_{2i-1,2j-1}\alpha_{2i,2j}|\alpha_{2i-1,2j},\alpha_{2i,2j-1})$$

$$\times\omega_2(\alpha_{2i,2j},\alpha_{2i+1,2j+1}|\alpha_{2i,2j+1},\alpha_{2i+1,2j}),\qquad(57)$$

where

$$\omega_1=\exp\{\tfrac{1}{2}[2P_y^v+I_y^v(\alpha_{2i-1,2j-1}\nabla_{x/2}^{(l_x)}\alpha_{2i-1,2j-1}$$
$$+\alpha_{2i-1,2j}\nabla_{x/2}^{(l_x)}\alpha_{2i-1,2j})$$
$$+J_{-x}^v(\alpha_{2i-1,2j-1}\nabla_{y/2}^{(l_y)}\alpha_{2i-1,2j-1}$$
$$+\alpha_{2i,2j-1}\nabla_{y/2}^{(l_y)}\alpha_{2i,2j-1})+L_y^v(\alpha_{2i-1,2j-1}\alpha_{2i,2j}$$
$$+\alpha_{2i-1,2j}\alpha_{2i,2j-1})]\}\delta(\alpha_{2i-1,2j-1}\alpha_{2i,2j}\alpha_{2i-1,2j}$$
$$\times\alpha_{2i,2j-1}-1),$$

$$\omega_2=\exp\{\tfrac{1}{2}[2P_x^v+I_x^v(\alpha_{2i,2j}\nabla_{y/2}^{(l_y)}\alpha_{2i,2j}$$
$$+\alpha_{2i+1,2j}\nabla_{y/2}^{(l_y)}\alpha_{2i+1,2j})+J_{-y}^v(\alpha_{2i,2j}\nabla_{x/2}^{(l_x)}\alpha_{2i,2j}$$
$$+\alpha_{2i,2j+1}\nabla_{x/2}^{(l_x)}\alpha_{2i,2j+1})+L_x^v(\alpha_{2i-1,2j-1}\alpha_{2i,2j}$$
$$+\alpha_{2i-1,2j}\alpha_{2i,2j-1})]\}\delta(\alpha_{2i,2j}\alpha_{2i+1,2j+1}\alpha_{2i+1,2j}$$
$$\times\alpha_{2j,2j+1}-1).$$

Here ω_1 and ω_2 describe the statistical weights of the two sublattices in the 8-vertex model. Obviously, at

$$P_x^v=P_y^v=P^v,\quad I_x^v=J_{-x}^v,\quad I_y^v=J_{-y}^v,\quad L_x^v=L_y^v=L^v\qquad(58)$$

the statistical weights ω_1 and ω_2 coincide, and we have the usual model.[10] For the sake of comparison we show the relationship between ω_1 and the usual vertex notation for the statistical weights in the 8-vertex model:

$$a=\omega_1(++|++)=\exp[P^v(r)+I_x^v(r)+I_y^v(r)+L^v(r)],$$
$$b=\omega_1(++|--)=\exp[P^v(r)-I_x^v(r)-I_y^v(r)+L^v(r)],$$
$$c=\omega_1(+-|-+)=\exp[P^v(r)+I_x^v(r)-I_y^v(r)-L^v(r)],$$
$$d=\omega_1(+-|+-)=\exp[P^v(r)-I_x^v(r)+I_y^v(r)-L^v(r)].\qquad(59)$$

The relation between the boundary conditions on the spin variables $\sigma(r)$, $\tau(\tilde r)$, and $\alpha_{k,l}$ is given by (56), from which it follows that

$$\nabla_{x/2}^{(l_x)}\alpha_{k,l}=\nabla_x^{(p_x)}\sigma(r)\nabla_x^{(q_x)}\tau(\tilde r),$$
$$\nabla_{y/2}^{(l_y)}\alpha_{k,l}=\nabla_y^{(p_y)}\sigma(r)\nabla_y^{(q_y)}\tau(\tilde r).$$

We see that the boundary condition indices are related by

$$l_x=p_x+q_x,\quad l_y=p_y+q_y,\qquad(60)$$

where the sum is assumed to be mod 2. Bearing this in mind, we obtain (55). Now, allowing for (54), (55), and (60), we can easily obtain the exact relation that between the partition functions of the Ashkin–Teller model and the 8-vertex model with the lattice obliquely wound around the torus:

$$Z_{AT}^{(p_\nu,q_\nu)}[P,I,J,L]=\sum_{\tilde q_x,\tilde q_y=0}^{1}\exp[i\pi(q_x\tilde q_y-q_y\tilde q_x)]$$
$$\times Z_{8V}^{(p_x+\tilde q_x,p_y+\tilde q_y)}[P^v,I^v,J^v,L^v],$$

where the coupling constants are related by (50)–(53).

We now returen to deriving the duality relation for the Ashkin–Teller model, for which we must use (5) and carry out the duality transformations of the partition functions Z_σ and Z_s in (42) simultaneously. As a result we have

$$Z_{AT}^{(p_\nu,q_\nu)}=\frac{1}{4}\sum_{\tilde p_\nu,\tilde q_\nu=0}^{1}\exp[i\pi(p_x\tilde p_y-p_y\tilde p_x+q_x\tilde q_y-q_y\tilde q_x)]$$
$$\times\prod_{r,\nu}\{\alpha_\nu(r)\exp(P_\nu(r))\}$$
$$\times\sum_{[\tilde\tau]}\sum_{[\tilde\sigma]}\prod_{r,\nu}[(\sinh 2\tilde I'_\nu(\tilde r))^{-1/2}$$
$$\times\exp\{\tilde I'_\nu(\tilde r)\tilde\sigma(\tilde r)\nabla_\nu^{(p_\nu)}\tilde\sigma(\tilde r)\}]$$
$$\times\sum_{[\tilde s]}\prod_{r,\nu}[(\sinh 2\tilde J'_\alpha(\tilde r))^{-1/2}$$
$$\times\exp\{\tilde J'_\nu(\tilde r)\tilde s(\tilde r)\nabla_\nu^{(q_\nu)}\tilde s(\tilde r)\}],$$

where $I'_\nu(r)$ and $J'_\mu(r)$ have been defined in (43), and

$$\tanh\tilde J'_{-\mu}(\tilde r)=\exp[-2J'_\nu(r)],$$
$$\tanh\tilde I'_{-\mu}(\tilde r)=\exp[-2I'_\nu(r)].$$

Now, repeating the calculations used to derive (47), we can easily obtain the following representation for the partition function of the Ashkin–Teller model:

$$Z_{AT}^{(p_\nu,q_\nu)}=\frac{1}{4}\sum_{\tilde p_\nu,\tilde q_\nu=0}^{1}\exp[i\pi(p_x\tilde p_y-p_y\tilde p_x+q_x\tilde q_y-q_y\tilde q_x)]$$
$$\times\sum_{[\tilde\sigma],[\tilde s]}\prod_{r,\nu}\exp[P_\nu(r)+I_\nu(r)+J_\nu(r)+L_\nu(r)]$$
$$\times\{1+\exp[-2I_\nu(r)-2L_\nu(r)]\tilde\sigma(\tilde r)\nabla_{-\mu}^{(\tilde p_\mu)}\tilde\sigma(\tilde r)$$
$$+\exp[-2J_\nu(r)-2L_\nu(r)]\tilde s(\tilde r)\nabla_{-\mu}^{(\tilde q_\mu)}\tilde s(\tilde r)$$
$$+\exp[-2I_\nu(r)-2J_\nu(r)]\tilde\sigma(\tilde r)\nabla_{-\mu}^{(\tilde p_\mu)}\tilde\sigma(\tilde r)\tilde s(\tilde r)$$
$$\times\nabla_{-\mu}^{(\tilde q_\mu)}\tilde s(\tilde r)\}.\qquad(61)$$

Using the identity (46), we reduce the partition function of the dual model,

$$\tilde Z_{AT}^{(\tilde p_\nu,\tilde q_\nu)}=\sum_{[\tilde\sigma],[\tilde s]}\exp\Bigg\{\sum_{r,\nu}[\tilde P_\nu(\tilde r)$$
$$+\tilde I_\nu(\tilde r)\tilde\sigma(\tilde r)\nabla_\nu^{(\tilde p_\nu)}\tilde\sigma(\tilde r)$$
$$+\tilde J_\nu(\tilde r)\tilde s(\tilde r)\nabla_\nu^{(\tilde q_\nu)}\tilde s(\tilde r)$$

$$+\tilde{L}_\nu(\tilde{r})\tilde{\sigma}(\tilde{r})\nabla_\nu^{(\tilde{p}_\nu)}\tilde{\sigma}(\tilde{r})\tilde{s}(\tilde{r})\nabla_\nu^{(\tilde{q}_\nu)}\tilde{s}(\tilde{r})]\Big\},$$

to a form similar to (49), and comparing the coefficients of the products of dual spins with the coefficients of the corresponding products in (61), we can relate the original and dual coupling constants:

$$\exp[-2J_\mu(r)-2L_\mu(r)]$$
$$=\frac{\tanh\tilde{J}_{-\nu}(\tilde{r})+\tanh\tilde{I}_{-\nu}(\tilde{r})\tanh\tilde{L}_{-\nu}(\tilde{r})}{1+\tanh\tilde{J}_{-\nu}(\tilde{r})\tanh\tilde{I}_{-\nu}(\tilde{r})\tanh\tilde{L}_{-\nu}(\tilde{r})},$$

$$\exp[-2I_\mu(r)-2L_\mu(r)]$$
$$=\frac{\tanh\tilde{I}_{-\nu}(\tilde{r})+\tanh\tilde{J}_{-\nu}(\tilde{r})\tanh\tilde{L}_{-\nu}(\tilde{r})}{1+\tanh\tilde{J}_{-\nu}(\tilde{r})\tanh\tilde{I}_{-\nu}(\tilde{r})\tanh\tilde{L}_{-\nu}(\tilde{r})},$$

$$\exp[-2I_\mu(r)-2J_\mu(r)]$$
$$=\frac{\tanh\tilde{L}_{-\nu}(\tilde{r})+\tanh\tilde{I}_{-\nu}(\tilde{r})\tanh\tilde{J}_{-\nu}(\tilde{r})}{1+\tanh\tilde{J}_{-\nu}(\tilde{r})\tanh\tilde{I}_{-\nu}(\tilde{r})\tanh\tilde{L}_{-\nu}(\tilde{r})},$$

$$\frac{1}{2}\exp(P_\mu+I_\mu+J_\mu+L_\mu)$$
$$=\exp(\tilde{P}_{-\nu})(\cosh\tilde{I}_{-\nu}\cosh\tilde{J}_{-\nu}\cosh\tilde{L}_{-\nu}$$
$$+\sinh\tilde{I}_{-\nu}\sinh J_{-\nu}\sinh\tilde{L}_{-\nu}), \tag{62}$$

and we obtain the self-duality relation for the Ashkin–Teller model:

$$Z_{AT}^{(p_\nu,q_\nu)}[P,I,J,L]$$
$$=\frac{1}{4}\sum_{\tilde{p}_\mu,\tilde{q}_\mu=0}^{1}T_{\tilde{p}_x,\tilde{p}_y}^{p_x,p_y}T_{\tilde{q}_x,\tilde{q}_y}^{q_x,q_y}\tilde{Z}_{AT}^{(\tilde{p}_x,\tilde{p}_y,\tilde{q}_x,\tilde{q}_y)}[\tilde{P},\tilde{I},\tilde{J},\tilde{L}]. \tag{63}$$

The matrix $\hat{T}$ is defined in (5) and (27).

Using these results, we can easily obtain the duality relations for the 8-vertex model on a lattice wound obliquely around the torus. Indeed, as (38) and (48) show, the Hamiltonians of the Ashkin–Teller model and the two-sublattice Ising model with four-spin interaction have the same structure of the interaction of the spin variables. Such a duality relation for the latter model can be derived in the same way as for the Ashkin–Teller model. As a result we arrive at an expression similar to (63):

$$Z_{\mathrm{IV}}^{(p_\nu,q_\nu)}[P^v,I^v,J^v,L^v]$$
$$=\frac{1}{4}\sum_{\tilde{p}_\mu,\tilde{q}_\mu=0}^{1}T_{\tilde{p}_x,\tilde{p}_y}^{p_x,p_y}T_{\tilde{q}_x,\tilde{q}_y}^{q_x,q_y}$$
$$\times\tilde{Z}_{\mathrm{IV}}^{(\tilde{p}_x,\tilde{p}_y,\tilde{q}_x,\tilde{q}_y)}[\tilde{P}^v,\tilde{I}^v,\tilde{J}^v,\tilde{L}^v],$$

where the coupling constants of the original and dual models are related by (62). If we now use this equation and Eq. (55), which relates the partition functions of the 8-vertex model on a lattice wound obliquely around a torus and the two-sublattice Ising model, and the relationship (60) between the boundary condition indices in these models, we can easily obtain the duality relation for the 8-vertex model:

$$Z_{8V}^{(l_x,l_y)}[P^v,I^v,J^v,L^v]$$
$$=\frac{1}{4}\sum_{p_\mu,q_\mu=0}^{1}\delta(p_x+q_x-l_x)\,\delta(p_y+q_y-l_y)$$
$$\times\frac{1}{4}\sum_{\tilde{p}_\mu,\tilde{q}_\mu=0}^{1}T_{\tilde{p}_x,\tilde{p}_y}^{p_x,p_y}T_{\tilde{q}_x,\tilde{q}_y}^{q_x,q_y}$$
$$\times\tilde{Z}_{8V}^{(\tilde{p}_x+\tilde{q}_x,\tilde{p}_y+\tilde{q}_y)}[\tilde{P}^v,\tilde{I}^v,\tilde{J}^v,\tilde{L}^v], \tag{64}$$

where the partition functions of the original and dual models have the form (57), summation of the boundary condition indices is assumed to be mod 2, and the coupling constants are related by (62). Simple algebraic manipulations show that the relations (62) are the same as the duality relations of the 8-vertex model given in Ref. 10 if conditions (58) and (59) are met:

$$\tilde{a}=\frac{1}{2}(a+b+c+d),\quad \tilde{b}=\frac{1}{2}(a+b-c-d),$$
$$\tilde{c}=\frac{1}{2}(a-b+c-d),\quad \tilde{d}=\frac{1}{2}(a-b-c+d).$$

5. CONCLUSION

The duality relations obtained in the present paper for $Z(N)$-symmetric models on a square lattice wound around a torus are of a general topological origin. Hence the proposed method can be generalized to spin models with a continuous symmetry group—for example, the XY-model on a torus.

Note that the duality relations (64) were obtained for the inhomogeneous 8-vertex model. Using the method suggested in Ref. 4, we can cut the torus along its cycles and obtain the duality relations for the model on a plane with different boundary conditions on the plane's edges. Moreover, using (64), we can obtain the duality relations for the correlation functions of the 8-vertex model on a square lattice obliquely wound around the torus, relations that generalize those obtained in Ref. 13 for infinite lattices.

We thank M. A. Lashkevich, S. Z. Pakulyak, and S. Khoroshkin for useful discussions. One of us (V.N.Sh.) thanks A. A. Belavin for his support and A. Yu. Morozov for his hospitality at the Institute of Theoretical and Experimental Physics in Moscow, where this work was completed. This work was supported financially by the Ukrainian State Committee for Science and Technology (Project No. 2.5.1/051) and by the INTAS Program (Grant No. 93-1038).

*)E-mail: abugrij@gluk.apc.org

[1] H. A. Kramers and G. H. Wannier, Phys. Rev. **60**, 252 (1941).
[2] L. P. Kadanoff and H. Ceva, Phys. Rev. B **3**, 3918 (1971).
[3] A. I. Bugriĭ and V. N. Shadura, Zh. Éksp. Teor. Fiz. **109**, 1024 (1996) [JETP **82**, 552 (1996)].
[4] A. I. Bugriĭ and V. N. Shadura, JETP Lett. **63**, 384 (1996).

[5] V. L. Berezinskiĭ, Ph.D. dissertation in Physics and Mathematics [in Russian], ITF Akad. Nauk SSSR, Moscow (1971).
[6] J. Villain, J. Phys. C **36**, 581 (1975).
[7] Al. B. Zamolodchikov, Zh. Éksp. Teor. Fiz. **75**, 341 (1978) [Sov. Phys. JETP **48**, 168 (1978)].
[8] V. S. Dotsenko, Zh. Éksp. Teor. Fiz. **75**, 1083 (1978) [Sov. Phys. JETP **48**, 546 (1978)].
[9] R. Savit, Rev. Mod. Phys. **52**, 453 (1980).
[10] R. Baxter, *Exactly Solved Models in Statistical Mechanics*, Academic Press, London (1982).
[11] E. Fradkin, V. A. Huberman, and S. H. Shenker, Phys. Rev. B **18**, 4789 (1978).
[12] L. P. Kadanoff, Ann. Phys. (San Diego) **120**, 39 (1979).
[13] M. Yu. Lashkevich, Mod. Phys. Lett. B **10**, 101 (1996).

Translated by Eugene Yankovsky

Order parameter quantum fluctuations in a two-dimensional system of mesoscopic Josephson junctions

A. I. Belousov, S. A. Verzakov, and Yu. E. Lozovik*)

Institute of Spectroscopy, Russian Academy of Sciences, 142092 Troitsk, Moscow Region, Russia
(Submitted 18 June 1997)
Zh. Éksp. Teor. Fiz. **113**, 261–277 (January 1998)

The boson lattice Hubbard model is used to study the role of quantum fluctuations of the phase and local density of the superfluid component in establishing a global superconducting state for a system of mesoscopic Josephson junctions or grains. The quantum Monte Carlo method is used to calculate the density of the superfluid component and fluctuations in the number of particles at sites of the two-dimensional lattice for various average site occupation numbers n_0 (i.e., number of Cooper pairs per grain). For a system of strongly interacting bosons, the phase boundary of the ordered superconducting state lies above the corresponding boundary for its quasiclassical limit—the quantum XY-model—and approaches the latter as n_0 increases. When the boson interaction is weak in the boson Hubbard model (i.e., the quantum fluctuations of the phase are small), the relative fluctuations of the order parameter modulus are significant when $n_0<10$, while quantum fluctuations in the phase are significant when $n_0<8$; this determines the region of mesoscopic behavior of the system. Comparison of the results of numerical modeling with theoretical calculations show that mean-field theory yields a qualitatively correct estimate of the difference between the phase diagrams of the quantum XY-model and the Hubbard model. For a quantitative estimate of this difference the free energy and thermodynamic averages of the Hubbard model are expanded in powers of $1/n_0$ using the method of functional integration. © *1998 American Institute of Physics.* [S1063-7761(98)01801-0]

1. INTRODUCTION

The development of microlithography has led to considerable progress in the investigation of mesoscopic systems,[1] the development of single-electron devices,[2] and the elucidation of new fundamental concepts in the theory of superconducting phase transitions in quantum dots.[3] There is also much interest in the study of properties of extended systems made up of mesoscopic objects. Among these systems are, for example, arrays of mesoscopic Josephson junctions, systems of ultrafine grains, and superconducting helium in a porous medium.[4a]

As a rule, granular superconductors and systems of Josephson junctions are described by using various modifications of the quantum XY-model (see below, Eq. (2)). However, this description is only correct when fluctuations in the modulus of the superconducting or superfluid order parameters are not significant.[4b] In order to investigate the role of fluctuations in the modulus of the order parameter in mesoscopic systems it is necessary to use other models that are in a certain sense more general.

In this paper, a system of interacting bosons on a square $N\times N$ lattice with a distance d between lattice sites is used as such a generalized model. In this model one site corresponds to one helium-filled pore, etc. The Hamiltonian of the system is chosen in the form

$$\hat{H}=\frac{t}{2}\sum_{\langle i,j\rangle}(2a_i^\dagger a_i-a_i^\dagger a_j-a_j^\dagger a_i)+\frac{U}{2}\sum_i(a_i^\dagger a_i-n_0)^2. \tag{1}$$

Here the operators $a_i^\dagger$ (a_i) are operators for creation (annihilation) of a boson at site $i=1,\ldots,N^2$, and are subject to the usual Bose commutation relations. The first term in the Hamiltonian corresponds to the ''kinetic'' energy of the particles; the hopping energy t for a particle of mass m can be written $t=\hbar^2/md^2$. The summation $\Sigma_{i,j}$ is over all nonrepeating pairs $\langle i,j\rangle$ of adjacent lattice sites. The second term in Eq. (1) describes the interaction between particles, with characteristic energy $U>0$. Qualitatively this Hamiltonian can also approximately describe a system of mesoscopic superconducting grains; in this formulation of the problem we avoid the question of how an individual grain enters the superconducting state, a problem of interest in its own right (see, for example, Ref. 5 and the citations therein) which recalls the problem of superfluidity in atomic nuclei in many ways.

A system with Hamiltonian (1) has a rich phase diagram[6,7] containing a Mott insulator phase (at $T=0$),[6,8] a normal metal, and a superconducting phase. In this paper we are interested in system (1) for integer-valued occupation, i.e., where the average number of bosons per site $n_0=\langle a_i^\dagger a_i\rangle$ is a whole number. Under these conditions, at $T=0$ the boson Hubbard model belongs to the same universality class (see Refs. 6–9) as the quantum XY-model:

$$\hat{H}_{XY}=J\sum_{\langle i,j\rangle}(1-\cos(\varphi_i-\varphi_j))-\frac{U}{2}\sum_i\left(\frac{\partial}{\partial\varphi_i}\right)^2, \tag{2}$$

where the phases of the order parameter $\varphi_i\in[0,2\pi]$. At finite temperature, the requirement that the average occupation number $n_0=k$ be an integer is weakened. In this case, the behavior of the system will depend continuously on the av-

 1063-7761/98/010146-10$15.00

erage occupation number; its critical properties remain unchanged in a band $n_0 = k + \delta n_0$, whose width $2\delta n_0$ should decrease with decreasing temperature.

The state of system (2), which has superconducting and normal (metallic) phases at finite temperature, is described by two dimensionless control parameters: the temperature in units of the coupling constant $T = k_b T/J$, and the dimensionless quantum parameter $q = \sqrt{U/J}$, which determines the characteristic zero-point fluctuations of the phase. The corresponding dimensionless control parameters for the Hubbard model (1) when $J = tn_0$ are the dimensionless temperature $T = k_b T/tn_0$ and quantum parameter $q = \sqrt{U/tn_0}$.

Model (2) can be obtained from Eq. (1) by neglecting relative fluctuations in the modulus of the order parameter $\Phi_i = \Delta_i e^{j\varphi_i}$, which is correct (for $U \neq 0$) at high particle densities n_0 (see below). Thus, the lattice boson Hubbard model can actually be regarded as more general when the effects of ordering in systems of granular superconductors, thin films, etc. are being investigated. For this reason there is interest in first-principles studies of how quantum fluctuations in the order parameter affect the establishment of a global superconducting state and, in particular, comparison of the phase diagrams of the two model systems (1) and (2) in the $\{q, T\}$ plane.

Another question of considerable interest involves the possibility of observing the phenomenon of reentrant superconductivity in a system of type (1), that is, the appearance of disorder not only with increasing but also with decreasing temperature T within a certain range of the quantum parameter q. Within the framework of the quantum XY-model the possible existence of reentrant effects is associated with the accurate specification of the domain of definition of the order parameter phase;[10] elsewhere it involves the effects of leakage or inclusion of mutual capacitance of grains in the bulk.[11,12] In this paper we show that inclusion of fluctuations of the modulus within the framework of model (1) does not lead to reentrant behavior, at least over the range of control parameters that we investigate.

This paper is organized in the following way.

In Sec. 2 we compare the phase diagrams of systems (1) and (2) obtained in the mean-field approximation. By expanding the free energy, the mean-square fluctuations, and the density of the superfluid component for the boson Hubbard model in the parameter $1/n_0$, we obtain corrections to the analogous quantities calculated within the framework of the quantum XY-model and determine quantitative criteria for the possibility of separating fluctuations in the modulus of the order parameter from fluctuations in its phase.

In Sec. 3 we present a short description of the quantum Monte Carlo method used to simulate the boson lattice system and the quantum XY-model.

In Sec. 4 we discuss and compare the results of our theoretical investigation with numerical models.

2. THEORETICAL ESTIMATES OF SIMILARITY BETWEEN THE BOSON HUBBARD MODEL AND THE QUANTUM *XY*-MODEL

2.1. Mean-field approximation

The mean-field approximation[10] consists of replacing the original Hamiltonian of the system, which can be written in the form (see Eqs. (1) and (2))

$$\hat{H} = \sum_i \hat{H}_i^0 + \sum_{\langle i,j \rangle} \hat{V}_{ij}, \tag{3}$$

by the mean-field Hamiltonian, which depends on a variational parameter α:

$$\hat{H}_{mf}(\alpha) = \sum_i \hat{H}_i^0 + \alpha \sum_i \hat{V}_i, \tag{4}$$

where the mean-field operator $\hat{V}$ is governed by the condition

$$\frac{\mathrm{Tr}(V_{ij} \exp(-\beta \hat{H}_{mf}))}{\mathrm{Tr}(\exp(-\beta \hat{H}_{mf}))} \equiv \langle \hat{V}_{ij} \rangle_{mf} = \frac{1}{K} \langle \hat{V}_i \rangle_{mf} \langle \hat{V}_j \rangle_{mf}.$$

The constant K depends on the specific form of the interaction energies $\hat{V}_{ij}$, $\hat{V}_i$. The parameter α is determined by using the variational Gibbs–Bogolyubov principle from the requirement that the trial energy

$$F^t(\alpha) \equiv F_{mf}(\alpha) + \langle \hat{H} - \hat{H}_{mf}(\alpha) \rangle_{mf}$$

be a minimum. Vanishing of the variational parameter α indicates that the system is disordered, and in the mean-field approximation the curve $\alpha(q,T) = 0$ can be regarded as an estimate of the boundaries of the ordered phase.

Simple calculations show that at the boundary of the ordered state the following general relation holds:

$$1 = -\frac{z}{K} \int_0^\beta dt \langle \hat{V}(t) \hat{V}(0) \rangle_{mf}, \quad \hat{V}(t) = e^{t\hat{H}} \hat{V} e^{-t\hat{H}}, \tag{5}$$

where z is the number of nearest neighbors ($z = 4$ for a two-dimensional square lattice).

For the quantum XY-model with Hamiltonian (2), it is natural to choose

$$\hat{V}_i = -zJ \cos \varphi_i/2$$

where

$$K = -Jz^2/4, \quad \alpha = 2\langle \cos \varphi_i \rangle_{mf}.$$

The use of Eq. (5) yields the well-known equation[10] for the boundaries of the ordered phase $T = T_c(q)$:

$$\frac{q^2}{z} = \frac{\sum_{n=-\infty}^{\infty} [\exp(-q^2 n^2/2T) - \exp(-q^2 (n+1)^2/2T)]/(2n+1)}{\sum_{n=-\infty}^{\infty} \exp(-q^2 n^2/2T)}. \tag{6}$$

The solution of this equation is shown in Fig. 1 as a solid curve.

In order to apply the method described above to the lattice Hubbard model, it is necessary to treat the system within the framework of the grand canonical ensemble, where

$$\hat{H}_i^0=\frac{U}{2}\,(a_i^\dagger a_i-n_0)^2+\frac{zt}{2}\,a_i^\dagger a_i-\mu a_i^\dagger a_i,$$

$$V_{ij}=-\frac{t}{2}\,(a_i^\dagger a_j-a_j^\dagger a_i), \tag{7}$$

and μ is the chemical potential of the system. The choice

$$\hat{V}_i=-zt(a_i^\dagger+a_i)/4$$

leads to

$$K=-tz^2/4, \quad \alpha=\langle a_i^\dagger+a_i\rangle_{mf}.$$

In this case, Eq. (5) takes the form

$$\frac{q^2}{z}=\frac{\Sigma_{n=-n_0}^{\infty}(n+n_0+1)[\exp(-q^2(n-\eta)^2/2T)-\exp(-q^2(n+1-\eta)^2/2T)]/(2n+1-2\eta)}{n_0\Sigma_{n=-n_0}^{\infty}\exp(-q^2(n-\eta)^2/2T)},$$

$$\eta=\frac{\mu}{U}-\frac{z}{2q^2n_0}. \tag{8}$$

The chemical potential μ can be found from the equation

$$\sum_{n=-n_0}^{\infty} n\,\exp(-q^2(n-\eta)^2/2T)=0, \tag{9}$$

which corresponds to choosing the average number of particles at a site to be n_0. Note that in the limit $n_0\to\infty$ we have $\eta=0$, and it is not difficult to show that Eq. (8) reduces to Eq. (6).

Solving Eqs. (8) and (9) simultaneously, we obtain the estimate of the ordered phase boundary shown in Fig. 1. For comparison we show plots for the average occupation numbers $n_0=1,2,6$. It is clear from Fig. 1 that for any interaction energy U the mean-field theoretical estimate of the ordered state boundary for the Hubbard model lies above the corresponding boundary for the XY-model, and approaches the latter as the average site occupation number n_0 increases. Calculations in the mean-field approximation show that approximate agreement of the two phase diagrams is obtained for $n_0\geqslant 25$.

2.2. Computation of corrections by the method of functional integration

We now attempt to refine the qualitative estimate of the difference between phase diagrams of models (1) and (2) obtained in the mean-field approximation. For this we will use the method of functional integration.

The partition function of model (1) can be written in terms of a functional integral over the components of the Bose field Φ in the form[9]

$$Z_h=\mathrm{Tr}(e^{-S})=\int D(\Phi,\Phi^*)e^{-S(\Phi,\Phi^*)},$$

$$S(\Phi,\Phi^*)=\int_0^\beta\Bigg[\sum_i \dot{\Phi}_i\Phi_i^*+\frac{t}{2}\sum_{\langle i,j\rangle}(|\Phi_i|^2+|\Phi_j|^2-\Phi_i^*\Phi_j-\Phi_j^*\Phi_i)+\frac{U}{2}\sum_i(|\Phi_i|^2-n_0)^2\Bigg]d\tau.$$

$$\Phi_i=\Phi_i(\tau), \quad \Phi_i^*=\Phi_i^*(\tau). \tag{10}$$

Making the variable substitution $\Phi_i=\sqrt{n_0+\delta n_i}e^{j\varphi_i}$ in Eq. (10), where n_0 is an integer, and taking into account the periodicity of the functions $\Phi_i(\tau)$, $\Phi_i^*(\tau)$ with respect to imaginary time, we have

$$Z_h=\int D(\delta n,\varphi)e^{-S(\delta n,\varphi)},$$

$$S(\delta n,\varphi)=\int_0^\beta\Bigg[\frac{U}{2}\sum_i(\delta n_i)^2+\tilde{j}\sum_i\delta n_i\dot{\varphi}_i+tn_0\sum_{\langle i,j\rangle}\Bigg(1+\frac{\delta n_i+\delta n_j}{2n_0}-\sqrt{\Big(1+\frac{\delta n_i}{n_0}\Big)\Big(1+\frac{\delta n_j}{n_0}\Big)}\cos(\varphi_i-\varphi_j)\Bigg)\Bigg]d\tau. \tag{11}$$

Increasing the average number of particles per site while fixing $J=tn_0$ and U reduces Eq. (11) to the partition function for the quantum XY-model. In fact, as $n_0\to\infty$ we have

$$Z_h\to\int D(\delta n,\varphi)\exp\Bigg\{-\int_0^\beta\Bigg[\tilde{j}\sum_i\delta n_i\dot{\varphi}_i+J\sum_{\langle i,j\rangle}(1-\cos(\varphi_i-\varphi_j))+\frac{U}{2}\sum_i(\delta n_i)^2\Bigg]d\tau\Bigg\}$$

$$=Z_n\int D(\varphi)\exp\Bigg\{-\int_0^\beta\Bigg[\sum_i\frac{\dot{\varphi}_i^2}{2U}$$

$$+J\sum_{\langle i,j\rangle}(1-\cos(\varphi_i-\varphi_j))d\tau=Z_nZ_{XY}, \qquad (12)$$

where the factor Z_n results from integrating (11) over fluctuations in the order parameter modulus. In the current approximation, this can be done separately from the integration over phase. It is clear from Eq. (12) that the critical behavior of the system of interacting bosons on a lattice (in the limit $n_0\to\infty$) is determined by fluctuations of the phases φ_i, and coincides with the critical behavior of the quantum XY-model.

Since we are interested in how close the boson Hubbard model is to the quantum XY-model (12) for large but finite average occupation numbers n_0, we expand the free energy, density of superfluid component, and mean-square fluctuations in the number of particles for model (1) in powers of $1/n_0$ to second order. The corrections we obtain to the corresponding quantities for the quantum XY-model (12) can serve as a quantitative measure of the difference between the phase diagrams of the two systems.

The following relations hold for the free energy F, the mean-square fluctuations in particle number δn^2, and the fraction of superfluid component ν_s (the helicity modulus γ in the case of the XY-model):

$$F_h=F_{xy}-\frac{1}{\beta}\ln(Z_n)+\frac{1}{2n_0^2}F^{(2)}+\ldots,$$

$$\delta n_h^2=\delta n_{xy}^2+\frac{1}{2n_0^2}\Psi^{(2)}+\ldots$$

$$\nu_s=\gamma+\frac{1}{2n_0^2}\Gamma^{(2)}+\ldots \qquad (13)$$

The first-order corrections vanish by virtue of the invariance of Eq. (12) of the XY-model under ''time'' reversal.

After integrating over the variables δn_i and discretizing the ''time,'' we obtain expressions for the corrections $F^{(2)}$, $\Psi^{(2)}$, $\Gamma^{(2)}$ in Eq. (13) that are suitable for subsequent estimation by the trajectory quantum Monte Carlo method based on the quantum XY-model (2). For the correction to the free energy we have

$$\beta F^{(2)}\equiv\langle\widetilde{F}^{(2)}\rangle_{xy}=\left\langle\frac{1}{4q^2}\sum_{p=0}^{P-1}\sum_{\langle i,j\rangle}\left[\left(2-\frac{PT}{q^2}(\dot\varphi_i^p-\dot\varphi_j^p)^2\right)\cos(\varphi_i^p-\varphi_j^p)\right]+\frac{1}{4q^4}\sum_{p=0}^{P-1}\left(\sum_i\dot\varphi_i^p\square_i^p\right)^2-\frac{1}{4PTq^2}\sum_{p=0}^{P-1}\sum_i(\square_i^p)^2\right\rangle_{xy}. \qquad (14)$$

Once we have determined $\delta n_{xy}^2\equiv\langle\widetilde{\Psi}\rangle$ (see Eq. (26)), we obtain the following expression for the correction $\Psi^{(2)}$ to the mean-square fluctuations in particle number within the XY-model:

$$\Psi^{(2)}=\langle\widetilde{\Psi}\rangle_{xy}\langle\widetilde{F}^{(2)}\rangle_{xy}-\langle\widetilde{\Psi}\widetilde{F}^{(2)}\rangle_{xy}-\frac{4T}{q^2N^2}\langle\widetilde{F}^{(2)}\rangle_{xy}$$

$$\Psi(2)=\langle\widetilde{\Psi}$$

Introducing a quantity $\widetilde{\Gamma}$ such that $\gamma\equiv\langle\widetilde{\Gamma}\rangle_{xy}$ (see Eq. (23)), we have for the correction $\Gamma^{(2)}$

$$\Gamma^{(2)}=\langle\widetilde{\Gamma}\rangle_{xy}\langle\widetilde{F}\rangle_{xy}-\langle\widetilde{\Gamma}\widetilde{F}\rangle_{xy}-\frac{1}{4N^2Pq^2}\left\langle\sum_{p=0}^{P-1}\sum_i Y_i^p\square_i^p\right\rangle_{xy}$$
$$-\frac{T}{8N^2q^2}\left\langle\sum_{p=0}^{P-1}\sum_{\langle i,j\rangle}\left(2-\frac{PT}{q^2}(\dot\varphi_i^p-\dot\varphi_j^p)^2\right)X_{ij}\right\rangle_{xy}$$
$$+\frac{T}{4N^2q^2}\left\langle\left[\sum_{p=0}^{P-1}\sum_i\dot\varphi_i^pS_i^p\right]^2\right\rangle_{xy}$$
$$-\frac{1}{4N^2Pq^2}\left\langle\sum_{p=0}^{P-1}\sum_i(S_i^p)^2\right\rangle_{xy}$$
$$+\frac{T}{4N^2q^4}\left\langle\left[\sum_{p=0}^{P-1}\sum_i\dot\varphi_i^pY_i^p\right]\left[\sum_{p=0}^{P-1}\sum_i\dot\varphi_i^p\square_i^p\right]\right\rangle_{xy}$$

$$X_{ij}^p=\cos(\varphi_i^p-\varphi_j^p)-\frac{2}{PT}\sin(\varphi_i^p-\varphi_j^p)\times\sum_{p=0}^{P-1}\sum_{\langle i,j\rangle}\sin(\varphi_i^p-\varphi_j^p),$$

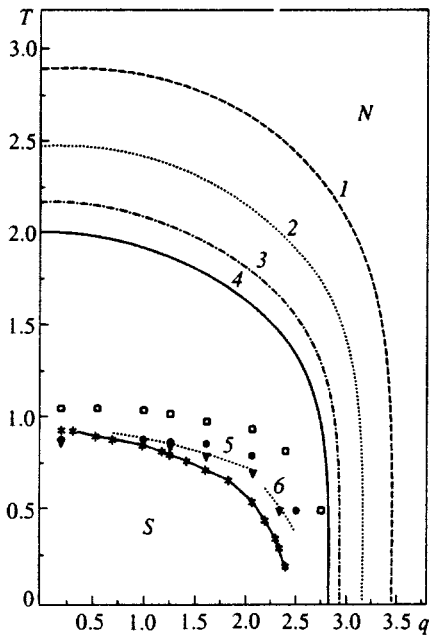


FIG. 1. Phase diagrams for the Hubbard model (1) and the quantum XY-model (2). S—superconducting state, N—normal state. Results of calculations based on mean-field theory: $n_0=1$ (*1*), 2 (*2*), 6 (*3*); *4*—quantum XY-model ($n_0=\infty$). Results of the $1/n_0$ expansion: $n_0=6$ (*5*), 14 (*6*). Here and in what follows the following symbols denote results of quantum Monte Carlo calculations: empty—$N=6$, filled in—$N=4$, with dots inside—$N=10$; squares—$n_0=1$, triangles—$n_0=2$, circles—$n_0=3$, rhombi—$n_0=4$, inverted triangles—$n_0=6$, stars—quantum XY-model.

$$Y_i^p = X_{ii+x}^p + X_{ii-x}^p + X_{ii+y}^p + X_{ii-y}^p,$$

$$S_i^p = \sin(\varphi_i^p - \varphi_{i+x}^p) + \sin(\varphi_i^p - \varphi_{i-x}^p) + \sin(\varphi_i^p - \varphi_{i+y}^p) + \sin(\varphi_i^p - \varphi_{i-y}^p). \quad (16)$$

Henceforth we use the notation

$$\dot{\varphi}_i^p = \lfloor \varphi_i^{p+1} - \varphi_i^p \rfloor_{[-\pi,\pi)}, \quad \varphi_i^P = \varphi_i^0,$$

$$\square_i^p = 4 - \cos(\varphi_i^p - \varphi_{i+x}^p) - \cos(\varphi_i^p - \varphi_{i-x}^p) - \cos(\varphi_i^p - \varphi_{i+y}^p) - \cos(\varphi_i^p - \varphi_{i-y}^p), \quad (17)$$

where $\lfloor f \rfloor_{[a,b)}$ denotes the reduction of the quantity f to the interval $[a,b)$.

Once we know the value of the coefficient $\Gamma^{(2)}$ in the expansion (13) for the fraction of superfluid component of the Hubbard model as a function of the control parameter, i.e., $\Gamma^{(2)} = \Gamma^{(2)}(q,T)$, we can estimate the correction to the curve $T_{xy}^c(q)$ for phase transitions of the quantum XY-model at large but finite average occupation numbers n_0. The corresponding estimate from above for the superconducting transition temperature $T_h^c(q;n_0)$ in the Hubbard model is given by the curve

$$T_h^c(q;n_0) \leqslant T_{xy}^c(q)\left(1 + \frac{\nu_s(q,T_{xy}^c) - \gamma(q,T_{xy}^c)}{\gamma(q,T_{xy}^c)}\right) = T_{xy}^c(q)\left(1 + \frac{\pi\Gamma^{(2)}(q,T_{xy}^c)}{4n_0^2 T_{xy}^c(q)}\right). \quad (18)$$

An estimate of Eq. (18) is easily obtained by assuming that on the curves $T_{xy}^c(q)$ and $T_h^c(q;n_0)$ of the topological Kosterlitz–Thouless phase transition, the following relations are satisfied for the ''universal discontinuity'' in the fraction of superfluid component[13]:

$$\gamma(q,T_{xy}^c) = 2T_{xy}^c/\pi, \quad \nu_s(q,T_h^c) = 2T_h^c/\pi.$$

Since $\Gamma^{(2)}(q,T)$ and $\gamma(q,T)$ are averages based on the quantum XY-model, they can easily be estimated from a quantum Monte Carlo calculation (see Sec. 3),[1)] which enables us to determine a family of phase diagrams for the boson Hubbard model Eq. (1) for various average boson numbers per grain n_0. The corresponding results are presented in Fig. 1. Estimates based on Eq. (18) show that in the range $0.7 < q < 1.5$, the phase transition curve for the Hubbard model reaches its limiting position—the phase transition curve for the quantum XY-model—for $n_0 = 8 \pm 1$. At larger values of the quantum parameter $q > 1.7$, the correction $\Gamma^{(2)}$ increases significantly, so that in this approximation the phase diagrams approximately coincide when $n_0 > 16 \pm 2$; however, direct numerical calculation of the model Eq. (1) leads to the conclusion (see below) that this estimate is much too large.

To conclude this section, it is necessary to make the following comment. It is easy to see that for all of these calculations, starting from representation (10) of the partition function of the boson Hubbard system within the grand canonical ensemble, the chemical potential equals zero. This enables us to make all of the required estimates analytically, avoiding the fact that the total number of particles is fixed in calculating the functional integrals over fluctuations in the order parameter modulus. In order to justify the assertion that the results of this section can be compared with the results of Monte Carlo calculations, it is necessary to show that the deviation of the average number of particles at a site calculated from n_0 is small compared to n_0 itself under the conditions mentioned above. To first order in $1/n_0$ this deviation has the form

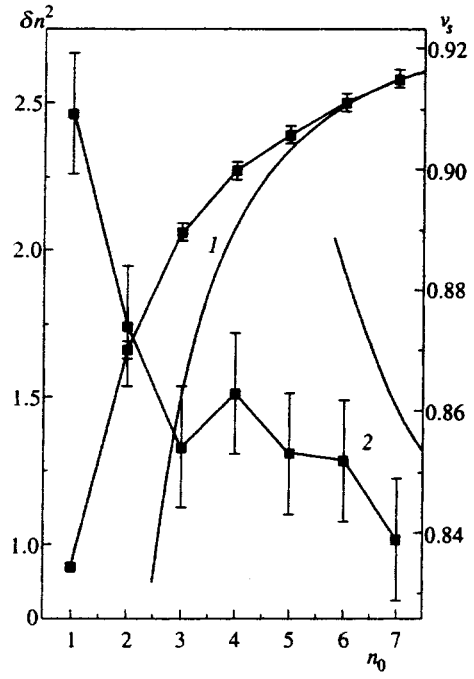


FIG. 2. Comparison of theoretical estimates (solid curves) for fluctuations in the order parameter modulus δn^2 (*1*) and fraction of superfluid component ν_s (*2*) obtained from the expansion (13) with the results of quantum Monte Carlo calculations. $q = 0.5$, $T = 0.6$. Statistical errors that are not shown were smaller than the size of the corresponding symbol.

$$\delta n = \frac{1}{n_0}\Delta^{(1)} + \ldots, \quad (19)$$

where we can use for the quantity $\Delta^{(1)}$ the following expression, which is convenient for subsequent estimation by the trajectory quantum Monte Carlo method within the quantum XY-model:

$$\Delta^{(1)} = \frac{T}{2N^2q^4}\left\langle \left[\sum_{p=0}^{P-1}\sum_i \dot{\varphi}_i^p\right]\left[\sum_{p=0}^{P-1}\sum_i \varphi_i^p \square_i^p\right] - \frac{q^2}{P}\sum_{p=0}^{P-1}\sum_i \square_i^p \right\rangle_{xy}. \quad (20)$$

Calculations show that in the range $q > 0.7$, $T < 1$, the quantity $\Delta^{(1)}$ does not exceed 0.08. Consequently, use of the grand canonical distribution with zero chemical potential is justified in deriving the coefficients of the expansion (13). As an illustration, Fig. 2 shows a comparison of the theoretical estimates for the mean-square fluctuations in particle number $\delta n_h^2(n_0)$ and the fraction of superfluid component $\nu_s(n_0)$ calculated using Eqs. (13)–(16) with the results of calculation by the quantum Monte Carlo method at the point

$\{q,T\}=\{0.5,0.6\}$, at which $\Delta^{(1)}=-0.5\pm0.1$. From the figure it is clear that for $n_0>5$ the experimental data obtained by using computer simulation are satisfactorily described by the expansion (13).

Conversely, in the region $q<0.4$, which corresponds to small fluctuations in phase and large fluctuations in the order parameter modulus, estimates based on Eq. (20) yield a larger deviation δn of the average number of bosons at lattice sites from n_0, and the approach described above is inappropriate.

3. THE QUANTUM MONTE CARLO METHOD. MEASURABLE QUANTITIES

The quantum trajectory Monte Carlo method is implemented by applying Trotter discretization, in which all thermodynamic averages of operators of a $D-$dimensional quantum system are estimated by using a classical $D+1$-dimensional quantum system generated by replicating the original system along the imaginary-time axis; the product of matrix elements of the high-temperature density matrix, which are calculated approximately, serves as the Boltzmann weight of a configuration of the corresponding classical system.[16] In this case, considerations of convenience dictate the choice of the representation in which the matrix elements are estimated, as well as the way the Hamiltonian is partitioned.

To investigate the properties of the Hubbard model (1) over the control parameter plane $\{q,T\}$, we used the quantum Monte Carlo method in its ''chessboard'' modification (a detailed explanation of Trotter discretization and the organization of Monte Carlo steps when a partition of the type discussed here is used can be found in Ref. 17). In this method, the degrees of freedom of the discretized system are the occupation numbers $\{n_i^p\}$ of sites of a three-dimensional $N\times N\times 4P$ lattice formed by a $4P$-fold replication of the original $N\times N$ lattice along the imaginary-time axis.

The properties of the quantum XY-model (2) are conveniently studied in a phase representation where the degrees of freedom of the discretized three-dimensional classical system are N^2P of the phases $\{\phi_i^p\}$.[18,19] To increase the efficiency of the calculation, we organized the Monte Carlo step by using a multigrid modification of the Metropolis procedure (see Ref. 19 and citations therein).

The Trotter discretization procedure can be used when the number P is large enough that all the thermodynamic averages calculated in the discretized system have errors of order $O(P^{-2})$. It is not difficult to show that the dimensionless parameters that determine the discretization error are the quantities $\epsilon_h=\max\{1/(T^2P^2n_0,q^2/(T^2P^2)\}$ for system (1) and $\epsilon_{xy}=q^2/(T^2P^2)$ for system (2). At each computation point $\{q,T\}$ the number of partitions P is chosen in such a way that $\epsilon_h,\epsilon_{xy}<0.05$. To monitor the convergence we compared the calculated results for various P. We found that the discretization error did not exceed 3% over the full range of parameters under study.

We studied the properties of this system along the lines $q=\text{const}$ and $T=\text{const}$, thereby moving from the ordered to the disordered phase. The results presented in this paper were obtained by averaging over 3–5 initial configurations formed by $4P$-fold replication (P-fold for the case of the XY-model) of an arbitrary configuration of bosons (or phases in the XY-model) on a $N\times N$ lattice.

3.1. Measurable quantities

Our attention was primarily focused on calculating the density of superfluid component ν_s. For the lattice Hubbard model this quantity can be found by two methods.

1) In terms of fluctuations of the ''winding number'' [7,16]:

$$\nu_s=\frac{T(W_x^2+W_y^2)_h}{2},$$

$$W_x=\sum_{p=0}^{4P}\sum_{i_y=1}^{N}(-1)^{i_x+p}n_i^p,$$

$$W_y=\sum_{p=0}^{4P}\sum_{i_x=1}^{N}(-1)^{i_y+p}n_i^p, \tag{21}$$

where n_i^p denotes the number of bosons at site i (with coordinates $\{i_x,i_y\}$) of layer p in the discretized classical system.

2) In terms of the correlation function for the paramagnetic current $\hat{j}^{(p)}$ (see Ref. 19):

$$\nu_s=-\frac{1}{n_0N^2}\langle\hat{T}_x\rangle_h-\frac{1}{n_0^2N^2TP}\sum_{\tau=0}^{P-1}\langle\hat{J}_x^{(p)}(\tau)\hat{J}_x^{(p)}(0)\rangle_h,$$

$$\hat{T}_x=-\frac{1}{2}\sum_i(a_{i+x}^\dagger a_i+a_i^\dagger a_{i+x}),$$

$$\hat{J}_x^{(p)}=-\frac{\tilde{j}}{2}\sum_i(a_{i+x}^\dagger a_i-a_i^\dagger a_{i+x}),$$

$$\hat{J}_x^{(p)}(\tau)=\exp(\tau\beta\hat{H}/P)\hat{J}_x^{(p)}\exp(-\tau\beta\hat{H}/P). \tag{22}$$

We found that the results of calculations based on Eqs. (21) and (22) coincided within the limits of statistical error over the entire range of control parameters we investigated.

Note that when the replacement $a_i\to\sqrt{n_0}e^{j\varphi}$ is made, Eq. (22) becomes the expression for the ''helicity modulus'' γ of the quantum XY-model[17]:

$$\gamma\equiv\langle\tilde{\Gamma}\rangle_{xy}=\frac{1}{N^2}\left\langle\sum_i\cos(\varphi_{i+x}^0-\varphi_i^0)\right\rangle_{xy}$$

$$-\frac{1}{N^2PT}\left\langle\sum_{p=0}^{P-1}\sum_{i,j}\sin(\varphi_{i+x}^0-\varphi_i^0)\sin(\varphi_{j+x}^p-\varphi_j^p)\right\rangle_{xy}$$

As the authors of Ref. 20 have noted, the temperature derivative of the density of superfluid component $\partial(\beta\nu_s)/\partial\beta$ yields additional information about the nature of the phase transition at temperature $T=T^c(q)$: within the framework of Kosterlitz–Thouless transition, the quantity $\partial(\beta\nu_s)/\partial\beta$ behaves like a Dirac delta-function $\delta(T-T^c)$ as the size of the system increases. Consequently, the position of the peak in the derivative of the density of superfluid component should not depend on the dimensionality of the system. This observation is very useful, because its computational complexity restricts our study of the Hubbard model to low-dimensional systems ($N\sim4-10$).

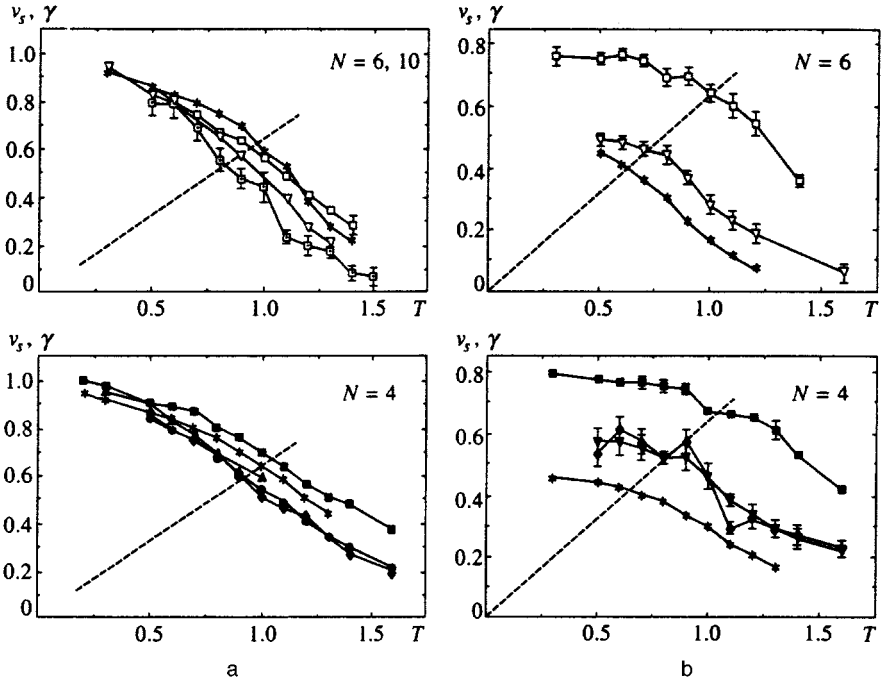


FIG. 3. a) Dependence of the fraction of superfluid component ν_s (helicity modulus γ in the case of the XY-model) on temperature T for $q=0.2$. The broken line shows the function $2T/\pi$. The curves are joined for convenience of representation. b) Plots of the fraction of superfluid component ν_s (helicity modulus γ for the case of the XY-model) as a function of temperature T for $q=2.0$.

To find the derivative $\partial(\beta\nu_s)/\partial\beta$ it is necessary to estimate the difference in average energies for systems that differ by a phase shift of $\pi/2$ (along one of the directions) at the boundary:

$$\frac{\beta(E(\pi/2)-E(0))}{n_0\beta t}\sim\nu_s+\beta\frac{\partial\nu_s}{\partial\beta}. \tag{24}$$

It can be shown that if the lattice bosons in the Hubbard model are assumed to be Cooper pairs with charge $2e$ within superconducting grains, this shift in phase is realized in a ''flux quantization'' scheme, where 1/8 of a flux quantum is emitted through the center of a torus on whose surface the system is located,[2)] leading to the appearance of a gauge phase for the Cooper pairs.[20]

One other observable quantity is the mean-square fluctuation of the number of bosons at a lattice site

$$\delta n_h^2=\frac{1}{4PN^2}\left\langle\sum_{p=0}^{4P-1}\sum_i(n_i^p-n_0)^2\right\rangle_h. \tag{25}$$

The corresponding quantity for the quantum XY-model is

$$\delta n_{xy}^2\equiv\langle\tilde{\Psi}\rangle_{xy}=\frac{T}{N^2q^2}\left\langle\sum_{p=0}^{P-1}\sum_i\left[1-\frac{PT}{q^2}(\dot{\varphi}_i^p)^2\right]\right\rangle_{xy}. \tag{26}$$

4. DISCUSSION OF RESULTS

Figure 3 shows the temperature dependence of the fraction of superfluid component $\nu_s(T)$ for the Hubbard model at values of the quantum parameter $q=0.2$ (Fig. 3a, classical region for the XY-model) and $q=2.0$ (Fig. 3b). For comparison, we also show the dependence of the helicity modulus (23) of the quantum XY-model in this figure. Analysis of data obtained for systems with various dimensions N and mean occupation numbers n_0 reveals that for the system of strongly interacting bosons ($q=2.0$), the results of numerical modeling are in qualitative agreement with the theoretical calculations in Sec. 2. In fact, it is clear from Fig. 1 and Fig. 3b that the boundary of the ordered superconducting phase for model (1) approaches the corresponding boundary for the quantum XY-model from the high-temperature side. The location of this boundary (i.e., the temperature $T^c(q)$ for a transition to the normal state) can be estimated from the universal jump relation for the fraction of superfluid component (the helicity modulus in the case of the XY-model): $\nu_s(T^c)=2T^c/\pi$. The Kosterlitz–Thouless transition temperature determined according to this universal discontinuity is in good agreement with the position of the peak in the temperature derivative of the superfluid density (24), which does not depend on the size of the system (within the limits of statistical errors in the calculations), and decreases with increasing average occupation number n_0, as is clear from Fig. 4.

Once we have calculated the helicity modulus and fraction of superfluid component for a system of finite size, we can obtain a more precise estimate of the topological phase transition temperature T^c by using the scaling procedure along with the Kosterlitz renormalization group.[21] Within the framework of this procedure we map the original system onto a two-dimensional Coulomb gas with an effective temperature $T_{CG}=T/(2\pi\nu_s^{(0)}(T)$, where $\nu_s^{(0)}(T)$ is the fraction of superfluid component in a region with no vortices (i.e., the fraction of local superfluid density). Thus, it is necessary to know $\nu_s^{(0)}$ as a function of the control parameters $\{q,T\}$ in order to extrapolate to a system of infinite size and determine the true transition temperature $T^c(q)$. This problem was solved in Ref. 22 for the two-dimensional classical XY-model, where the relation $\nu_s^{(0)}=\langle\cos(\varphi_i-\varphi_{i+x})\rangle$ holds. We may assume that this relation remains valid in the quantum XY-model for small enough values of q. However,

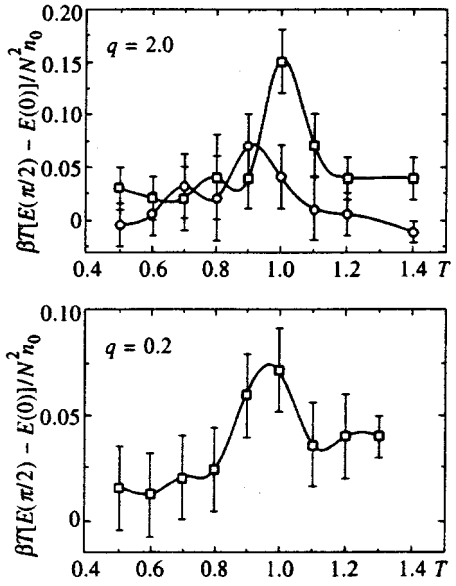


FIG. 4. Results of calculating the temperature derivative of the superfluid density (24) in the flux quantization method for $q=0.2$, 2.0.

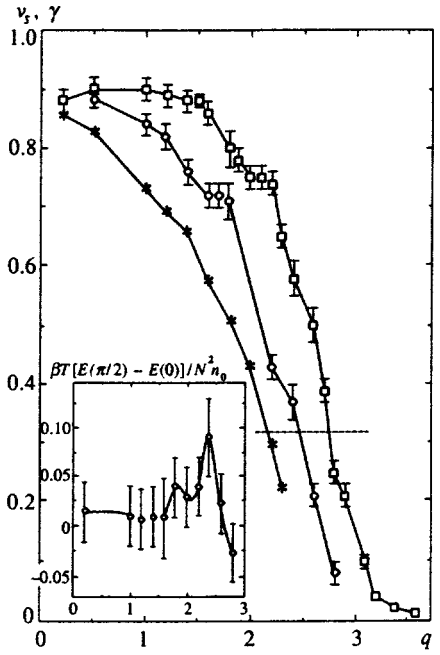


FIG. 5. Dependence of the fraction of superfluid component ν_s (helicity modulus γ in the case of the XY-model) on the value of the quantum parameter q for $T=0.5$. The broken line shows the line $1/\pi$. The inset shows the results of calculating the temperature derivative of the superfluid density (24) using the flux quantization method.

when $q \geqslant 1$, quantum fluctuations play a significant role, and this relation is incorrect, if only because simulations show that

$$\lim_{T\to 0} \nu_s/\nu_s^{(0)} = \lim_{T\to 0} \gamma/\langle \cos(\varphi_i - \varphi_{i+x})\rangle \neq 1.$$

Analogous considerations can be derived for the boson Hubbard model as well.

In using the scaling procedure described above for these quantum models, the customary approaches require that either a) the fraction of local superfluid density be set equal to unity[21]: $\nu_s^{(0)}(T)=1$, or b) it is postulated[7] that $\nu_s^{(0)} \times(T) = \lim_{T\to 0} \nu_s(T)$. Obviously, these two approaches are valid only at low temperatures $T \ll 1$. We therefore assume that it is more fundamental to estimate the position of the boundary for an ordered superconducting state either by using the universal discontinuity of the superfluid density (which yields fairly good results even for a system with dimensions $N>10$; see Refs. 17 and 18), or based on the position of the peak in its temperature derivative. Note that the latter method results in a much larger statistical error (see Figs. 3 and 4).

It is clear from Fig. 3a that the transition temperature to the superconducting state of a system of weakly interacting bosons ($q=0.2$) is lower than the ordering temperature of the quantum XY-model. This tendency is maintained as the size of the system increases. The theoretical approach used above does not help us to understand the reason for this behavior, because as we pointed out in Sec. 2.2, in the region $q<0.4$, where the interaction only weakly suppresses fluctuations of the order parameter modulus, the theoretical estimates (13)–(20) work badly, as evidenced by the large values of the corrections $\Psi^{(2)}$, $\Gamma^{(2)}$, and $\Delta^{(2)}$.

Figure 5 shows the results of calculating the fraction of superfluid component ν_s for the boson Hubbard model and the helicity modulus γ of the quantum XY-model as functions of the quantum parameter q for $T=0.5$, $N=6$. Having determined the phase transition point q^c according to the intersections of the functions $\nu_s(q)$ and $\gamma(q)$ with the line $2T/\pi = 1/\pi$, we find that the boundary of the superconducting state for the Hubbard model penetrates the large-q region and approaches the phase transition curve for the XY-model with increasing occupation number n_0.

The inset of Fig. 5 shows the results of calculating the temperature derivative of the superfluid density (24) obtained according to the ''flux quantization'' scheme. The position of the peak in the derivative ($q \approx 2.35$ for $n_0=3$) is in fairly good agreement with the critical point $q^c|_{n_0=3} \approx 2.4$ based on the universal jump in the superfluid density.

The dependence of the relative fluctuations in particle number at lattice sites of the system on temperature is shown in Figs. 6 and 7. In particular, Fig. 7 illustrates the important role played by interactions between bosons in the transition from the Hubbard model (1) to the quantum XY-model (2). In fact, for finite occupation numbers n_0 the spectrum of the operator $\hat{n}_i - n_0$ may be treated as unbounded if the relative fluctuations in particle number are small: $\delta n^2/n_0^2 \ll 1$. Then the particle number operator $\hat{n}_i - n_0$ may be assumed to be conjugate to the phase operator $\hat{\varphi}_i$, as is usually done in discussing Josephson and granular systems in terms of the model (2). Increasing the boson interaction (i.e., the quantum parameter q) leads to suppression of the relative fluctuations in the order parameter modulus, which is confirmed by Fig. 7. Note that for large q the fluctuations in the number of particles for the Hubbard model (25) are larger than the corresponding fluctuations for the quantum XY-model (26), and approach the latter as the occupation number n_0 increases.

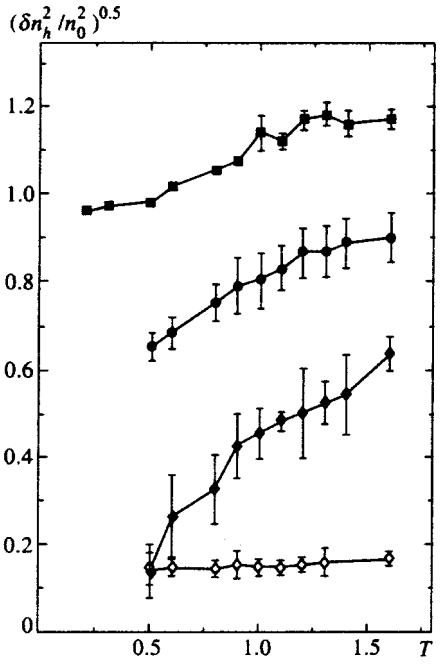


FIG. 6. Dependence of the relative fluctuations in the boson number at lattice sites on temperature T for $q=0.2$ ($N=4$, filled-in symbols) and $q=2.0$ ($N=6$, open symbols).

The inset to Fig. 7 shows the dependence of the relative fluctuations $\delta n_h^2/n_0^2$ on the quantum parameter q for $T=0.5$. The figure shows that increasing the interaction and occupation number n_0 significantly suppresses the relative fluctuations in the order parameter modulus.

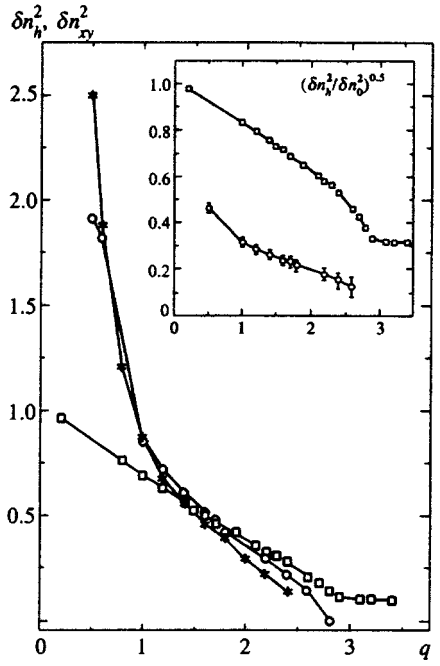


FIG. 7. Dependence of the mean-square fluctuations in the order parameter modulus on the value of the quantum parameter q for $T=0.5$. The inset shows the relative fluctuations in the modulus for various average occupation numbers n_0.

With increasing temperature T at a fixed value of the quantum parameter q, we observe an increase in the relative fluctuations in particle number (see Fig. 6) that is typical of a transition to the normal state (metallic in the case of the grain bulk). As is clear from the figure, at $q=0.2$ the occupation number $n_0=4$ is insufficient for appreciable suppression of the relative fluctuations in the order parameter modulus, but by $q=2.0$ we observe a sharp drop-off in the relative fluctuations due to the increasing role of the interaction.

In conclusion, we mention once more that at finite temperatures the critical behavior of our boson lattice Hubbard model coincides with the critical behavior of the quantum XY-model over a certain range of mean occupation number n_0 for the system sites (in contrast to the $T=0$ system). Consequently, the results of experiments on systems of superconducting grains on an insulating substrate should not differ qualitatively from experiments on a granular system within the framework of the grand canonical ensemble, where the chemical potential of a pair can be controlled, e.g., by applying a voltage to a conducting substrate.

We now summarize the basic results of this paper. Within the framework of the boson lattice Hubbard model we have analyzed the effect of quantum fluctuations in the phase and modulus of the order parameter on the character of ordering in two-dimensional mesoscopic Josephson and granular systems. Theoretical calculations based on the mean-field method and using the method of functional integration show that the models under consideration have identical critical behavior in the $\{q,T\}$ plane of dimensionless control parameters; the quantum XY-model (2) may be regarded as the semiclassical limit of the boson model (1) for $n_0 \gg 1$ and $U \neq 0$. In this case suppression of relative fluctuations in the modulus of the order parameter by the interaction enables us to treat the particle number operator $\hat{n}_i$ and phase $\hat{\varphi}_i$ as approximately canonically conjugate in this region: $\hat{n}_i - n_0 = \tilde{j}\partial/\partial\varphi_i$. Estimates arrived at using the quantum Monte Carlo method show that in the region $q<1$, which is ''classical'' for the quantum XY-model and a region of strong fluctuations in the order parameter modulus for the Hubbard model (1), the latter reduces to the quantum XY-model (2) for $n_0 \geqslant 10$, whereas in the region $q\sim 2$ of significant quantum fluctuation in the phase of the order parameter, the corresponding limit is reached by $n_0 \sim 8$. The effects of reentrant superfluidity (superconductivity) or low-temperature instability were not observed over the entire range $0.2<q<3.8$, $0.2<T<1.6$ of the control parameters we investigated.

This work was partially supported by a grant from the Russian Fund for Fundamental Research and the program ''Physics of Solid-State Nanostructures.''

*)E-mail: lozovik@isan.troitsk.ru

1)They are still easy to estimate theoretically, for example, when a different type of self-consistent approximation is used.[14]

2)Thus, in order to find the derivative of the density of superfluid component it is necessary to calculate the mean energy of system (1) in a field of the form $\mathbf{A}=(A,0,0)$. We have developed the algorithm necessary to implement this modification.

[1] Y. Imry and U. Sivan, Solid State Commun. **92**, 83 (1994).
[2] B. L. Altshuler and A. G. Aronov, in *Modern Problems in Condensed Matter Sciences*, V. M. Agranovich and A. A. Maradudin (eds.), Vol. 10, p. 1, North-Holland, Amsterdam (1985).
[3] A. F. Andreev, JETP Lett. **63**, 1018 (1996); JETP Lett. **64**, 664 (1996).
[4a] M. H. W. Chan, K. I. Blum, J. D. Reppy *et al.*, Phys. Rev. Lett. **61**, 1950 (1988); D. B. Haviland, Y. Liu, and A. M. Goldman, Phys. Rev. Lett. **62**, 2180 (1989).
[4b] C. Bruder, R. Fazio *et al.*, Phys. Scr. **42**, 159 (1992).
[5] A. L. Dobryakov, Yu. E. Lozovik *et al.*, Appl. Phys. A **54**, 100 (1992).
[6] M. C. Cha, M. P. A. Fisher, S. M. Girvin *et al.*, Phys. Rev. B **44**, 6883 (1991).
[7] W. Krauth, N. Trivedi, and D. Ceperley, Phys. Rev. Lett. **67**, 2703 (1991); W. Krauth and N. Trivedi, Europhys. Lett. **14**, 627 (1991).
[8] V. A. Kashurnikov, A. V. Krasavits, and B. V. Svistunov, JETP Lett. **64**, 99 (1996).
[9] M. P. A. Fisher and G. Grinstein, Phys. Rev. Lett. **60**, 208 (1988); M. P. A. Fisher, P. B. Weichman, G. Grinstein, and D. S. Fisher, Phys. Rev. B **40**, 546 (1989).
[10] E. Simanek, Phys. Rev. B **23**, 5762 (1983); Phys. Rev. B **32**, 500 (1985).
[11] R. S. Fishman and D. Stroud, Phys. Rev. B **37**, 1499 (1988); E. Roddick and D. Stroud, Phys. Rev. B **48**, 16600 (1993).
[12] J. B. Kim and M. Y. Choi, Phys. Rev. B **52**, 3624 (1995).
[13] P. Minnhagen, Rev. Mod. Phys. **59**, 1001 (1987).
[14] Yu. E. Lozovik and S. G. Akopov, J. Phys. C **14**, L31 (1981); S. G. Akopov and Yu. E. Lozovik, J. Phys. C **15**, 4403 (1982); S. A. Verzakov and Yu. E. Lozovik, Fiz. Tverd. Tela (St. Petersburg) **39**, 818 (1997) [Phys. Solid State **39**, 724 (1997)].
[15] R. M. Fye, Phys. Rev. B **33**, 6271 (1986).
[16] A. Blaer and J. Han, Phys. Rev. A **46**, 3225 (1992).
[17] M. Jacobs, J. V. Jose, M. A. Novotny, and A. M. Goldman, Phys. Rev. B **38**, 4562 (1988).
[18] A. I. Belousov and Yu. E. Lozovik, Solid State Commun. **100**, 421 (1996); A. I. Belousov and Yu. E. Lozovik, Matem. Model. **9**(4), 39 (1997).
[19] G. G. Batrouni, B. Larson, R. T. Scalettar, J. Tobochnik, and J. Wang, Phys. Rev. B **48**, 9628 (1993).
[20] F. F. Assad, W. Hanke, and D. J. Scalapino, Phys. Rev. B **50**, 12835 (1994).
[21] D. M. Ceperley and E. L. Pollock, Phys. Rev. B **39**, 2084 (1989).
[22] P. Olsson, Phys. Rev. B **52**, 4511 (1995).

Translated by Frank J. Crowne

Optical echo spectroscopy and phase relaxation of Nd^{3+} ions in CaF_2 crystals

T. T. Basiev, A. Ya. Karasik, and V. V. Fedorov

Institute of General Physics, Russian Academy of Sciences, 117942 Moscow, Russia

K. W. Ver Steeg

Microelectronics Research Center, Iowa State University, Ames, Iowa 50011, U.S.A.
(Submitted 18 June 1997)
Zh. Éksp. Teor. Fiz. **113**, 278–290 (January 1998)

Accumulated photon echoes have been used to investigate the mechanisms of optical dephasing in CaF_2 crystals activated by Nd^{3+} ions. Tunable picosecond laser radiation, which permits the selective excitation of various Nd^{3+} optical centers in the $^4I_{9/2}\rightarrow{}^4G_{5/2},{}^2G_{7/2}$ transition, is used. The optical phase relaxation times measured at temperatures from 9 to 50 K permit determination of the homogeneous widths of the transitions between the low-lying $^4I_{9/2}$ Stark level and three excited $^4G_{5/2},{}^2G_{7/2}$ levels, and calculation of the constants of the inter-Stark relaxation transitions in the ground and excited multiplets for the rhombic N and M Nd^{3+} centers in CaF_2 crystals. An analysis of the temperature dependence of the homogeneous linewidth of the transitions between low-lying Stark levels of the ground and excited states shows that the mechanism of optical dephasing in the crystals investigated is described well by direct relaxation processes with resonant inter-Stark absorption of one phonon in the ground and excited states. At $T=9$ K, the homogeneous linewidth Γ_h in CaF_2 crystals is almost an order of magnitude smaller than Γ_h in disordered CaF_2–YF_3 crystals. This difference can be attributed to the significantly greater spectral phonon density of states in disordered crystals.
 [S1063-7761(98)01901-5]

1. INTRODUCTION

The investigation of fundamental optical dephasing processes in organic and inorganic glasses and crystals activated by trivalent rare-earth ions has been the subject of numerous investigations,[1,2] since understanding the relaxation processes of excited states of impurity ions in a solid is not only a very important scientific problem, but also a necessary condition for creating efficient solid-state lasers. In the present work the optical phase relaxation of Nd^{3+} ions in CaF_2 crystals was investigated using accumulated photon echoes.[3,4]

The structure of the optical Nd^{3+} centers in CaF_2 crystals was fully investigated in earlier studies. In Refs. 5–7 the concentration method and selective laser excitation were successfully employed to elucidate the composition of the Nd^{3+} centers in CaF_2 and to determine the positions of the Stark sublevels for each type of center. In Refs. 8 and 9 the symmetry of the principal Nd^{3+} centers in CaF_2 crystals was investigated using ESR.

Accumulated photon echoes permit direct measurement of the phase relaxation kinetics and determination of the homogeneous linewidth Γ_h of a transition, which is not distorted by the inhomogeneous broadening of the spectra Γ_{ih}. In the present work we investigate the temperature dependence of the homogeneous linewidth of the $^4I_{9/2}\rightarrow{}^4G_{5/2},{}^2G_{7/2}$ transition in the dimeric and trimeric Nd^{3+} cluster centers in the temperature range 9–50 K. The experimental plot of $\Gamma_h(T)$ can be described within terms of direct relaxation transitions between Stark sublevels of the ground and excited states, with the absorption and emission of one phonon.

2. INHOMOGENEOUS SPLITTING AND BROADENING OF THE ABSORPTION LINES OF Nd^{3+} IN CaF_2 CRYSTALS

The CaF_2:Nd^{3+} crystals were grown by a modified Bridgeman method in a metered fluorine atmosphere in the absence of oxygen-bearing compounds.[5] The concentration of NdF_3 was varied from 0.1 to 7 wt. %. In the cubic fluorite structure Nd^{3+} ions can be distributed among sites of differing structure, which have different spectra.[5–7] According to Ref. 7, triply charged Nd^{3+} ions replace doubly charged Ca^{2+} ions during growth in a fluorinating atmosphere, and the crystal is saturated by interstitial fluorine ions to maintain electroneutrality. At very low Nd^{3+} concentrations, the charge compensation can be nonlocal (it can be effected by distant interstitial F_i^- ions) and can leave the original cubic symmetry of the ligand field around each Nd^{3+} ion unchanged. As the concentration of impurity Nd^{3+} ions and, therefore, of the F_i^- ions compensating them rises, they are attracted to one another to form dipoles, and the symmetry of the local environment of each Nd^{3+} ion changes from cubic to tetragonal (an L center forms, in which the F_i^- ion occupies a nearby interstitial site[7]). As the concentration is further increased, the mutual attraction of the Nd^{3+}–F_i^- dipoles leads to the formation of dimeric $(Nd^{3+}–F_i^-)_2$ clusters or M centers and tetrameric $(Nd^{3+}–F_i^-)_4$ clusters or N centers with rhombic symmetry. In these clusters two (or four) Nd^{3+} ions replace two (or four) Ca^{2+} ions. In this case two or four fluorine ions occupy nearby free interstitial sites to achieve local charge compensation.

The relative concentration of particular optical centers in CaF_2:Nd^{3+} crystals depends primarily on the total concen-

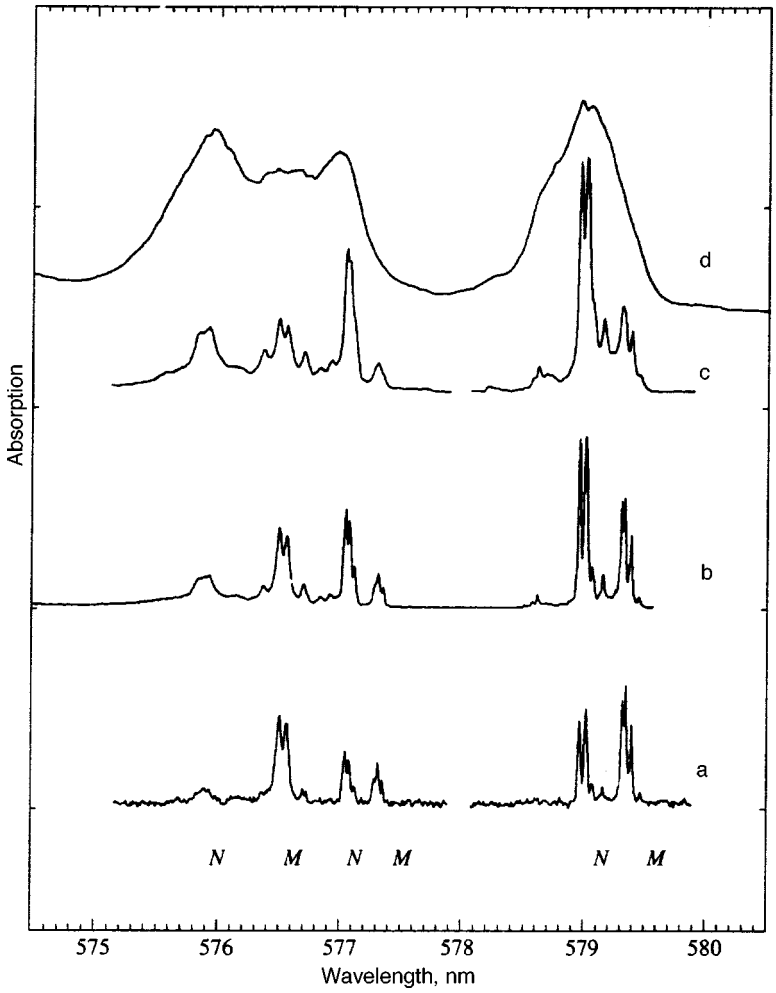


FIG. 1. Absorption spectra of the M and N Nd^{3+} centers in CaF_2 crystals in the $^4I_{9/2}\rightarrow{}^4G_{5/2},{}^2G_{7/2}$ transition at $T=9$ K for various concentrations of NdF_3, wt. %: a) 0.1, b) 0.3, c) 1, d) 7.

tration of NdF_3 introduced into the CaF_2 crystal during its synthesis. Figure 1 presents the absorption spectra of neodymium ions between the low-lying Stark sublevels of the $^4I_{9/2}\rightarrow{}^4G_{5/2},{}^2G_{7/2}$ transition in a CaF_2 crystal at $T=9$ K. According to Refs. 5–7, absorption at these wavelengths corresponds to a transition between the ground $^4I_{9/2}$ level and the first three Stark sublevels of the $^4G_{5/2},{}^2G_{7/2}$ multiplets of the M and N centers. It is clear from the figure that the absorption associated with M centers dominates at an NdF_3 concentration equal to 0.1%. Increasing the NdF_3 concentration from 0.1 to 1.0 wt. % results in an increase in the concentration of the tetrameric N centers in comparison with the dimeric M centers.[7] When the NdF_3 concentration is increased significantly (to more than 1 wt. %), the system should be treated as a crystalline solid solution. In this case NdF_3 is regarded not as an impurity, but as one of the components of the solid solution.

Solid solutions are characterized by statistical disorder, under which it becomes difficult to speak about order in distant coordination spheres. A significant increase in the concentration of NdF_3 to 7 wt. % is manifested by an increase in the number of centers, and passage from inhomogeneous line splitting to inhomogeneous broadening of the spectra (Fig. 1, curve d), where weakly structured broad bands appear instead of sets of narrow lines. In this case, the concentrations of the various components of the solid solution are already comparable, and the M and N cluster centers enlarge, possibly forming tetrameric, pentameric, and more aggregated clusters. The symmetry of the new clusters can be lower than that of the M and N cluster centers, and the statistics of the Nd^{3+} states in crystal fields of different symmetry and strength lead to considerable inhomogeneous broadening of the lines (comparable to glasses), amounting ~ 30 cm^{-1} (Fig. 1, curve d).

Figure 2 presents the absorption spectra of the CaF_2:Nd^{3+} crystals investigated in the $^4I_{9/2}\rightarrow{}^4G_{5/2},{}^2G_{7/2}$ transition between the low-lying Stark levels as the NdF_3 concentration is varied from 0.1 to 1 wt. % at $T=9$ K, which were recorded with a resolution of 0.22 cm^{-1}. As seen from the figure, the absorption spectra consist of two groups of lines with $\lambda=579.4$ and 579.0 nm. The minimum values of the linewidth in Fig. 2 are determined by the spectral resolution. It can be seen from the figure that varying the NdF_3 concentration from 0.1 to 1 wt. % results in variation of the specific concentrations of the M and N centers toward the more aggregated N centers, while the relative splitting of the absorption coefficients within each group of lines remains constant. In the figure, each of these groups corresponds to four lines with splitting for the M centers ranging from 0.9 to 4.3 cm^{-1} (between the outermost lines) and splitting for the N centers ranging from ~ 1.8 to 6 cm^{-1}. The lack of similar splitting in the absorption spectra in other transitions sug-

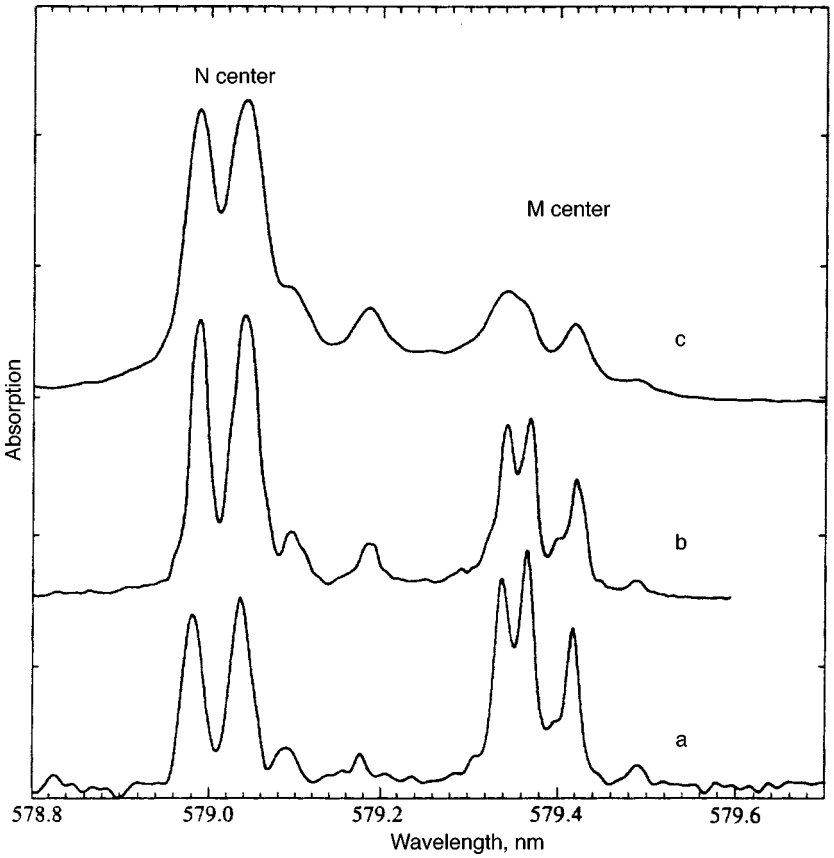


FIG. 2. Fragments of the absorption spectra of the M and N Nd^{3+} centers in CaF_2 crystals in the $^4I_{9/2} \rightarrow {}^4G_{5/2}, {}^2G_{7/2}$ transition at $T=9$ K for various concentrations of NdF_3, wt. %: a) 0.1, b) 0.3, c) 1.

gests that the complex structure of each group of lines is a consequence of the splitting of the $^4G_{5/2}, {}^2G_{7/2}$ excited state of the Nd^{3+} ion in the dimeric and tetrameric cluster centers. This conclusion is supported by Ref. 6, in which splitting of the low-lying levels of the $^4G_{5/2}, {}^2G_{7/2}$ multiplet of the Nd^{3+} ions in N and M centers was observed with a magnitude of the order of $\sim 1\ \mathrm{cm}^{-1}$ (the first three sublevels for the M centers and the first two for the N centers). In Ref. 10 we showed that the splitting of the levels can be attributed to a coherent interaction of paired Nd^{3+} ions in M and N cluster centers. Figure 3 presents energy diagrams of the Stark levels of an Nd^{3+} ion, which were obtained from the absorption spectra of the crystals investigated. The positions of the levels are consistent with the data presented in Refs. 5–7.

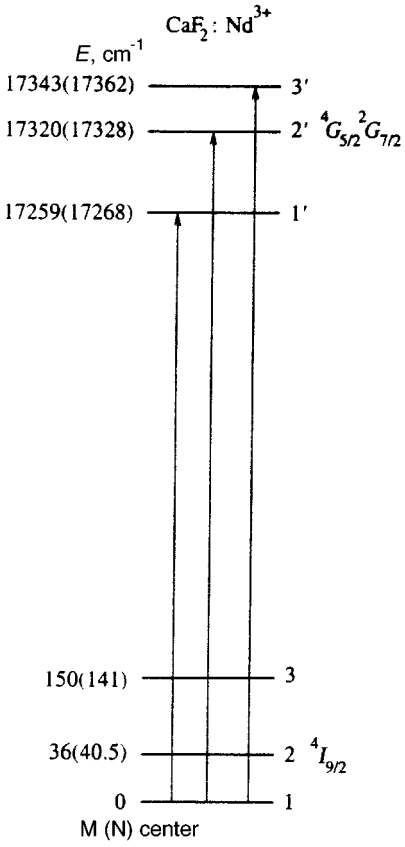


FIG. 3. Energy diagrams of the low-lying Stark levels of the $^4I_{9/2}$ and $^4G_{5/2}, {}^2G_{7/2}$ multiplets of Nd^{3+} ions in the M and N centers in a CaF_2 crystal.

3. ACCUMULATED PHOTON ECHOES IN CALCIUM FLUORIDE CRYSTALS

Accumulated photon echoes,[3,4] which are based on nonlinear resonant four-wave mixing, were used to measure the phase relaxation time T_2. The resonant excitation of Nd^{3+} ions was effected in the $^4I_{9/2} \rightarrow {}^4G_{5/2}, {}^2G_{7/2}$ transition using a Rhodamine 6G dye laser that is tunable in the range 560–600 nm. The dye laser was synchronously pumped by the second-harmonic emission ($\lambda_p = 532$ nm) of a YAG:Nd^{3+} laser operating in an active mode-locking regime with a frequency of 82 MHz. Compression of the pump laser pulses using a fiber-grating compressor was employed to shorten the duration of the output pulses of the dye laser from $\tau = 18$ to 0.5 ps. When the duration of the output pulses was $\tau = 18$ ps, the width of the spectrum of the pulses was

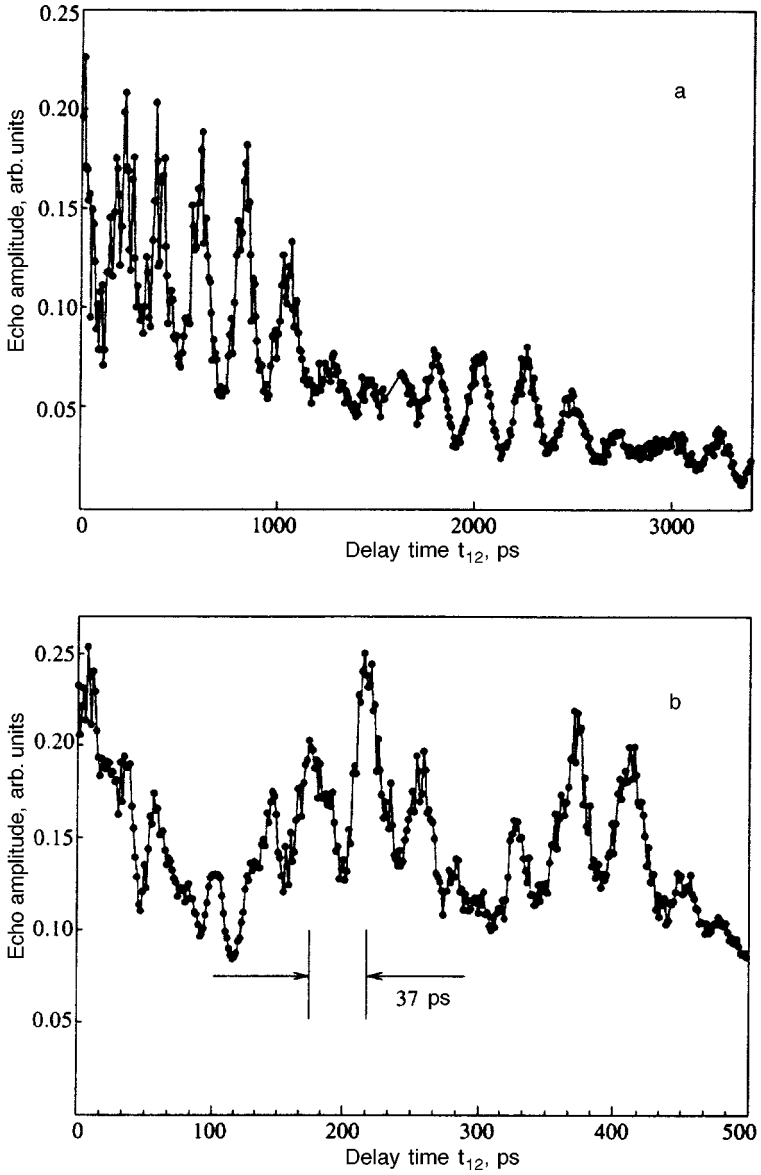


FIG. 4. Kinetics of the accumulated photon-echo signal (a) upon excitation of M centers ($\lambda = 579.36$ nm) in a CaF_2:Nd^{3+}(0.3 wt. %) crystal at $T = 9$ K; b) sample of the kinetics of accumulated photon echoes.

$\Delta\nu = 1\ \mathrm{cm}^{-1}$, and for $\tau = 0.5$ ps it was $\Delta\nu = 40\ \mathrm{cm}^{-1}$. The emission of the 20–100 mW dye laser was divided into orthogonally polarized pump beams and a probe beam in a 7:1 ratio. The probe pulses were delayed relative to the pump pulses by a time t_{12} ranging from 0 to 3.4 ns using an optical delay line. The pump beam was modulated by an acousto-optical modulator with a modulation frequency of 4 MHz. The pump beam and the probe beam then converged at a 1.5° angle and were focused onto the sample. The amplitude of the photon echo pulses was recorded in the direction of the probe beam using synchronous detection at the modulation frequency of the pump beam.

After excitation, the Nd^{3+} ions undergo rapid radiationless relaxation through several multiplets to the metastable $^4F_{3/2}$ level. The lifetime of the metastable $^4F_{3/2}$ level in the crystals investigated is long (hundreds of microseconds) in comparison with the time between pairs of exciting pulses (12 ns). This very relationship is responsible for accumulation of the echo signal. In the experiment, the amplitude of the accumulated photon-echo signal was investigated as a function of the delay between the probe pulses and pump pulses. In the experimental geometry, the echo signal decays according to $S \propto \exp(-2t_{12}/T_2)$, where t_{12} is the delay between the probe and pump pulses, and T_2 is the phase relaxation time ($T_2 = 1/(\pi\Gamma_h)$).[3]

Figure 4 presents the dependence of the accumulated photon-echo signal on the delay t_{12} for a CaF_2 crystal containing 0.3 wt. % NdF_3 at $T = 9$ K and an excitation wavelength $\lambda = 579.36$ nm. This wavelength falls between the two strongest absorption peaks assigned to the M centers (Fig. 2). As we see from the figure, the kinetics of the accumulated photon-echo signal exhibit pronounced amplitude modulation. The oscillations with a period of 37 ps correspond to the splitting between the absorption peaks of the M centers ($0.9\ \mathrm{cm}^{-1}$), which can be seen in the absorption spectrum (Fig. 2). Oscillations at a longer period correspond to smaller splitting values and cannot be seen in the absorption spectrum due to the inhomogeneous broadening of the lines and the limited spectral resolution of the monochromator. As we showed in Ref. 10, such modulation of the

TABLE I. Phase relaxation times T_2 and homogeneous linewidths $[\Gamma_h = (\pi T_2)^{-1}]$ of the ${}^4I_{9/2} \rightarrow {}^4G_{5/2}$, ${}^2G_{7/2}$ transitions of Nd^{3+} ions in the M and N centers in a CaF_2 crystal.

Transition (center)	T_2, ps	Γ_h, GHz	Wavelength, nm	T, K
$1 \rightarrow 1'(M)$	6300	0.05	579.43	9
$1 \rightarrow 1'(M)$	1200	0.27	579.43	18
$1 \rightarrow 1'(N)$	3500	0.09	579.09	9
$1 \rightarrow 1'(N)$	1000	0.32	579.09	18
$1 \rightarrow 2'(M)$	110	3.0	577.29	9
$1 \rightarrow 2'(N)$	150	2.1	577.00	9
$1 \rightarrow 3'(M)$	45	7.0	576.54	9
$1 \rightarrow 3'(N)$	30	10	575.91	9

accumulated photon-echo signal is due to fine splitting ($0.1-1\ cm^{-1}$) of the ground ${}^4I_{9/2}$ and excited ${}^4G_{5/2}, {}^2G_{7/2}$ multiplets as a result of the coherent exchange, magnetic dipole–dipole, and electric quadrupole–quadrupole interactions of the Nd^{3+} ions in the dimeric M and tetrameric N centers. The beat contrast in the kinetics depended on the excitation wavelength and decreased with increasing temperature or increasing concentration of the Nd^{3+} ions. At the same time, no variation of the decay time of the accumulated photon-echo signal was noted in the present experiments (for concentrations of NdF_3 equal to 0.3 and 1.0 wt. %). The decrease in the oscillation amplitude as the impurity-ion concentration increases might result from variation of the fine splittings in the centers that emerges as the inhomogeneous linewidth increases. Variation of the splittings leads to variation of the oscillation periods and, as a result of averaging, to a decrease in the oscillation amplitude. In addition, as can be seen from Fig. 4, the high-frequency modulation depth decreased as the delay t_{12} increased.

The echo kinetics measured in the experiment at $T=9$ K were not exponential. The reasons for this nonexponential behavior were discussed in Refs. 11 and 12, and might be associated with saturation effects at high laser pump power, as well as with the simultaneous excitation of several optical centers having different values of T_2. To minimize the influence of saturation effects, we selected a low radiated pump power of ~ 20 mW. In particular, we controlled and monitored the excitation selectivity of the centers and the transitions by varying the width of the laser excitation spectrum.

The influence of the optical density can also lead to distortion of the initial stage of decay of accumulated photon echoes.[13] It is difficult to determine the optical density of CaF_2 crystals, since the spectral width of the absorption line at $T=9$ K (Fig. 1) is less than the spectral resolution of the monochromator. Taking into account all the facts just cited, we measured the dephasing time T_2 in the final stage of decay of the accumulated photon-echo signal. In addition, to minimize the influence of oscillations and reduce the influence of optical density, the M and N centers were excited at the edge of the absorption band of each center (Fig. 2).

Table I lists T_2 and the linewidth Γ_h for transitions from the ground-state ${}^4I_{9/2}$ multiplet to the low-lying Stark sublevels of the ${}^4G_{5/2}, {}^2G_{7/2}$ multiplet. As can be seen from the table, there is a sharp increase in linewidth in the transitions to the high-lying Stark sublevels. For example, the linewidth of the transition to the third Stark sublevel (7 GHz for the M centers) is more than two orders of magnitude greater than the width of the line corresponding to the transition between the low-lying Stark sublevels (0.05 GHz for the M centers). In addition, a slight difference is observed between the values of Γ_h for the N and M centers.

The dependence of the decay time of the kinetics of the ${}^4I_{9/2} \rightarrow {}^4G_{5/2}, {}^2G_{7/2}$ transition between the low-lying Stark levels in a CaF_2:Nd^{3+} crystal was investigated in this work in the temperature range 9–50 K. Figures 5 and 6 show the experimental kinetics of the accumulated photon echoes for the excitation of M and N centers in a CaF_2 crystal (with excitation wavelengths $\lambda = 579.09$ nm and $\lambda = 579.43$ nm for the N and M centers, respectively) with a concentration of NdF_3 equal to 0.3 wt. % at $T=9$ (a) and 18 K (b). Measurements of the dephasing time and the values of the homogeneous width of the spectrum calculated from them are presented as a function of temperature for the N and M centers in CaF_2 in Fig. 7. As is seen from the figure, the widths of the N and M lines are similar over the measured temperature range, and increase monotonically with increasing temperature. At the same time, it should be noted that at $T=10$ K, the homogeneous linewidth Γ_h for the ${}^4I_{9/2} \rightarrow {}^4G_{5/2}, {}^2G_{7/2}$ transition of Nd^{3+} ions in CaF_2 crystals ($\Gamma_h \sim 0.05$ GHz) is almost an order of magnitude less than Γ_h in the disordered CaF_2–YF_3:Nd^{3+} crystals ($\Gamma_h \sim 0.05$ GHz) that we investigated in Ref. 12.

4. MECHANISM OF OPTICAL DEPHASING IN CALCIUM FLUORIDE CRYSTALS

To ascertain the mechanism of optical dephasing in these media, we analyzed the temperature dependence of the homogeneous linewidth (Fig. 7) in direct relaxational transitions between Stark sublevels involving the resonant absorption and emission of one phonon. In this approximation, the equation for the homogeneous linewidth $\Gamma_{11'}$ of the ${}^4I_{9/2} \rightarrow {}^4G_{5/2}, {}^2G_{7/2}$ transition between the low-lying Stark levels of the ground and excited multiplets (Fig. 6) can be written in the form[14]

$$\Gamma_{11'} = \frac{W_{12}^0}{\exp(\Delta E_{12}/kT)-1} + \frac{W_{13}^0}{\exp(\Delta E_{13}/kT)-1} + \frac{W_{1'2'}^0}{\exp(\Delta E_{1'2'}/kT)-1} + \frac{W_{1'3'}^0}{\exp(\Delta E_{1'3'}/kT)-1} + \frac{1}{2\pi T_1}, \quad (1)$$

where W_{12}^0, W_{13}^0, $W_{1'2'}^0$, and $W_{1'3'}^0$ are amplitude parameters that characterize the rates of radiationless relaxation between levels 1 and 2 and between levels 1 and 3 of the ground ${}^4I_{9/2}$ multiplet and between levels $1'$ and $2'$ and between levels $1'$ and $3'$ of the excited ${}^4G_{5/2}, {}^2G_{7/2}$ multiplet; ΔE_{ij} is the energy gap between the corresponding Stark sublevels (Fig. 3); and T_1 is the relaxation time of the population of the excited level. It was shown in Ref. 15 that the

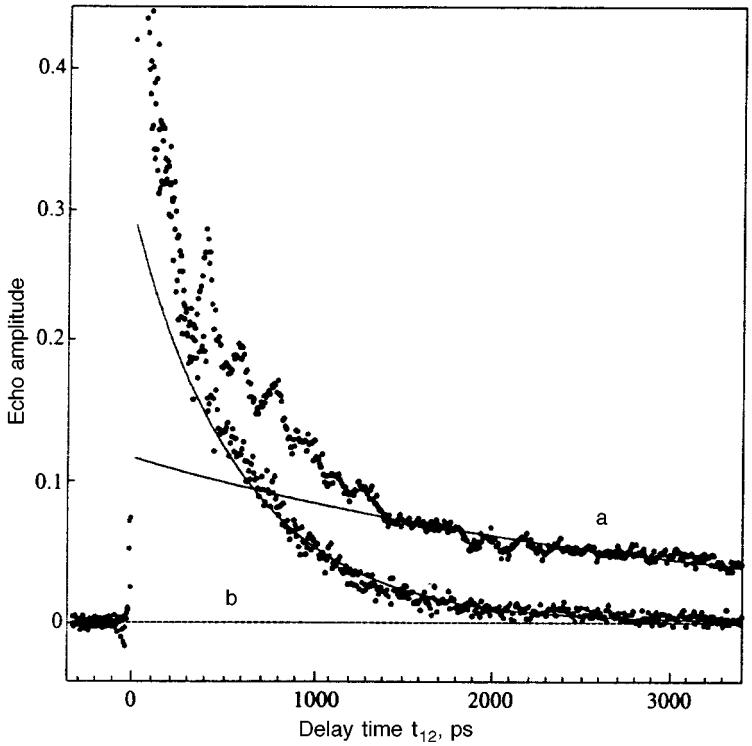


FIG. 5. Kinetics of the accumulated photon-echo signal upon excitation of the M centers ($\lambda = 579.43$ nm) in a CaF_2:Nd^{3+}(0.3 wt. %) crystal at $T = 9$ K (a) and $T = 18$ K (b). The solid lines correspond to exponential decay of the accumulated photon-echo signal; $T_2 = 6.3$ ns (a) and $T_2 = 1.2$ ns (b).

multiphonon relaxation rate $(2\pi T_1)^{-1}$ varies only slightly between liquid helium and liquid nitrogen temperatures. For this reason, we assume that the multiphonon relaxation rate $(2\pi T_1)^{-1}$ is the same for the first three Stark sublevels of the ${}^4G_{5/2}, {}^2G_{7/2}$ multiplet. Thus, in Eq. (1) the temperature-dependent terms describe only processes involving the absorption of a phonon.

We confined ourselves to transitions between three low-lying Stark sublevels, whose transition rates can dominate the relaxation process at $T < 50$ K. In analyzing Eq. (1) we also neglected transitions within the system of the fine-structure splitting of the M and N centers with $\Delta E < 1$ cm^{-1}, since the phonon density of states at energies $h\nu = \Delta E < 1$ cm^{-1} is low. While we only consider processes

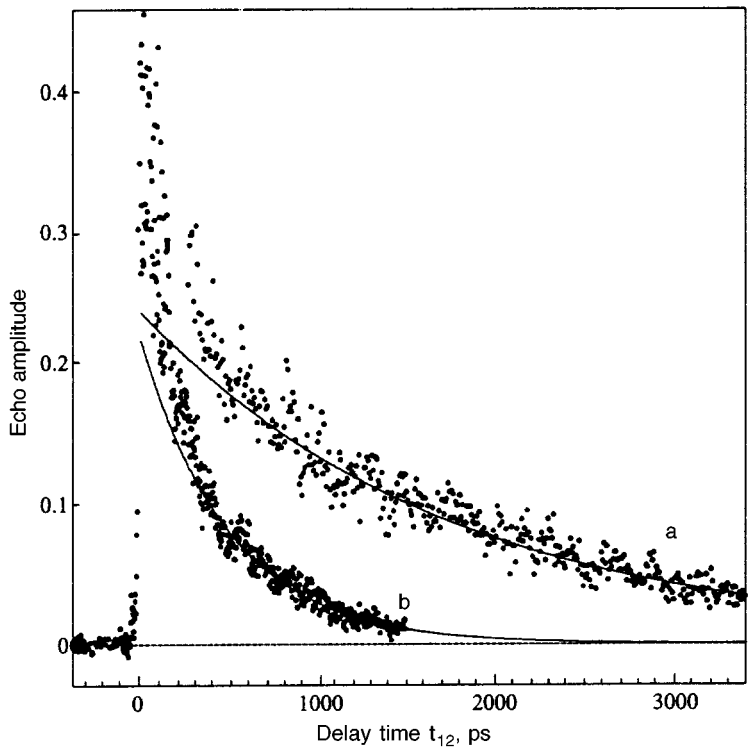


FIG. 6. Kinetics of the accumulated photon-echo signal upon excitation of the N centers ($\lambda = 579.09$ nm) in a CaF_2:Nd^{3+}(0.3 wt. %) crystal at $T = 9$ K (a) and $T = 18$ K (b). The solid lines correspond to exponential decay of the accumulated photon-echo signal; $T_2 = 3.5$ ns (a) and $T_2 = 1.0$ ns (b).

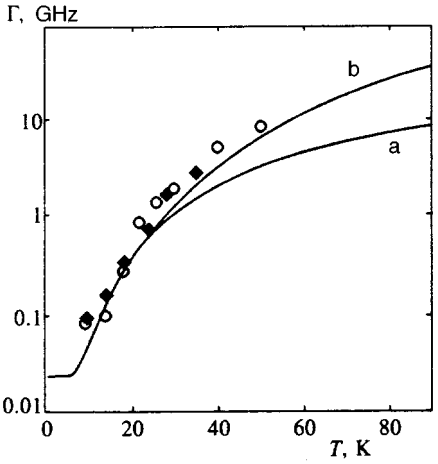


FIG. 7. Temperature dependence of the homogeneous linewidth Γ_h of the ${}^4I_{9/2}\rightarrow{}^4G_{5/2},{}^2G_{7/2}$ transition between low-lying Stark sublevels of the M and N Nd^{3+} centers (○ and ◆, respectively) in a CaF_2 crystal. The solid curves are theoretical plots obtained without (a) and with (b) consideration of W_{13}^0.

involving the absorption of a single phonon when a low-lying Stark sublevel of the ${}^4G_{5/2},{}^2G_{7/2}$ state is excited, we should take processes that entail phonon emission when high-lying Stark sublevels of this compound are excited into account. For example, for the $1\rightarrow 2'$ transition to the second Stark component, we can write

$$\Gamma_{12'}=\frac{W_{12}^0}{\exp(\Delta E_{12}/kT)-1}+\frac{W_{13}^0}{\exp(\Delta E_{13}/kT)-1}+\frac{W_{2'1'}^0\exp(\Delta E_{1'2'}/kT)}{\exp(\Delta E_{1'2'}/kT)-1}+\frac{W_{2'3'}^0}{\exp(\Delta E_{2'3'}/kT)-1}+\frac{1}{2\pi T_1}. \quad (2)$$

Here, along with terms that describe processes involving the absorption of a single phonon, the third term describes the spontaneous and stimulated emission of one phonon in the $2'-1'$ transition. The expression for the homogeneous linewidth $\Gamma_{13'}$ upon excitation of the third Stark component (the $1'-3'$ transition, Fig. 3) has the form

$$\Gamma_{13'}=\frac{W_{12}^0}{\exp(\Delta E_{12}/kT)-1}+\frac{W_{13}^0}{\exp(\Delta E_{13}/kT)-1}+\frac{W_{3'1'}^0\exp(\Delta E_{1'3'}/kT)}{\exp(\Delta E_{1'3'}/kT)-1}+\frac{W_{3'2'}^0\exp(\Delta E_{2'3'}/kT)}{\exp(\Delta E_{2'3'}/kT)-1}+\frac{1}{2\pi T_1}. \quad (3)$$

In Eqs. (1), (2), and (3), the energy gaps and the amplitude terms are symmetric under interchange of the level indices: $W_{ij}=W_{ji}$. This greatly simplifies the solution of the problem. Because of the exponentially small term $1/[\exp(\Delta E/kT)-1]$ at $T=9$ K ($kT=6.3$ cm^{-1}), Eqs. (2) and (3) can be simplified for Stark splittings $\Delta E>30$ cm^{-1} (Fig. 3). In particular, all terms except the third in (2) and the third and fourth in (3) can be neglected. In this approximation, $\Gamma_{12'}$ is equal to $W_{1'2'}^0$, since the ratio between the exponential terms in the third term in (2) for $T=9$ K and the energy gaps given is 1.

Hence, using the data in Table I, we at once obtain the radiationless relaxation constants $W_{1'2'}^0=3$ GHz (for the M centers) and $W_{1'2'}^0=2.1$ GHz (for the N centers). For this same reason, from (3) we obtain the expression $\Gamma_{13'}=W_{3'1'}^0+W_{3'2'}^0$. For subsequent evaluations we need the relationship between $W_{3'1'}^0$ and $W_{3'2'}^0$. According to the theory of electron–phonon interactions, W_{ij} is proportional to the product of the square of the matrix element of the electron–phonon interaction and the spectral phonon density of states. In the Debye approximation, the latter is proportional to the cube of the frequency of the phonon participating in the interaction. On this basis we can estimate the ratio $W_{3'2'}^0/W_{3'1'}^0=(\Delta E_{2'3'}/\Delta E_{1'3'})^3$, which is equal to 0.02 and 0.06 for the M and N centers, respectively. As a result of the evaluations made, in (3) we can neglect the term containing $W_{3'2'}^0$ and determine the constants $W_{3'1'}^0=\Gamma_{13'}$, which are 7 and 10 GHz for the M and N centers, respectively. The ratios between the resulting values are $W_{3'1'}^0/W_{2'1'}^0=2.3$ (4.3) for the M (N) centers. A comparison of these ratios with the ratio $(\Delta E_{3'1'}/\Delta E_{2'1'})^3=2.7$ (4.3) exhibits good agreement with the Debye approximation.

The significant difference between ΔE_{12} (~ 35 cm^{-1}) and ΔE_{13} (~ 150 cm^{-1}) for the ground-state ${}^4I_{9/2}$ multiplet causes the second term in (1) to be negligible in comparison with the first at $T<20$ K. It should be noted that the determination of W_{12}^0 from (1) at $T=9$ K is very sensitive to the accuracy of the measurements. First, the maximum delay between the pump pulses and probe pulses in the experimental system is $t_{12}=3.4$ ns, and this complicates the determination of a damping decrement of an accumulated photon-echo signal that is comparable to this time. Second, the values of T_1 obtained in Ref. 16 were 7.2 ns and 3.3 ns for the M and N centers in a CaF_2:Nd^{3+} crystal, respectively; therefore, a difference between two similar values (for example, $T_1(M)=7.2$ ns and $T_2(M)=6.3$ ns, see Table I) must be taken to calculate $W_{12}^0\sim(\Gamma_{11'}-(2\pi T_1)^{-1})$. Consequently, we used the results of kinetic measurements of the decay of the accumulated photon-echo signal at $T=18$ K, where $T_2(M)=1.2$ ns, to calculate W_{12}^0. The value of $\Gamma_{11'}$ in these experiments was 0.27 (0.30 GHz) for the M (N) centers. As a result, W_{12}^0 calculated from Eq. (1) was 3.4 and 5.9 GHz for the M and N centers, respectively.

In Fig. 7 (curve a) the solid line shows the dependence of $\Gamma(T)$ obtained from Eq. (1) for the M centers in a CaF_2:Nd^{3+} crystal with exact consideration of the transitions to the second and third Stark levels of the excited state and to the second level of the ground state, but without considering the relaxation processes associated with phonon absorption between levels 1 and 3 of the ground state ($W_{13}^0=0$). The figure reveals a sharp increase in the difference between the experimental points and the theoretical curve at $T>30$ K. This difference may be associated with neglect of relaxation processes between levels 1 and 3 of the ground state, which

becomes unjustified as the temperature rises. In order to take these processes into account, we determined W_{13}^0 by assuming that the phonon density of states has a Debye distribution. This assumption enables us to determine the relationship between W_{13}^0 and W_{12}^0 as $W_{13}^0 = W_{12}^0(\Delta E_{13}/\Delta E_{12})^3$ ($W_{13}^0 = 260$ (270) GHz for the M (N) centers). A theoretical curve that takes these parameters into account is depicted in Fig. 7 (curve b). The experimental values clearly correspond to a direct relaxational dephasing process in the temperature range considered.

Thus, the mechanism of optical dephasing in multicenter ordered crystals over the range $T = 10-50$ K is described well by a single-phonon resonant relaxation process. We previously showed that the temperature dependence of optical dephasing in disordered CaF_2-YF_3 crystals is also described by single-phonon processes.[12] The tenfold difference between the relaxation rate in disordered CaF_2-YF_3(10%):Nd^{3+} crystals with $\Gamma_h = 0.5$ GHz and in ordered CaF_2:Nd^{3+} crystals ($\Gamma_h = 0.05$ GHz) at $T = 9$ K can be associated with the difference between the probabilities W_{ij}^0 (Eq. (1)). In Ref. 17 we measured the absorption spectra of these crystals in the far infrared ($10-100\ \mathrm{cm}^{-1}$), and showed that the absorption coefficient in the frequency range corresponding to direct single-phonon transitions ($10-100\ \mathrm{cm}^{-1}$) increases by about an order of magnitude upon passage from an ordered crystal of fluorite, CaF_2, to disordered crystals of yttrofluorite, CaF_2-YF_3(12 wt. %). Assuming that there is a correlation between the IR absorption coefficient and the density of the phonon states participating in direct single-phonon transitions (which was previously demonstrated for amorphous semiconductors[18]), we can thus explain the indicated difference between the values of the probabilities W_{ij}^0, and therefore between Γ_h and the relaxation rates in ordered and disordered media.

5. CONCLUSIONS

Picosecond accumulated photon echoes have been used to investigate the temperature dependence of the homogeneous width of the spectrum of the $^4I_{9/2} \rightarrow {}^4G_{5/2}, {}^2G_{7/2}$ transition of Nd^{3+} ions in ordered CaF_2 crystals in the range 9–50 K with selective excitation of rhombic M and N centers. The homogeneous linewidths of the transitions between the low-lying Stark level of the ground $^4I_{9/2}$ multiplet and three levels of the excited $^4G_{5/2}, {}^2G_{7/2}$ multiplet have been measured, permitting calculation of the force constants of the relaxational single-phonon inter-Stark transitions in the ground and excited multiplets. An analysis of temperature dependence has shown that optical dephasing in the ground and excited states of the ordered crystals investigated is described well by direct relaxation processes that involve resonant inter-Stark absorption of one phonon. At $T = 9$ K, the homogeneous linewidth Γ_h in the disordered crystals is almost an order of magnitude greater than the analogous value in the ordered crystals. This difference can be accounted for by a significantly higher spectral phonon density of states in the disordered crystals.

This research was performed with support from the Russian Fund for Fundamental Research (Project 95-02-04328-a).

[1] R. M. Macfarlane and R. M. Shelby, in *Spectroscopy of Solids Containing Rare Earth Ions*, A. A. Kaplyanskii and R. M. Macfarlane (eds.), Elsevier, Amsterdam (1987), Vol. 21, p. 51.
[2] O. K. Alimov, T. T. Basiev, and S. B. Mirov, *Selective Laser Spectroscopy (Proceedings of the Institute of General Physics of the Academy of Sciences of the USSR, Vol. 9)* [in Russian], Nauka, Moscow (1987).
[3] W. H. Hesselink and D. A. Wiersma, Phys. Rev. Lett. **43**, 1991 (1979).
[4] H. de Vries and D. A. Wiersma, J. Chem. Phys. **80**, 657 (1984).
[5] Yu. K. Voron'ko, A. A. Kaminskiĭ, and V. V. Osiko, Zh. Éksp. Teor. Fiz. **49**, 420 (1965) [Sov. Phys. JETP **22**, 295 (1966)].
[6] T. P. J. Han, G. D. Jones, and R. W. Syme, Phys. Rev. B **47**, 14 706 (1993).
[7] V. V. Osiko, Yu. K. Voron'ko, and A. A. Sobol, in *Crystals: Growth, Properties and Applications, Vol. 10*, Springer-Verlag, Berlin–Heidelberg (1984), pp. 37–86.
[8] N. E. Kask, L. S. Kornienko, and E. G. Lariontsev, Fiz. Tverd. Tela (Leningrad) **8**, 2572 (1966) [Sov. Phys. Solid State **8**, 2058 (1967)].
[9] N. E. Kask, L. S. Kornienko, and M. Fakir, Fiz. Tverd. Tela (Leningrad) **6**, 549 (1964) [Sov. Phys. Solid State **6**, 430 (1964)].
[10] V. V. Fedorov, T. T. Basiev, A. Ya. Karasik *et al.*, in *Conference Handbook ICL '96*, Prague (1996), p. 12.
[11] K. W. Ver Steeg, A. Ya. Karasik, R. J. Reeves *et al.*, J. Lumin. **60/61**, 741 (1994).
[12] K. W. Ver Steeg, A. Ya. Karasik, R. J. Reeves *et al.*, Phys. Rev. B **51**, 6085 (1995).
[13] S. Saikan, H. Miyamoto, Y. Tosaki *et al.*, Phys. Rev. B **36**, 5074 (1987).
[14] *Laser Spectroscopy of Solids*, W. M. Yen and P. M. Selver (eds.), Springer-Verlag, Berlin (1981).
[15] T. T. Basiev, A. Yu. Dergachev, Yu. V. Orlovskiĭ *et al.*, in *Kinetic Laser-Fluorescence Spectroscopy of Laser Crystals* (*Proceedings of the Institute of General Physics of the Russian Academy of Sciences, Vol.* **46**) [in Russian], Nauka, Moscow (1994).
[16] Y. V. Orlovskii, S. B. Abalakin, and I. N. Vorob'ev, in *Abstracts of DPC-97: 11th International Conference on Dynamical Processes in Excited States of Solids*, Mittelberg, Austria–Germany (1997).
[17] A. Ya. Karasik, T. T. Basiev, A. A. Volkov *et al.*, in *Conference Handbook ICL '96*, Prague (1996), p. 2.
[18] *Light Scattering in Solids I (Topics in Applied Physics, Vol. 8)*, M. Cardona (ed.), Springer, Berlin (1975).

Translated by P. Shelnitz

Long-range correlations upon wave propagation in random media under the conditions of strong internal reflection from their boundaries

L. V. Korolev and D. B. Rogozkin

Moscow Engineering Physics Institute, 115409 Moscow, Russia
(Submitted 24 June 1997)
Zh. Éksp. Teor. Fiz. **113**, 291–312 (January 1998)

The large-scale behavior of the spatial distribution of radiation in a random medium is investigated under the assumption of strong internal reflection from its boundaries. The qualitative variations of the angular coherent backscattering spectrum and long-range spatial intensity correlations in the transmitted and reflected radiation fluxes are predicted. © *1998 American Institute of Physics.* [S1063-7761(98)02001-0]

1. INTRODUCTION

The long-range correlations of multiply scattered waves resulting from interference are manifested in numerous unusual effects, which are observed when electrons and light are transported in media with random inhomogeneities. They include weak localization of the electronic conductivity and other kinetic coefficients in metals and semiconductors,[1–3] universal fluctuations of the total conductivity of samples of small dimensions,[4,5] and enhanced backscattering and strong fluctuations upon propagation of coherent light in random media.[5–7]

One of the factors which strongly influences the distribution of radiation in a medium is the internal reflection of waves from its boundaries.[8,9] The appearance of a radiation flux reflected from a boundary can be caused by ordinary Fresnel reflection, which is a result of the difference between the refractive indices of the medium and a vacuum.[10,11] The limiting case of strong reflection is achieved when scatterers are situated in a specular resonator.[12] In either case internal reflection leads to the effective ''confinement'' of waves within a scattering layer. In addition, the internal reflection coefficient r becomes—along with the layer thickness L, the mean free path l, and the absorption length l_a—one more parameter that controls the dynamics of wave propagation in a medium.

The analysis of the influence of internal reflection on radiation transport in a scattering medium has been the subject of numerous publications.[9–21] This question has been considered within the ordinary theory of radiative transport,[9,11,21] and it has recently been treated in investigations of coherent phenomena.[10–20]

The problem of reflecting boundaries was addressed in the context of an analysis of the form of the angular coherent backscattering spectrum for the first time in Ref. 13. A new form of boundary conditions for the diffusion equation, in which internal reflection from the boundaries of the layer is taken into account qualitatively, was proposed in Ref. 13. An approach differing from the one in Ref. 13 to the description of wave propagation with internal reflection was proposed in Refs. 14 and 18. The radiation distribution within the medium was not considered in Refs. 14 and 18, and fluxes undergoing different numbers of reflections from the boundaries were summed at the very beginning. A correct derivation of the boundary conditions on the diffusion equation for the case of specular internal reflection was given in Ref. 10, where the correlation function of the scattered waves was calculated and it was shown that consideration of internal reflection significantly improves agreement between theory and experiment.

Various parameters of a scattering medium under the conditions of internal reflection have been determined experimentally using the approach in Ref. 10 (see Refs. 10, 15, 19, and 20). Long-range intensity correlations in the presence of strong internal reflection were considered in Refs. 16 and 17. The results in Refs. 16 and 17 were obtained by different methods. A comparison reveals that they contradict one another. The conclusions of both studies are based on incorrect starting relations for the intensity correlation functions, and should therefore be revised. This applies, in particular, to the frequency dependence of the correlations of the radiation fluxes transmitted through a layer.

In the present work the effects caused by the long-range correlations of multiply scattered waves, viz., coherent backscattering and long-range spatial intensity correlations, are considered under the conditions of strong internal reflection from the boundaries of a random layer. The influence of internal reflection on the distribution of radiation in a medium is investigated in detail, and it is shown that the wave-propagation regime is determined by the competition between the transmissive capacities of the boundaries and the random medium itself. The various approaches[10,14] to solving the problem of wave propagation in a medium in the presence of internal reflection are compared. Exact expressions for the diffusion asymptote of the distribution of radiation in a medium and for the angular dependence in the vicinity of the coherent backscattering peak are found in the limit of strong internal reflection. The qualitative variation of the dependence of the form of the peak on the thickness of the scattering sample is found. It is also shown that the ''confinement'' of radiation in the presence of strong internal reflection leads to significant alteration of the form of the fluctuation spectrum of the transmitted and reflected intensity. A maximum corresponding to reversal of the sign of the intensity correlation function appears in the spectrum. The asymptote of the frequency-dependent intensity correlations found here differs from those obtained in Refs. 16 and 17

 1063-7761/98/010164-13$15.00

and is described by a $(\Delta\omega)^{-3/2}$ law under the conditions of strong internal reflection.

The results obtained in the present work are of interest for investigations of the transport properties of dense random systems by optical methods.

2. INFLUENCE OF INTERNAL REFLECTION ON RADIATION PROPAGATION IN A RANDOM MEDIUM

As we know,[6,22,23] the angular distribution of radiation in the presence of coherent backscattering and the spectrum of long-range intensity correlations can be expressed in terms of ladder propagators, i.e., solutions of the linear transport equation:[8,24,25]

$$(\mathbf{\Omega}\cdot\nabla)I(\mathbf{r},\mathbf{\Omega})+\frac{1}{l_{\rm tot}}I(\mathbf{r},\mathbf{\Omega}) = \frac{1}{4\pi l}\int_{4\pi} I(\mathbf{r}',\mathbf{\Omega}')d\mathbf{\Omega}'+S(\mathbf{r},\mathbf{\Omega}), \tag{1}$$

where $I(\mathbf{r},\mathbf{\Omega})$ is the radiation intensity at the point $\mathbf{r}$ in the direction $\mathbf{\Omega}$, $S(\mathbf{r},\mathbf{\Omega})$ is the distribution of sources, $l_{\rm tot}^{-1}=(l_a^{-1}+l^{-1})$, l is the elastic mean free path, and l_a is the absorption length. For simplicity, in Eq. (1) we assume that single scattering is isotropic.

In the case of the internal reflection of radiation from the boundaries of a layer located between the $z=0$ and $z=L$ planes, the distribution of sources can be written in the form

$$S(\mathbf{r},\mathbf{\Omega})=\delta(\mathbf{r}-\mathbf{r}')\,\delta(\mathbf{\Omega}-\mathbf{\Omega}')+\delta(z)\int_{(\mathbf{n}\cdot\mathbf{\Omega}')>0} r(\mathbf{\Omega},\mathbf{\Omega}') \times(\mathbf{n}\cdot\mathbf{\Omega}')I(\boldsymbol{\rho},z=0,\mathbf{\Omega})d\mathbf{\Omega}'+\delta(L-z) \times\int_{(\mathbf{n}\cdot\mathbf{\Omega}')>0} r(\mathbf{\Omega},\mathbf{\Omega}')(\mathbf{n}\cdot\mathbf{\Omega}')I(\boldsymbol{\rho},z=L,\mathbf{\Omega}')d\mathbf{\Omega}', \tag{2}$$

where $r(\mathbf{\Omega},\mathbf{\Omega}')$ is the reflection coefficient of the surface, $\mathbf{n}$ is the exterior normal to a boundary of the layer, and $\boldsymbol{\rho}$ is the component of the vector $\mathbf{r}$ that is parallel to the boundaries of the layer.

The first term in (2) corresponds to a source of radiation of unit power in the bulk of the medium, and the second and third terms describe radiation reflected from the boundaries.

Finding the radiation intensity from Eq. (1) with source (2) and with an arbitrary reflection coefficient is a complicated mathematical problem.[8,9,21,25] An analytic solution can be obtained for it only in certain limiting cases, specifically in the case of diffuse (or Lambert) reflection (see Appendix A).

To elucidate the qualitative features that internal reflection imparts to the spatial distribution of radiation with the intention of subsequently considering long-range correlations (over distances much greater than the mean free path l), we use the simplest (i.e., diffusion) approximation in (1). Under the conditions of weak absorption ($l_a\gg l$) this approximation yields the correct qualitative description of the intensity on spatial scales exceeding l. The results obtained in the diffusion approximation will be refined later on in the limit of strong internal reflection, where the exact solution found under the assumption of diffusely reflecting boundaries can be utilized.

We go over to the diffusion approximation in Eq. (1) according to the usual procedure.[24,25] Under the assumption that the angular distribution of the radiation is weakly anisotropic, i.e.,

$$I(\mathbf{r},\mathbf{\Omega})=\frac{1}{4\pi}(F(\mathbf{r})+3\mathbf{\Omega}\cdot\mathbf{J}(\mathbf{r})), \tag{3}$$

where

$$F(\mathbf{r})=\int_{4\pi} I(\mathbf{r},\mathbf{\Omega}')\,d\mathbf{\Omega}', \quad \mathbf{J}(\mathbf{r})=\int_{4\pi}\mathbf{\Omega}'I(\mathbf{r},\mathbf{\Omega}')\,d\mathbf{\Omega}',$$

from Eq. (1) we obtain the relation $\mathbf{J}=-(l/3)\nabla F(\mathbf{r})$ and the diffusion equation for the radiation energy density $F(\mathbf{r})$ in the medium,

$$\Delta F(\mathbf{r})-\frac{1}{l_D^2}F(\mathbf{r})+\frac{3}{l}\delta(\mathbf{r}-\mathbf{r}')=0, \tag{4}$$

where $l_D=(ll_a/3)^{1/2}$ is the diffusion length.

The boundary conditions for the diffusion equation express the balance between the fluxes on a boundary of the medium.[8,9] With consideration of the surface part of the source (2) they have the form

$$\int_{(\mathbf{n}\cdot\mathbf{\Omega})<0} d\mathbf{\Omega}\,|\mathbf{n}\cdot\mathbf{\Omega}|I(\mathbf{r},\mathbf{\Omega})|_{z=0,L} = \int_{(\mathbf{n}\cdot\mathbf{\Omega})<0} d\mathbf{\Omega}\int_{(\mathbf{n}\cdot\mathbf{\Omega}')>0} d\mathbf{\Omega}' \times(\mathbf{n}\cdot\mathbf{\Omega}')r(\mathbf{\Omega},\mathbf{\Omega}')I(\mathbf{r},\mathbf{\Omega}')|_{z=0,L}\,. \tag{5}$$

Substituting (3) into (5), we obtain

$$\left(\frac{1}{2}F(\mathbf{r})+\frac{l}{3}\mathbf{n}\cdot\nabla F(\mathbf{r})\right)\Bigg|_{z=0,L} =\left(\frac{1}{2}r^{(1)}F(\mathbf{r})-\frac{l}{3}r^{(2)}\mathbf{n}\cdot\nabla F(\mathbf{r})\right)\Bigg|_{z=0,L}, \tag{6}$$

where

$$r^{(n)}=(n+1)\int_0^1 d\mu\int_0^1 d\mu'\mu'^n \times\int_0^{2\pi}\frac{d\varphi}{2\pi}\int_0^{2\pi}d\varphi' r(\mu,\mu',\varphi,\varphi'),$$

$\mu=|\Omega_z|$, $\mu'=|\Omega_z'|$, and φ and φ' are the azimuthal angles of the vectors $\mathbf{\Omega}$ and $\mathbf{\Omega}'$.

Equation (6) generalizes the usual boundary conditions for the diffusion equation[24,25] to the case of an arbitrary reflection coefficient.

If radiation is reflected specularly from the surface (for example, according to Fresnel's law), i.e., if

$$r(\mu,\mu',\varphi,\varphi')=r(\mu)\,\delta(\mu-\mu')\,\delta(\varphi-\varphi'), \tag{7}$$

the boundary conditions (6) take the form

$$F(z,\boldsymbol{\rho})+z_0 n_z \frac{\partial F(z,\boldsymbol{\rho})}{\partial z}\Big|_{z=0,L}=0, \tag{8}$$

where

$$z_0=\frac{2}{3}l\frac{1+r^{(2)}}{1-r^{(1)}}, \quad r^{(n)}=(n+1)\int_0^1 r(\mu)\mu^n d\mu. \tag{9}$$

Conditions (8) and (9) were proposed in Ref. 10.

If the boundary of the scattering medium is a diffusely reflecting, rough surface, then

$$r(\mu,\mu',\varphi,\varphi')=\frac{r}{\pi}\mu, \quad z_0=\frac{2}{3}l\frac{1+r}{1-r}, \tag{10}$$

where r is the albedo of the diffusely reflecting surface (the ratio of the total reflected flux to the incident flux).

As follows from (6) and (8)–(10), in the diffusion wave-propagation regime the boundary conditions are insensitive to the detailed character of internal reflection from the surface and depend only on the integrated characteristics of the reflection coefficient. In the case of strong internal reflection $(1-r^{(n)}\ll 1)$ Eqs. (9) and (10) yield $z_0\approx(4/3)l(/(1-r^{(1)})\gg l$. In the opposite limiting case $(r=0)$, $z_0=2l/3$ is the familiar value of the extrapolated length in the diffusion approximation.[24,25]

In Eq. (4) with boundary condition (8) it is convenient to take the Fourier transform with respect to the difference $\boldsymbol{\rho}-\boldsymbol{\rho}'$. Then this solution can be represented in the form

$$F(\mathbf{q},z,z')=\int d(\boldsymbol{\rho}-\boldsymbol{\rho}')\,\exp(-i\mathbf{q}\cdot(\boldsymbol{\rho}-\boldsymbol{\rho}'))F(\boldsymbol{\rho}-\boldsymbol{\rho}',z,z')$$
$$=\frac{3}{\gamma l}\frac{P((z+z'-|z-z'|)/2)P(L-(z+z'+|z-z'|)/2)}{[(\gamma z_0)^2+1]\sinh\,\gamma L+2\gamma z_0\,\cosh\,\gamma L}, \tag{11}$$

where $P(z)=\sinh\,\gamma z+\gamma z_0\cosh\,\gamma z$ and $\gamma^2=q^2+l/l_D^2$. In the case of a semi-infinite medium $(L\to\infty)$, it follows from (11) that

$$F(q,z,z')=\frac{3}{\gamma l}\frac{1}{1+\gamma z_0}P\left(\frac{z+z'-|z-z'|}{2}\right)$$
$$\times\exp\left(-\gamma\frac{z+z'+|z-z'|}{2}\right). \tag{12}$$

When there is weak internal reflection $(z_0\leqslant l)$, the result (11) transforms into the formula obtained without consideration of the reflecting boundaries (see, for example, Refs. 26 and 27):

$$F(q,z,z')=\frac{3}{\gamma l}\frac{1}{\sinh\,\gamma(L+2z_0)}P\left(\frac{z+z'-|z-z'|}{2}\right)$$
$$\times\exp\left(L-\frac{z+z'+|z-z'|}{2}\right), \tag{13}$$

where $P(z)=\sinh\,\gamma(z+z_0)$.

Under the conditions of strong internal reflection and large values of z_0 $(z_0\gg l)$, a new region $(z_0^{-1}\ll q\ll l^{-1})$ can be isolated in the behavior of (12), where the distribution of F behaves as in a medium with a perfectly reflecting surface:

$$F(q,z,z')=\frac{3}{\gamma l}\cosh\,\gamma\left(\frac{z+z'-|z-z'|}{2}\right)$$
$$\times\exp\left(-\gamma\frac{z+z'+|z-z'|}{2}\right). \tag{14}$$

From a physical standpoint, the transition to (14) describes the confinement of radiation in a medium with strong reflection from its boundaries. The probability of leaving through a boundary becomes appreciable only after many reflections. Therefore, close to the source $(\rho\sim q^{-1}<z_0)$, where the effective number of reflections from the boundary is small, the radiation distribution corresponds to an opaque boundary.

In the absence of absorption, Eq. (12) leads to the following spatial dependence of the density $F(z,z',\rho)$ near the boundary:

$$F(z\leqslant l,\ z'\leqslant l,\rho)=\frac{3}{2\pi l}\begin{cases}1/\rho, & l<\rho<z_0\\ z_0^2/\rho^3, & z_0<\rho\end{cases}. \tag{15}$$

According to (15), at small ρ the value of F does not depend on z_0 and is determined by the same expression as in a medium with a perfectly reflecting boundary. At large ρ the value of F is $(z_0/l)^2\gg 1$ times greater than in the absence of internal reflection.

In the case of a finite layer, new effects appear for large values of the reflection coefficient, for which the inequality $z_0\gg L$ holds. When $z_0\gg L$ under the assumption that there is no absorption, expression (11) can be written in the range of small spatial frequencies $q\ll L^{-1}$ in the following form:

$$F(q,z,z')\approx\frac{3z_0}{2l}\frac{1}{1+q^2z_0L/2}. \tag{16}$$

It can be seen at once from this formula that the long-range asymptotic behavior of the spatial distribution of the density is altered under the conditions of strong internal reflection. Whereas in the absence of reflection $(z_0=2l/3)$, the law $F(z,z',\rho)\approx\exp(-\pi\rho/L)$ follows from (11), in the limit $z_0\gg L$ relation (16) gives

$$F(z,z',\rho)\approx\frac{3}{lL}\frac{1}{2\pi}K_0\left(-\rho\sqrt{\frac{2}{z_0L}}\right). \tag{17}$$

Dependence (17) is a consequence of the confinement of radiation in the medium in the case of strong reflection from its boundaries.

Using (11), we can calculate the transmission coefficient of a radiation flux through a layer of a random medium in the usual manner (see, for example, Ref. 8). In the case of elastic scattering with normal incidence of the waves at the surface

$$T=\frac{l+z_0}{L+2z_0}. \tag{18}$$

The reflection coefficient is determined in this case from the flux conservation condition, $R+T=1$. According to (18), under the conditions of strong internal reflection $(z_0\gg L)$ the transmission and reflection coefficients are equal to one another: $T=R=1/2$. Under these conditions, radiation repeatedly crosses the scattering layer before leaving it, and ''forgets'' through which boundary it entered the medium.

It follows from (11)–(14) that in the diffusion approximation the structure of the expressions for the propagators has a universal character, and the reflectivity of the boundary influences only the value of z_0. In addition, under the conditions of strong internal reflection, $z_0 \simeq 4l/3(1-r^{(1)})$ does not depend on whether reflection from the boundary is specular or diffuse (compare (9) and (10)). Thus, under the conditions of strong internal reflection, where the scattered flux is reflected repeatedly from the boundary, the angular distribution of the radiation becomes nearly isotropic and ceases to be sensitive to the details of the interaction of the waves with the boundary of the medium. The same phenomenon was noted in Ref. 28, in which numerical calculations of resonant radiation transport in a flat resonator were performed.

In the case of diffuse (Lambert) reflection from the boundaries, in addition to the solution in the diffusion approximation (11) an exact analytic solution of the problem can be found (see Appendix A). If we compare it to solution (11), we can, on the one hand, evaluate the accuracy of the diffusion approximation, and, on the other hand, we can, more importantly, understand the mechanism of the influence of internal reflection on the distribution of radiation in a medium.

Unlike the diffusion solution (11), the rigorous solution of the transport equation obtained in Appendix A enables us to separate the contributions due to waves that do or do not undergo internal reflection from the boundaries, and to express the distribution of radiation in a medium with reflecting boundaries in terms of the solution of the transport equation in the absence of internal reflection. The latter can be found exactly, without any additional assumptions regarding the conditions on the boundaries (unlike, for example, the solution of the diffusion equation (4) with an approximate boundary condition like (6)).

In solutions (A15) and (A16), the contribution of the terms that appear because of internal reflection becomes significant when

$$1-r \ll R_q = 1 - \frac{4}{3}\gamma l \coth \gamma L \tag{19}$$

in the case of reflection from a semi-infinite medium (see (A15)), or when

$$1-r \ll T_q = \frac{4}{3}\frac{\gamma l}{\sinh \gamma L} \tag{20}$$

in the case of the passage of radiation through a finite layer (see (A16)), where R_q and T_q are the reflection and transmission coefficients (for a medium without reflecting boundaries) for the qth harmonic of the spatial distribution of the radiation. The expressions for R_q and T_q in (19) and (20) are valid for $\gamma l \ll 1$.

Thus, the mechanism for the influence of internal reflection on the wave-propagation regime in a medium is clearly seen from solutions (A15) and (A16). The regime is governed by the competition between the reflective (or transmissive) capacities of the boundaries and the random medium itself. In the absence of absorption, (19) is equivalent to the condition $qz_0 \gg 1$, under which the radiation propagation regime in a medium with a perfectly reflecting boundary is effectively realized, and (20) transforms into the inequality $z_0 \gg L$, under which the strong confinement of radiation in the resonator is observed.

The expressions for $F(z,z',q)$ in Appendix A have exactly the same structure as (11) and (12). In fact, if we substitute the corresponding quantities calculated in the thick-layer approximation (see, for example, Refs. 8 and 23) into (A15) and (A16), we arrive at formulas which coincide in form with (11) and (12). The differences are of the next order with respect to the small parameters γl and l/L, and are beyond the range of accuracy of the diffusion approximation, within which Eqs. (11) and (12) were derived.

3. COHERENT BACKSCATTERING

One of the most striking manifestations of the long-range correlations of waves in coherent transport in a random medium is enhanced backscattering.[6,7] The peak in the angular spectrum of the backscattered radiation appears as a consequence of the exact matching of the phases of waves propagating in the medium along identical trajectories, but in mutually opposite directions. The calculation of the coherent backscattering intensity reduces to summation of a series of maximally crossed or ''fan'' diagrams.[6,7,22]

The contribution of coherently scattered radiation to the angular spectrum can be written in the form

$$J_c(\mathbf{k},\mathbf{k}_0) = \frac{1}{4\pi l^2 A}\int d^3r \int d^3r' \Psi(\mathbf{r},-\mathbf{k})\Psi^*(\mathbf{r},\mathbf{k}_0)F \times(\mathbf{r},\mathbf{r}')\Psi^*(\mathbf{r}',-\mathbf{k})\Psi(\mathbf{r}',\mathbf{k}_0), \tag{21}$$

where $F(\mathbf{r},\mathbf{r}')$ is the radiation energy density (see Eq. (3)), $\Psi(\mathbf{r},\mathbf{k}_0)$ is the mean field within the medium when a plane wave with the wave vector $\mathbf{k}_0$ impinges on it, $\mathbf{k}_0$ and $\mathbf{k}$ are the wave vectors of the incident and backscattered waves, respectively, and A is the surface area of the scattering medium. The integration in (21) is carried out over the volume of the medium.

We confine ourselves to consideration of normally incident radiation, in which the effects of refraction at the boundary are negligible, and everything is determined by the wave propagation dynamics within the medium. The features associated with oblique incidence are discussed in Ref. 29.

For normal incidence and small deviation angles θ from the ''backward'' direction, Eq. (21) takes the form

$$J_c(\theta) = \frac{1-r}{4\pi l^2}\int_0^L dz \int_0^L dz' \exp\left(-\frac{z+z'}{l}\right) F(z,z',q), \tag{22}$$

where $F(z,z',q)$ is the Fourier transform of the radiation energy density (3), $q = |(\mathbf{k}+\mathbf{k}_0)_\parallel| = k_0\theta$, $\mathbf{k}_{0\parallel}$ and $\mathbf{k}_\parallel$ are the components of $\mathbf{k}_0$ and $\mathbf{k}$ that are parallel to the boundaries of the layer, $k_0 = 2\pi/\lambda$, λ is the wavelength of the radiation, and $r = r(\mu = 1)$ is the specular reflection coefficient (7) for normal incidence. The distribution (22) is normalized to unit incident flux passing through the boundary.

Substituting the propagator (11), which was calculated in the diffusion approximation, into (22), for a semi-infinite medium we obtain

$$J_c(\theta)=\frac{3}{8\pi}(1-r)\frac{(1+2z_0/l)-\nu(2+3z_0/l)}{1+\nu z_0/l}, \qquad (23)$$

where $\nu=k_0 l\theta\ll 1$. This generalizes previous results[6,7,22,26] to the case of finite internal reflection. In the vicinity of the peak ($\nu z_0/l\ll 1$), the backscattering spectrum has the form

$$J_c(\theta)\approx\frac{3(1-r)}{8\pi}\left(1+2\frac{z_0}{l}-2\nu\left(1+\frac{z_0}{l}\right)^2\right), \qquad (24)$$

and does not differ in structure from the results obtained in the absence of reflection.[6,7,22,26,29] It follows from (24) that in the case of strong reflection ($1-r\ll 1$), where z is proportional to $(1-r)^{-1}$, the height of the peak does not vary, while the coefficient of ν (i.e., the slope of the angular dependence $J_c(\theta)$) increases. Thus, narrowing of the angular dependence $J_c(\theta)$ similar to that discussed in Ref. 13 is observed. In the limit of strong reflection from the boundary ($1-r\ll 1$), the formula

$$J_c(\theta)=\frac{1-r}{\pi(1-r+4\nu/3)}=\begin{cases}\dfrac{1}{\pi}\left(1-\dfrac{4}{3}\dfrac{\nu}{1-r}+\ldots\right), & \nu\ll 1-r\\ \dfrac{3}{4\pi}\dfrac{1-r}{\nu}, & \nu\gg 1-r\end{cases} \qquad (25)$$

is asymptotically accurate with respect to the large parameter $(1-r)^{-1}$.[11] This can be seen by comparing the result with the solution in Appendix A and the exact result for an infinite medium.[30]

The expression for the angular coherent backscattering spectrum from a layer of finite thickness with strong internal reflection has the form

$$J_c(\theta)=\frac{1-r}{\pi}\frac{(1-r)\sinh(\nu L/l)+(4/3)\nu\cosh(\nu L/l)}{((1-r)^2+(4\nu/3)^2)\sinh(\nu L/l)+(8/3)\nu(1-r)\cosh(\nu L/l)}. \qquad (26)$$

In (26) it is assumed that $L/l\gg 1$.

According to (26) the finite character of the layer is manifested at $\nu\ll l/L$. In this region the expression for $J_c(\nu)$ can be written

$$J_c(\theta)=\frac{1-r}{\pi}\frac{((1-r)L/l+4/3)+\nu^2(L/l^2)((1-r)L/l+4)/6}{(1-r)((1-r)L/l+8/3)+4\nu^2L((1-r)L/l+4/3+(1-r)^2(L/l)^2/8)/3l}. \qquad (27)$$

When $L/l>1/(1-r)$, the linear dependence of J_c on ν is replaced at $\nu<l/L<1-r$ by the quadratic law

$$J_c(\theta)=\frac{1}{\pi}\left(1-\frac{4l}{3(1-r)L}-\frac{4}{9}\nu^2\frac{L}{(1-r)l}+\ldots\right). \qquad (28)$$

In the limit of strong internal reflection ($L/l<1/(1-r)$), the linear portion of $J_c(\nu)$ is totally absent, since the dependence $J_c(\nu)\sim 1/\nu$ can be replaced at $\nu<l/L$ by a dependence of the form

$$J_c(\theta)=\frac{1-r}{2\pi}\frac{1}{(1-r)+2\nu^2L/3l}=\begin{cases}\dfrac{1}{2\pi}\left(1-\dfrac{2}{3}\dfrac{L}{l}\dfrac{\nu^2}{1-r}+\ldots\right), & \nu<\sqrt{\dfrac{(1-r)l}{L}}\\ \dfrac{3}{4\pi}\dfrac{l}{L}\dfrac{l-r}{\nu^2}, & \dfrac{l}{L}>\nu>\sqrt{\dfrac{(1-r)l}{L}}\end{cases}. \qquad (29)$$

In the case under consideration, truncation of the coherent backscattering peak occurs at $\nu\sim\sqrt{(1-r)l/L}<l/L$, which is associated with the confinement of radiation in the medium and the resultant lengthening of the wave propagation trajectories. Another manifestation of radiation confinement is the doubling of the scattering intensity exactly in the backward direction. This can be attributed to the equality between the probabilities of the emergence of radiation through the front and rear boundaries of the medium under the conditions of strong internal reflection. The variation of the form of the angular coherent backscattering spectrum with decreasing thickness of the scattering layer is shown in Fig. 1.

We assumed above that there is no absorption. To take weak absorption in the medium into account, ν must be replaced by $\sqrt{\nu^2+3(1-\omega)}\ll 1$, where $\omega=l_{\text{tot}}/l$ is the albedo for single scattering,[6,7] in the results obtained above.

4. LONG-RANGE SPATIAL INTENSITY CORRELATIONS

Let us now consider the long-range correlations between the radiation intensities reflected or transmitted through a layer. Like coherent backscattering, long-range intensity correlations are sensitive to the confinement of waves in the medium under the conditions of internal reflection from the boundaries.[16,17,23]

We define the intensity correlation function using the relation

$$C(\boldsymbol{\rho},z_f)=\langle E(\boldsymbol{\rho},z=z_f)E(0,z=z_f)\rangle-\langle E(0,z=z_f)\rangle^2, \qquad (30)$$

where $E(\boldsymbol{\rho},z_f)$ is the flux density through the surface at the point $(\boldsymbol{\rho},z_f)$, z_f is the coordinate of the plane of observation:

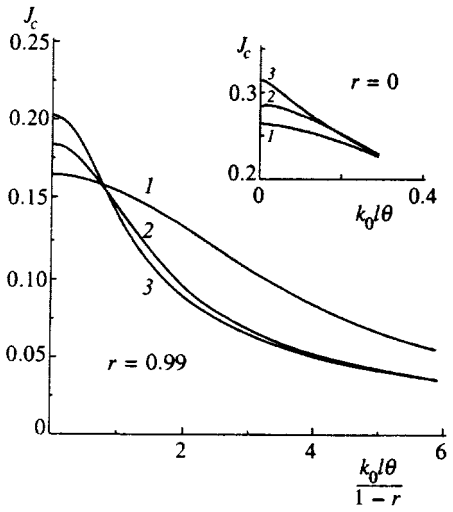


FIG. 1. Angular coherent backscattering spectrum from a layer of a random medium under the conditions of strong internal reflection. The numbers on the curves correspond to different values of the layer thickness $L/l=10$ (*1*), 50 (*2*), and 100 (*3*); the internal reflection coefficient of the boundaries $r=0.99$. For comparison the angular spectrum in the absence of internal reflection is shown in the inset ($r=0$ and $L/l=5$ (*1*), 10 (*2*), and 100 (*3*)).

$z_f=0$ for reflection and $z_f=L$ for transmission. The angle brackets in (30) denote averaging over the disposition of the scatterers.

The Fourier transform of the correlation function, i.e., the fluctuation spectrum

$$M(\mathbf{q},z_f)=\int C(\boldsymbol{\rho},z_f)\exp(-i\mathbf{q}\cdot\boldsymbol{\rho})\,d^2\rho, \tag{31}$$

obeys the relation[23]

$$\begin{aligned}M(\mathbf{q},z_f)=&\left(\frac{2\pi}{k_0}\right)^2\frac{1}{4\pi l}\\&\times\int_0^L dz\int_{4\pi}d\boldsymbol{\Omega}\int_{4\pi}d\boldsymbol{\Omega}'|\langle E_f(z_f,\mathbf{q}|z,\boldsymbol{\Omega})\rangle\\&-\langle E_f(z_f,\mathbf{q}|z,\boldsymbol{\Omega}')\rangle|^2\langle I_i(z,\boldsymbol{\Omega}|z_i\\&=0,\boldsymbol{\Omega}_0)\rangle\langle I_i(z,\boldsymbol{\Omega}'|z_i=0,\boldsymbol{\Omega}_0)\rangle,\end{aligned} \tag{32}$$

where $\langle I_i(z,\boldsymbol{\Omega}|z_i=0,\boldsymbol{\Omega}_0)\rangle$ is the mean intensity in the direction $\boldsymbol{\Omega}$ at depth z for radiation incident upon the surface of the medium at $z_i=0$ in the direction $\boldsymbol{\Omega}_0$,

$$\begin{aligned}\langle E_f(z_f,\mathbf{q}|z,\boldsymbol{\Omega})\rangle=&\int d^2\rho\,\exp(-i\mathbf{q}\cdot\boldsymbol{\rho})\int_{(\mathbf{n}\cdot\boldsymbol{\Omega}_1)>0}d\boldsymbol{\Omega}_1\\&\times(\mathbf{n}\cdot\boldsymbol{\Omega}_1)\langle I_f(\boldsymbol{\rho},z_f,\boldsymbol{\Omega}_1|z,\boldsymbol{\Omega})\rangle,\end{aligned} \tag{33}$$

$\langle I_f(\boldsymbol{\rho},z_f,\boldsymbol{\Omega}_1|z,\boldsymbol{\Omega})\rangle$ is the mean intensity in the direction $\boldsymbol{\Omega}_1$ at the point $(\boldsymbol{\rho},z_f)$ on the surface at z_f from a point source located at depth z emitting in the direction $\boldsymbol{\Omega}$. The mean intensity satisfies the transport equation (1) with appropriate sources.

Equation (32) was derived in Ref. 23 in the Born approximation for the amplitude of one-center scattering, in which only pairwise correlations of the wave fields are taken into account. For isotropic scattering centers of small radius, Eq. (32) makes the main contribution to the fluctuation spectrum.[31] The values of the ''outgoing'' propagators $\langle E_f\rangle$ in Eq. (32) are taken outside the medium in immediate proximity to its boundary. The values of $\langle I_i\rangle$ are normalized to unit incident flux after passage through the boundary of the medium.

Equation (32) simultaneously takes into account the contributions of short-range and long-range intensity correlations.[23,31] As was shown in Refs. 23 and 31, owing to the conservation of radiation flux upon elastic scattering, the contributions of short-range and long-range correlations to the fluctuation spectrum partially cancel, and as a result the spectrum (32) is determined by integration over scales of variation of the variable z that exceed the mean free path l. The diffusion approximation (3) can be used to calculate the spectrum. Repeating the transformations made in Refs. 23 and 31, but with consideration of the boundary conditions (8) and (9), we obtain the following expression for the fluctuation spectrum in the diffusion approximation:

$$\begin{aligned}M(q,z_f)=&\frac{2\pi(I+z_0)^2l^3}{27k_0^2}\int_0^L dz\left(\left|\frac{\partial}{\partial z_f}\frac{\partial}{\partial z}F(q,z_f,z)\right|^2\right.\\&\left.+q^2\left|\frac{\partial}{\partial z_f}F(q,z_f,z)\right|^2\right)\left|\frac{\partial}{\partial z_i}F(q=0,z,z_i)\right|^2,\end{aligned} \tag{34}$$

where the function $F(q,z,z')$ is defined by (11). It is assumed in (34) that $ql\ll 1$.

Below we analyze the spectrum $M(q,z_f)$ for purely elastic scattering under the conditions of strong internal reflection ($1-r\ll 1$ or $z_0\gg l$). The result of the integration in (34) for the general case, in which the incident waves differ in frequency and absorption occurs in the medium, is presented in Appendix B.

In the case of reflection from a semi-infinite nonabsorbing medium, the intensity fluctuation spectrum has the form ($z_i=z_f=0$)

$$M_R(q)=\frac{6\pi}{k_0^2}\left(\frac{z_0}{l}\right)^2\frac{ql}{(1+qz_0)^2}. \tag{35}$$

At $q=0$ Eq. (35) specifies the dispersion of the reflection coefficient,[23] $\langle(\delta R)^2\rangle=M_R(q=0)/A$, and vanishes because of the conservation of flux upon elastic scattering. The spectrum (35) peaks, with $(M_R)_{\max}=3\pi z_0/(2k_0^2l)$, at $q=1/z_0$, and decays as $1/q$ at $q>1/z_0$. In the range $1/z_0<q<1/l$ the decay of the spectrum is described by the relation

$$M_R(q)=6\pi/k_0^2ql, \tag{36}$$

which characterizes the intensity fluctuations in a medium with a perfectly reflecting boundary.

The correlation function corresponding to the spectrum (35) is

$$C_R(\rho)=\frac{3}{k_0^2}\left(\frac{z_0}{l}\right)^2\begin{cases}\dfrac{l}{z_0^3}\left(\dfrac{z_0}{\rho}-2\ln\dfrac{z_0}{\rho}+\ldots\right), & \rho\ll z_0\\[2ex] -\dfrac{l}{\rho^3}, & \rho\gg z_0\end{cases}. \tag{37}$$

According to (37), under the conditions of strong internal reflection, the main contribution to the correlation function for $l<\rho\ll z_0$ does not depend on z_0: $C_R(\rho)\approx 3l/(k_0^2 l^2\rho)$. At such values of ρ the probability of photons leaving the medium is negligible. When $\rho\gg z_0$, the probability of leaving the medium after multiple reflections becomes appreciable, and the correlation function begins to decay according to the same law as in the absence of reflection at the boundary.[23] However, because of the large number of reflections of the radiation from the boundary, the amplitude of long-range intensity correlations increases (by a factor of $(z_0/l)^2\approx(4/3(1-r))^2\gg 1$ for $\rho\gg z_0$).

In the case of the reflection of waves from a layer of finite thickness, the intensity fluctuation spectrum takes the form

$$M_R(q)=\frac{6\pi}{k_0^2}\left(\frac{z_0}{l}\right)^2\frac{l}{L}F_R(qL,qz_0), \tag{38}$$

where

$$\begin{aligned}F_R(x,x_0)=&\frac{1}{2}\frac{x}{(x+2x_0)^2((1+x_0^2)\sinh x+2x_0\cosh x)^2}\\&\times\left\{\left((x+x_0)^2+\frac{1}{2}\right)((1+x_0^2)\sinh 2x+2x_0\right.\\&\times\cosh 2x)-(x+x_0)((1+x_0^2)\cosh 2x\\&\left.+2x_0\sinh 2x)-x_0^3\right\}.\end{aligned} \tag{39}$$

In the limit of large layer thickness ($l\ll z_0\ll L$), the dependence of $M_R(q)$ on L is significant only at low spatial frequencies ($1<1/L$), where

$$M_R(q)=\frac{2\pi}{k_0^2}\left(\frac{z_0}{l}\right)^2\frac{l}{L}\left(1+\frac{13}{15}(qL)^2+\ldots\right). \tag{40}$$

Equation (40) is distinguished from the analogous equation obtained with neglect of internal reflection at the boundary[23] only by the enhancement factor $(z_0/l)^2$. It describes the enhancement of the amplitude of long-range correlations, which has been noted above. At $q>l/L$ the spectrum $M_R(q)$ behaves just as in the case of reflection from a semi-infinite medium.

Such behavior of the fluctuation spectrum means that the intensity correlations for the reflection of radiation in the thick-layer limit behave as in a medium with a perfectly reflecting boundary, if $\rho<z_0$, while for $\rho>z_0$, they mimic the dependence obtained in the absence of internal reflection,[23] enhanced by a factor $(z_0/l)^2$.

In the case of strong internal reflection, a situation in which the length z_0 is the maximum spatial scale ($l\ll L\ll z_0$) is possible. It is possible, if the transmission of the boundaries $t=1-r$ is significantly lower than the diffuse transmission of the medium itself $T=4l/3L$ (see inequality (20)). In this case the character of the correlations is radically altered. The spectrum can be represented in the form

$$\begin{aligned}M_R(q)=&\frac{6\pi}{k_0^2}\left(\frac{z_0}{l}\right)^2\frac{l}{L}\\&\times\begin{cases}\dfrac{1}{16}\left(\dfrac{L}{z_0}\right)^2(1+(qz_0)^2), & q\ll 1/z_0\\[2ex]\dfrac{1}{8}\dfrac{qL\sinh 2qL}{(2\cosh qL+qz_0\sinh qL)^2}, & q\gg 1/z_0\end{cases}.\end{aligned} \tag{41}$$

According to (41), the spectrum $M_R(q)$ has a maximum at $q=(2/z_0L)^{1/2}$. At $q\gg(2/z_0L)^{1/2}$ the spectrum decays according to

$$M_R(q)=\frac{3\pi}{4k_0^2}\frac{1}{ql}\frac{\sinh 2qL}{\sinh^2 qL}, \tag{42}$$

and does not depend on z_0. Equation (42) describes the fluctuation spectrum in a layer of a medium bounded by perfectly reflecting walls. It is noteworthy that the dependence $M_R(q)\approx 3\pi/(2k_0^2ql)$, which is distinguished from (36) only by the additional factor 1/4, follows from (42) at $q\gg l/L$. It is fairly simple to understand the origin of this factor. When $z_0\gg L$, radiation is reflected repeatedly from the boundaries and ''forgets'' through which boundary it entered the medium; the reflection and transmission coefficients both equal 1/2. Likewise, the fluctuation spectrum of the reflected intensity at $q\gg 1/L$ should coincide with the analogous expression for a semi-infinite medium, which, however, corresponds to the incident intensity diminished by a factor of 2.

The correlation function corresponding to the spectrum (41) behaves in the following manner:

$$\begin{aligned}C_R(\rho)&\\&=\frac{3}{(k_0l)^2}\begin{cases}\dfrac{1}{4}\dfrac{l}{\rho}, & l<\rho<L\\[2ex]\dfrac{1}{8}\dfrac{l}{L}\ln\dfrac{2z_0L}{\rho^2}, & L<\rho<\sqrt{z_0L/2}\\[2ex]-\dfrac{1}{8}\dfrac{l}{L}\sqrt{\dfrac{\pi\rho}{\sqrt{2Lz_0}}}\exp\left(-\rho\sqrt{\dfrac{2}{Lz_0}}\right), & \\ \rho>\sqrt{z_0L/2}.\end{cases}\end{aligned} \tag{43}$$

A comparison with the results in Ref. 23 shows that the behavior of $C_R(\rho)$ is totally different from that which we observed in the absence of internal reflection at the boundaries. The regions of positive and negative intensity correlations are interchanged.

At relatively short distances, $l<\rho<L$, $C_R(\rho)$ behaves in accordance with the laws discussed above. At large ρ ($\rho>L$), the correlation function decreases smoothly, changes sign at $\rho\approx\sqrt{Lz_0}$, passes through a minimum at $\rho\approx 1.7\sqrt{Lz_0}$, and then tends to zero exponentially, remaining negative all the time, in contrast to Ref. 23.

We now discuss the influence of strong internal reflection on the fluctuation spectrum of the intensity transmitted through a layer ($z_i=0$, $z_f=L$),

$$M_T(q)=\frac{6\pi}{k_0^2}\left(\frac{z_0}{l}\right)^2\frac{l}{L}F_T(qL,qz_0), \tag{44}$$

where

$$F_T(x,x_0)=\frac{1}{2}\frac{x}{(x+2x_0)^2((1+x_0^2)\sinh x+2x_0\cosh x)^2}$$
$$\times\left\{\left(x_0^2+\frac{1}{2}\right)((1+x_0^2)\sinh 2x+2x_0\cosh 2x)\right.$$
$$+x_0((1+x_0^2)\cosh 2x+2x_0\sinh 2x)$$
$$\left.-(1+x_0^2)(x+x_0)-2x_0(x+x_0)^2-x_0\right\}. \tag{45}$$

In the thick-layer limit ($l\ll z_0\ll L$), from (44) and (45) we obtain

$$M_T(q)=\frac{6\pi}{k_0^2}\left(\frac{z_0}{l}\right)^2\frac{l}{L}\begin{cases}\frac{1}{3}\left(1-\frac{2}{15}(qL)^2+\ldots\right), & q\ll 1/L\\ \frac{1}{2qL}\left(1+\frac{(qz_0)^2}{(1+qz_0)^2}\right), & q\gg 1/L\end{cases}. \tag{46}$$

The amplitude of intensity correlations of the transmitted radiation is increased by a factor of $(z_0/l)^2$ in comparison to the case in which internal reflection at the boundaries is insignificant.[27,32,33] This is a consequence of multiple reflection from the boundaries of the layer. In addition, at large q ($q>1/z_0$) the amplitude of the spectrum is additionally increased by a factor of 2, since a correlation regime as in a medium with a perfectly reflecting boundary sets in. At $q>1/z_0$ the calculations can be described by the appropriate equation, which is discussed above for the reflection geometry (see Eq. (36)), with the sole difference being that the power of the ''source'' must be taken into account in it, i.e., it must be additionally multiplied by the square of the transmission coefficient through the layer $T\sim z_0/L$ (see Eq. (18) for $l\ll z_0\ll L$).

In the limiting case of strong internal reflection ($l\ll L\ll z_0$), as noted above, the transmission of the scattering medium itself (without the boundaries) exceeds the transmission of the boundaries of the layer. During repeated traversals of the layer, photons forget through which of the boundaries they penetrated the scattering medium. Therefore, fluctuations at both boundaries display the same behavior (compare (44) and (45) with (41)–(43)).

Qualitative variations in the form of $M_R(q)$ and $M_T(q)$ as the internal reflection coefficient increases are shown in Fig. 2.

Now consider fluctuations in the total reflection and transmission coefficients of the medium. The dispersion of these parameters is specified by the value of the spectrum at $q=0$.[23] In purely elastic scattering, with which we are concerned, $R+T=1$, and therefore

$$\langle(\delta R)^2\rangle=\langle(\delta T)^2\rangle=\frac{1}{A}M_{R,T}(q=0). \tag{47}$$

Substituting the expressions (38) and (39) or (44) and (45) into (47), we obtain

$$\langle(\delta R)^2\rangle=\langle(\delta T)^2\rangle=\frac{2\pi l}{Ak_0^2}\left(\frac{z_0}{l}\right)^2\frac{(L+z_0)^3-z_0^3}{(L+2z_0)^4}. \tag{48}$$

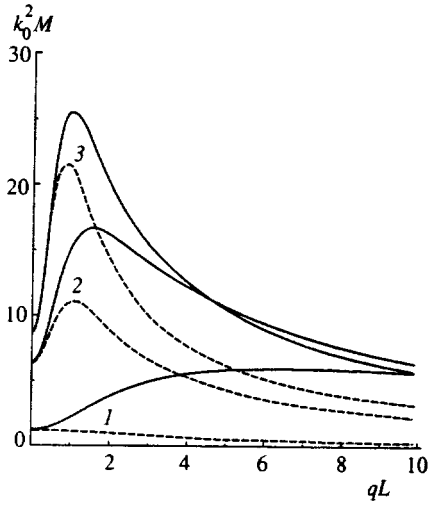


FIG. 2. Fluctuation spectra of reflected (solid curves) and transmitted (dashed curves) radiation for various values of the internal reflection coefficient ($r=0.5$ (*1*), 0.9 (*2*), and 0.95 (*3*)). The layer thickness $L/l=10$.

Expression (48) depends nonmonotonically on the layer thickness L. When $L\ll z_0$, the dispersion increases linearly with the thickness

$$\langle(\delta R)^2\rangle=\langle(\delta T)^2\rangle\approx\frac{3\pi}{2Ak_0^2}\frac{L}{l},$$

while in the thick-layer limit the dispersion decreases:

$$\langle(\delta R)^2\rangle=\langle(\delta T)^2\rangle\approx\frac{2\pi}{Ak_0^2}\left(\frac{z_0}{l}\right)^2\frac{1}{L}.$$

The dispersion reaches its maximum value when $L\approx 2z_0$:

$$\langle(\delta R)^2\rangle=\langle(\delta T)^2\rangle\approx\frac{\pi}{5Ak_0^2}\frac{z_0}{l}.$$

As for the relative magnitude of fluctuations in the transparency of the medium, $\langle(\delta T)^2\rangle/\langle T\rangle^2$ increases linearly with thickness L:

$$\frac{\langle(\delta T)^2\rangle}{\langle T\rangle^2}=\frac{2\pi}{2Ak_0^2}\left(\frac{L}{l}\right)\frac{L^2+3Lz_0+3z_0^2}{(L+2z_0)^2}$$
$$=\frac{2\pi}{2Ak_0^2}\left(\frac{L}{l}\right)\begin{cases}\frac{3}{4}, & L\ll z_0\\ 1, & L\gg z_0\end{cases}. \tag{49}$$

We have thus far assumed that waves incident upon the medium are monochromatic. However, this condition does not hold in many experiments.[7,15] The correlations between radiation fluxes at different frequencies have been a subject of investigation.

If the frequencies of the incident waves differ by $\Delta\omega$, the corresponding expression for the fluctuation spectrum can be obtained from (32) by substituting into (32) ''incoming'' propagators of the form

$$\langle I_i(z,\mathbf{\Omega}|z_i=0,\mathbf{\Omega}_0,\Delta\omega)\rangle\langle I_i(z,\mathbf{\Omega}'|z_i=0,\mathbf{\Omega}_0,-\Delta\omega)\rangle,$$

where $\langle I_i(z,\mathbf{\Omega}|z_i=0,\mathbf{\Omega}_0,\pm i\Delta\omega)\rangle$ satisfies a transport equation like (1) for a signal modulated at frequency $\pm\Delta\omega$. The

corresponding equation is obtained from (1) after replacing l_a^{-1} by $l_a^{-1} \pm i\Delta\omega/c$.[7] At small values of $\Delta\omega$ ($\Delta\omega \ll c/l$), the intensity propagators can be calculated in the diffusion approximation. In this approximation one obtains an expression for the spectrum that differs from (34) only in that the ''incoming'' F functions must be taken (see Eq. (11)) for complex absorption ($l_a^{-1} \pm i\Delta\omega/c$).

An investigation of the frequency dependence of the correlation between the total reflected or transmitted fluxes would be of great interest.[7,16,17] In the case of purely elastic scattering in the medium, the correlations between the transmitted and reflected radiation fluxes coincide:

$$\begin{aligned} C(\Delta\omega) &= \langle (T_{\omega_0+\Delta\omega/2} - \langle T_{\omega_0+\Delta\omega/2}\rangle) \\ &\quad \times (T_{\omega_0-\Delta\omega/2} - \langle T_{\omega_0-\Delta\omega/2}\rangle)\rangle \\ &= \langle (R_{\omega_0+\Delta\omega/2} - \langle R_{\omega_0+\Delta\omega/2}\rangle)(R_{\omega_0-\Delta\omega/2} \\ &\quad - \langle R_{\omega_0-\Delta\omega/2}\rangle)\rangle. \end{aligned} \tag{50}$$

This is a consequence of flux conservation: $R+T=1$.

When $\Delta\omega \ll c/l$, we obtain the following expression for $C(\Delta\omega)$ in the diffusion approximation:

$$C(\Delta\omega) = \frac{6\pi}{Ak_0^2}\left(\frac{z_0}{L+2z_0}\right)^2 \left(\frac{L}{l}\right)\Phi\left(\frac{L}{l_{\Delta\omega}}, \frac{z_0}{L}\right), \tag{51}$$

where $l_{\Delta\omega} = (2lc/3\Delta\omega)^{1/2}$, and the universal function $\Phi(x,y)$ has the form

$$\begin{aligned} \Phi(x,y) &= \{2x^2y^2(\sinh 2x + \sin 2x) + (\sinh 2x - \sin 2x) \\ &\quad + 2xy(\cosh 2x - \cos 2x)\}\{2x[4xy(\sinh 2x \\ &\quad + \sin 2x) + 8x^3y^3(\sinh 2x - \sin 2x) \\ &\quad + 8x^2y^2(\cosh 2x + \cos 2x) + (1 + 4x^4y^4) \\ &\quad \times (\cosh 2x - \cos 2x)]\}^{-1}. \end{aligned} \tag{52}$$

The length $l_{\Delta\omega}$ determines the depth to which the bulk speckle structure produced by interference of the incident waves penetrates. Plots of $C(\Delta\omega)$ for various values of the internal reflection coefficient are shown in Fig. 3. Equations (51) and (52) can be simplified in two limiting cases.

When $L \ll l_{\Delta\omega}$, the dependence of C on the frequency is appreciable only with very strong internal reflection ($z_0 \gg L$). In this case

$$\begin{aligned} C(\Delta\omega) &\approx \frac{3\pi}{8Ak_0^2}\left(\frac{L}{l}\right)\frac{l_{\Delta\omega}^4}{l_{\Delta\omega}^4 + L^2z_0^2} = \frac{3\pi}{8Ak_0^2}\left(\frac{L}{l}\right) \\ &\quad \times \begin{cases} 1, & \sqrt{Lz_0} \ll l_{\Delta\omega} \\ \dfrac{l_{\Delta\omega}^4}{L^2z_0^2}, & L \ll l_{\Delta\omega} \ll \sqrt{Lz_0}, \end{cases} \end{aligned} \tag{53}$$

i.e., the frequency dependence is manifested for $l_{\Delta\omega} < \sqrt{Lz_0}$, and the correlations weaken as $\Delta\omega$ increases according to a $\Delta\omega^{-2}$ law.

When $l_{\Delta\omega} \ll L$, the relations (51) and (52) take the form

$$C(\Delta\omega) \approx \frac{3\pi}{Ak_0^2}\left(\frac{z_0}{L+2z_0}\right)^2 \frac{l_{\Delta\omega}^3}{l((l_{\Delta\omega}+z_0)^2 + z_0^2)}$$

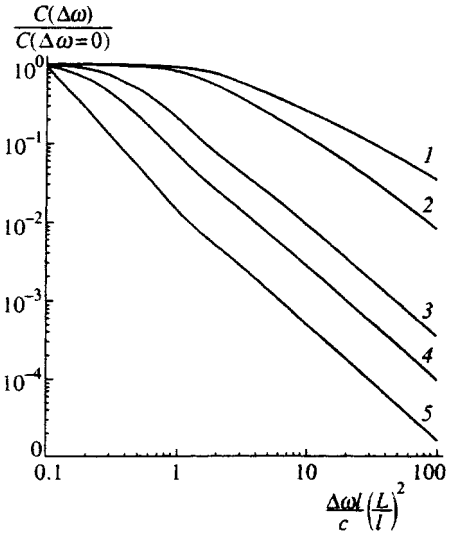


FIG. 3. Frequency-dependent flux correlation function for various values of the internal reflection coefficient ($r=0$ (*1*), 0.5 (*2*), 0.9 (*3*), 0.95 (*4*), and 0.99 (*5*)). The layer thickness is $L/l=10$.

$$= \frac{3\pi}{Ak_0^2}\left(\frac{z_0}{L+2z_0}\right)^2 \begin{cases} \dfrac{l_{\Delta\omega}}{l}, & z_0 \ll l_{\Delta\omega} \\ \dfrac{l_{\Delta\omega}^3}{2lz_0^2} & l_{\Delta\omega} \ll z_0 \end{cases}. \tag{54}$$

According to (54), when internal reflection is not very strong (or the frequency shift $\Delta\omega$ is not very large), the behavior of the correlation function $C(\Delta\omega)$ is qualitatively the same as when there is no internal reflection.[7,27,32,33] The behavior of $C(\Delta\omega)$ does, however, change at large $\Delta\omega$. The correlations between fluxes at differing frequencies decrease with increasing $\Delta\omega$ according to a $\Delta\omega^{-3/2}$ law. This result differs from the conclusions in Refs. 16 and 17, where other laws, viz, $\Delta\omega^{-1}$ and $\Delta\omega^{-1/2}$, respectively, were predicted.

The disparity with Refs. 16 and 17 can be explained as follows. In Ref. 16, ''surface'' sources of radiation fluctuations were introduced phenomenologically in an incorrect manner. The way in which this was done in Ref. 16 is inconsistent with flux conservation upon elastic scattering. In Ref. 17, the expression for the intensity correlation function obtained in that work is also inconsistent with flux conservation. In deriving the correlation function, van Rossum and Nieuwenhuizen[17] assumed that it is symmetric under interchange of the ''incoming'' and ''outgoing'' propagators. Such symmetry exists only for correlations of diffusion fluxes in an infinite homogeneous medium.[34] In the case of a finite geometry, the symmetry of the intensity correlation function becomes more complicated.[33] Only part of the total correlation function derived in Ref. 33 contributes to the frequency-dependent flux correlation function. It does not have the indicated symmetry.

5. CONCLUSIONS

We have shown that internal reflection from the boundaries of a random system alters the wave propagation regime

on scales exceeding the mean free path, and has an appreciable influence on effects due to long-range correlation of multiply scattered waves.

Under the conditions of strong internal reflection $(1-r\ll 1)$, significant restructuring of the spatial distribution of the radiation in the medium occurs. This is evidenced by calculations performed both in the diffusion approximation and under the assumption of diffusely reflecting boundaries. In the limit of strong internal reflection, the two approaches yield identical results. The structure of the exact relations enables us to understand on a qualitative level that the radiation propagation regime in the medium is determined by the competition between the reflectivity (or transmissivity) of the boundaries, and of the scattering medium itself. When $1-r\ll 1$, two situations can be distinguished.

In the thick-layer limit, where the transmission of the boundaries is higher than that of the scattering medium itself $(1-r\gg l/L)$, internal reflection simply increases the effective length of the path that waves traverse before leaving the medium. The propagation of radiation along comparatively short trajectories (shorter than $l/(1-r)$) takes place as in a medium with a perfectly reflecting boundary.

Qualitative changes in the radiation distribution appear in the opposite limiting case, where the transmission of the boundaries is lower than the transmission of the scattering medium itself $(1-r\ll l/L)$. Radiation crosses the layer many times from one boundary to the other, and as a result, ''forgets'' from which side it entered the medium. The difference between the reflection geometry and the transmission geometry vanishes. Before leaving the medium, the waves traverse a path of the order of $(lL/(1-r))^{1/2}\gg L$.

The features of the large-scale behavior of the radiation distribution are reflected in the shape of the angular coherent backscattering spectrum. The spectrum narrows as a whole. However, in the limit of strong internal reflection, the shape of the spectrum differs qualitatively from that observed in the absence of reflecting boundaries.

Internal reflection also leads to appreciable variation of the spatial dependence of the long-range intensity correlations in the reflected and transmitted radiation fluxes. When $1-r\ll l/L$, regions of positive and negative intensity correlation are interchanged, compared with those for the case in which there is no reflection from the boundaries. The shape of the fluctuation spectrum also changes. The spectrum for $q\neq 0$ exhibits a maximum, which corresponds to reversal of the sign of the correlation function. New qualitative features appear in the correlations between fluxes of coherent radiation at different frequencies. In the limit of strong internal reflection $(1-r\ll l/L)$, the correlations weaken more rapidly as the frequency shift increases than in the absence of reflecting boundaries: they go as $\Delta\omega^{-2}$ when $c(1-r)/L<\Delta\omega<cl/L^2$, and as $\Delta\omega^{-3/2}$ when $\Delta\omega>cl/L^2$.

We thank E. E. Gorodnichev and A. I. Kuzovlev for their interest in this work and valuable comments. This research was carried out with partial support from the International Science Foundation (Grant N3U000 and N3U300) and the Russian Fund for Fundamental Research (Project 95-02-05530).

APPENDIX A

In a semi-infinite medium with a diffusely reflecting boundary, transport equation (1) with source (2) and reflection coefficient (10) can be solved exactly. We write (1) in integral form:[8]

$$I(\mathbf{r},\mathbf{\Omega}|\mathbf{r}',\mathbf{\Omega}')=I^{(0)}(\mathbf{r},\mathbf{\Omega}|\mathbf{r}',\mathbf{\Omega}')+\frac{1}{4\pi l}\int d^3r''\times\int_{4\pi}d\mathbf{\Omega}''\int_{4\pi}d\mathbf{\Omega}'''I^{(0)}\times(\mathbf{r},\mathbf{\Omega}|\mathbf{r}'',\mathbf{\Omega}'')I(\mathbf{r}'',\mathbf{\Omega}'''|\mathbf{r}',\mathbf{\Omega}')+\frac{1}{\pi}r\int d^2\rho''E^{(0)}(\mathbf{r},\mathbf{\Omega}|z''=0,\boldsymbol{\rho}'')E(z''=0,\boldsymbol{\rho}''|\mathbf{r}',\mathbf{\Omega}''),\tag{A1}$$

where

$$I^{(0)}(\mathbf{r},\mathbf{\Omega}|\mathbf{r}',\mathbf{\Omega}')=\delta(\mathbf{\Omega}-\mathbf{\Omega}')\delta\left(\mathbf{\Omega}-\frac{\mathbf{r}-\mathbf{r}'}{|\mathbf{r}-\mathbf{r}'|}\right)\times\frac{1}{|\mathbf{r}-\mathbf{r}'|^2}\exp\left(-\frac{|\mathbf{r}-\mathbf{r}'|}{l_{\mathrm{tot}}}\right),\tag{A2}$$

$$E^{(0)}(\mathbf{r},\mathbf{\Omega}|z''=0,\boldsymbol{\rho}'')=\int_{(\mathbf{n}\cdot\mathbf{\Omega}'')<0}d\mathbf{\Omega}''|\mathbf{n}\cdot\mathbf{\Omega}''|\times I^{(0)}(\mathbf{r},\mathbf{\Omega}|z''=0,\boldsymbol{\rho}'',\mathbf{\Omega}''),\tag{A3}$$

$$E(z''=0,\boldsymbol{\rho}''|\mathbf{r}',\mathbf{\Omega}')=\int_{(\mathbf{n}\cdot\mathbf{\Omega}'')>0}d\mathbf{\Omega}''(\mathbf{n}\cdot\mathbf{\Omega}'')\times I(z''=0,\boldsymbol{\rho}'',\mathbf{\Omega}''|\mathbf{r}',\mathbf{\Omega}').\tag{A4}$$

It is convenient to rewrite the solution of Eq. (A1) in terms of the solution $\tilde{I}$ that is valid in the absence of internal reflection from the boundary:

$$I(\mathbf{r},\mathbf{\Omega}|\mathbf{r}',\mathbf{\Omega}')=\tilde{I}(\mathbf{r},\mathbf{\Omega}|\mathbf{r}',\mathbf{\Omega}')+\frac{1}{\pi}r\int d^2\rho''\times\tilde{E}(\mathbf{r},\mathbf{\Omega}|z''=0,\boldsymbol{\rho}'')E(z''=0,\boldsymbol{\rho}''|\mathbf{r}',\mathbf{\Omega}'),\tag{A5}$$

where $\tilde{I}$ satisfies Eq. (A1) with $r=0$, and

$$\tilde{E}(\mathbf{r},\mathbf{\Omega}|z''=0,\boldsymbol{\rho}'')=\int_{(\mathbf{n}\cdot\mathbf{\Omega}'')<0}d\mathbf{\Omega}''|\mathbf{n}\cdot\mathbf{\Omega}''|\times\tilde{I}(\mathbf{r},\mathbf{\Omega}|z''=0,\boldsymbol{\rho}'',\mathbf{\Omega}'').\tag{A6}$$

Physically, the transition from (A1) to (A5) corresponds to the separation of multiple scattering within the medium and reflection from the surface.

Solving Eq. (A5) iteratively, we can easily obtain a series expansion of the intensity in the multiplicity of reflection from the boundary. Summation of a series of a similar type was employed in Refs. 14 and 18 to find the total reflected and transmitted flux under the conditions of multiple internal reflection from the boundaries of the medium.

Taking the Fourier transform with respect to the coordinate difference $\boldsymbol{\rho}-\boldsymbol{\rho}'$,

$$I(z,\mathbf{\Omega},\mathbf{q}|z',\mathbf{\Omega}')=\int d(\boldsymbol{\rho}-\boldsymbol{\rho}')\times\exp\{-i\mathbf{q}\cdot(\boldsymbol{\rho}-\boldsymbol{\rho}')\}\times I(z,\mathbf{\Omega},\boldsymbol{\rho}-\boldsymbol{\rho}'|z',\mathbf{\Omega}'),\tag{A7}$$

Eq. (A5) takes the form

$$I(z,\mathbf{\Omega},\mathbf{q}|z',\mathbf{\Omega}')=\tilde{I}(z,\mathbf{\Omega},\mathbf{q}|z',\mathbf{\Omega}')+\frac{1}{\pi}r\tilde{E}(z,\mathbf{\Omega},\mathbf{q}|0)E(0,\mathbf{q}|z',\mathbf{\Omega}').\tag{A8}$$

With consideration of (A4), the solution of Eq. (A8) is

$$I(z,\mathbf{\Omega},\mathbf{q}|z',\mathbf{\Omega}')=\tilde{I}(z,\mathbf{\Omega},\mathbf{q}|z',\mathbf{\Omega}')+\frac{r}{\pi}\frac{\tilde{E}(z,\mathbf{\Omega},\mathbf{q}|0)\tilde{E}(0,\mathbf{q}|z',\mathbf{\Omega}')}{1-rR_q},\tag{A9}$$

where

$$R_q=\frac{1}{\pi}\int_{(\mathbf{n}\cdot\mathbf{\Omega}'')>0}d\mathbf{\Omega}|\mathbf{n}\cdot\mathbf{\Omega}|\times\int_{(\mathbf{n}\cdot\mathbf{\Omega}'')<0}d\mathbf{\Omega}'|\mathbf{n}\cdot\mathbf{\Omega}'|\tilde{I}(z=0,\mathbf{\Omega},\mathbf{q}|z'=0,\mathbf{\Omega}')\tag{A10}$$

is the qth spatial harmonic of the reflection coefficient for a diffuse point source of radiation located on the surface.

Generalization of the result (A9) to the case of a finite layer does not present any difficulty. It calls for supplementing Eq. (A8) with a term corresponding to reflection from the boundary of the layer at $z=L$:

$$I(z,\mathbf{\Omega},\mathbf{q}|z',\mathbf{\Omega}')=\tilde{I}(z,\mathbf{\Omega},\mathbf{q}|z',\mathbf{\Omega}')+\frac{1}{\pi}r\tilde{E}(z,\mathbf{\Omega},\mathbf{q}|0)E(0,\mathbf{q}|z',\mathbf{\Omega}')+\frac{1}{\pi}r\tilde{E}(z,\mathbf{\Omega},\mathbf{q}|L)E(L,\mathbf{q}|z',\mathbf{\Omega}').\tag{A11}$$

The solution of (A11) can be written in the form

$$I(z,\mathbf{\Omega},\mathbf{q}|z',\mathbf{\Omega}')=\tilde{I}(z,\mathbf{\Omega},\mathbf{q}|z',\mathbf{\Omega}')+\frac{1}{\pi N}\{r\tilde{E}(z,\mathbf{\Omega},\mathbf{q}|0)[\tilde{E}(0,\mathbf{q}|z',\mathbf{\Omega}')\times(1-rR_q)+rT_q\tilde{E}(L,\mathbf{q}|z',\mathbf{\Omega}')]+r\tilde{E}(z,\mathbf{\Omega},\mathbf{q}|L)[\tilde{E}(L,\mathbf{q}|z',\mathbf{\Omega}')\times(1-rR_q)+rT_q\tilde{E}(0,\mathbf{q}|z',\mathbf{\Omega}')]\},\tag{A12}$$

where

$$N=(1-rR_q)^2-r^2T_q^2,\tag{A13}$$

and

$$T_q=\frac{1}{\pi}\int_{(\mathbf{n}\cdot\mathbf{\Omega})>0}d\mathbf{\Omega}(\mathbf{n}\cdot\mathbf{\Omega})\int_{(\mathbf{n}\cdot\mathbf{\Omega}')<0}d\mathbf{\Omega}'|\mathbf{n}\cdot\mathbf{\Omega}'|\times\tilde{I}(z'=L,\mathbf{\Omega},\mathbf{q}|z'=0,\mathbf{\Omega}')\tag{A14}$$

is the qth spatial harmonic of the transmission coefficient through a scattering layer for a diffuse point source.

Equations (A9) and (A12) specify the radiation intensity within a scattering medium at any distance z from the boundary. If we set $q=0$, $z'=0$, and $z=L$ in (A9) and (A12), we obtain the familiar relations[25] for the reflected or transmitted intensity for radiation of infinite extent in the xy plane that is incident upon a layer of a medium.

Integrating (A9) and (A12) over $\mathbf{\Omega}$ and $\mathbf{\Omega}'$, we can easily obtain expressions for the radiation energy density $F(q,z,z')$:

$$F(z,z',q)=\tilde{F}(z,z',q)+\frac{r}{\pi}\frac{\Phi(z,q|0)\Phi(0,q|z')}{1-rR_q}\tag{A15}$$

in a semi-infinite medium, and

$$F(z,z',q)=\tilde{F}(z,z',q)+\frac{1}{\pi N}\{r\Phi(z,q|0)\times[\Phi(0,q|z')(1-rR_q)+rT_q\Phi(L,q|z')]+r\Phi(z,q|L)[\Phi(L,q|z')(1-rR_q)+rT_q\Phi(0,q|z')]\},\tag{A16}$$

in a finite layer, respectively, where

$$F(z,z',q)=\int_{4\pi}d\mathbf{\Omega}\int_{4\pi}d\mathbf{\Omega}'I(z,\mathbf{\Omega},q|z',\mathbf{\Omega}'),\tag{A17}$$

$$\Phi(z,q|0)=\int_{4\pi}d\mathbf{\Omega}\int_{(\mathbf{n}\cdot\mathbf{\Omega}')<0}d\mathbf{\Omega}'|\mathbf{n}\cdot\mathbf{\Omega}'|\times\tilde{I}(z,\mathbf{\Omega},\mathbf{q}|z'=0,\mathbf{\Omega}'),\tag{A18}$$

$$\Phi(0,q|z')=\int_{4\pi}d\mathbf{\Omega}'\int_{(\mathbf{n}\cdot\mathbf{\Omega})>0}d\mathbf{\Omega}(\mathbf{n}\cdot\mathbf{\Omega})\times\tilde{I}(z=0,\mathbf{\Omega},\mathbf{q}|z',\mathbf{\Omega}').\tag{A19}$$

Formulas (A15) and (A16) enable us to express the radiation energy density in the medium in terms of a known (for example, in the case of point scatterers[29]) solution obtained in the absence of internal reflection on the boundaries.

APPENDIX B

In the general case, where the layers of the medium absorb radiation and the incident waves differ in frequency, the integration in Eq. (34) with consideration of (11) leads to the following results for the fluctuation spectrum. The spectrum can be represented in the form

$$M(q)=\frac{6\pi}{k_0^2}\frac{(l+z_0)^2}{l^2}Q(q),\tag{B1}$$

where the function $Q(q)$ is

$$Q_R(q)=\frac{1}{A^2(\gamma_q)}\frac{1}{|A(\gamma_{\Delta\omega})|^2}\{M_1(q)F(\gamma_q)+M_2(q)G(\gamma_q)+M_3(q)F(\gamma_q=0)\} \tag{B2}$$

in the case of reflection and

$$Q_T(q)=\frac{1}{A^2(\gamma_q)}\frac{1}{|A(\gamma_{\Delta\omega})|^2}\{N_1(q)F(\gamma_q)-N_2(q)G(\gamma_q)+N_3(q)F(\gamma_q=0)\} \tag{B3}$$

in the case of transmission. The quantities appearing in (B2) and (B3) are

$$\begin{aligned}
&M_1(q)=\frac{1}{2}(1+(\gamma_q z_0)^2)((lq)^2+(l\gamma_q)^2),\\
&M_2(q)=\gamma_q z_0((lq)^2+(l\gamma_q)^2),\\
&M_3(q)=N_3(q)=\frac{1}{2}(1-(\gamma_q z_0)^2)((l\gamma_q)^2-(lq)^2),\\
&N_1(q)=M_1(q)\cosh 2\gamma_q L+M_2(q)\sinh 2\gamma_q L,\\
&N_2(q)=M_1(q)\sinh 2\gamma_q L+M_2(q)\cosh 2\gamma_q L,\\
&A(x)=(1+(xz_0)^2)\sinh xL+2xz_0\cosh xL,\\
&\gamma_q=\sqrt{q^2+\frac{3}{ll_a}},\quad \gamma_{\Delta\omega}=\sqrt{\frac{3}{l}\left(\frac{1}{l_a}+i\frac{\Delta\omega}{c}\right)}.
\end{aligned} \tag{B4}$$

The functions $F(x)$ and $G(x)$ in (B2) and (B3) are

$$\begin{aligned}
F(x)=\frac{1}{4l}\Bigg\{&(1+z_0^2(\alpha^2+\beta^2))\\
&\times\Bigg(\frac{x}{x^2-\alpha^2}\sinh 2xL\cosh 2\alpha L-\frac{\alpha}{x^2-\alpha^2}\cosh 2xL\sinh 2\alpha L\Bigg)\\
&+2z_0\alpha\Bigg(\frac{x}{x^2-\alpha^2}\sinh 2xL\sinh 2\alpha L-\frac{\alpha}{x^2-\alpha^2}(\cosh 2xL\cosh 2\alpha L-1)\Bigg)\\
&-(1-z_0^2(\alpha^2+\beta^2))\Bigg(\frac{x}{x^2+\beta^2}\sinh 2xL\cosh 2\beta L+\frac{\beta}{x^2+\beta^2}\cosh 2xL\sin 2\beta L\Bigg)\\
&+2z_0\beta\Bigg(\frac{x}{x^2+\beta^2}\sinh 2xL\sin 2\beta L-\frac{\beta}{x^2+\beta^2}(\cosh 2xL\cos 2\beta L-1)\Bigg)\Bigg\},
\end{aligned} \tag{B5}$$

$$\begin{aligned}
G(x)=\frac{1}{4l}\Bigg\{&(1+z_0^2(\alpha^2+\beta^2))\\
&\times\Bigg(\frac{x}{x^2-\alpha^2}(\cosh 2xL\cosh 2\alpha L-1)-\frac{\alpha}{x^2-\alpha^2}\sinh 2xL\sinh 2\alpha L\Bigg)\\
&+2z_0\alpha\Bigg(\frac{x}{x^2-\alpha^2}\cosh 2xL\sinh 2\alpha L-\frac{\alpha}{x^2-\alpha^2}\sinh 2xL\cosh 2\alpha L\Bigg)-(1-z_0^2(\alpha^2\\
&+\beta^2))\Bigg(\frac{x}{x^2+\beta^2}(\cosh 2xL\cos 2\beta L-1)+\frac{\beta}{x^2+\beta^2}\sinh 2xL\sin 2\beta L\Bigg)\\
&+2z_0\beta\Bigg(\frac{x}{x^2+\beta^2}\cosh 2xL\sin 2\beta L-\frac{\beta}{x^2+\beta^2}(\sinh 2xL\cos 2\beta L-1)\Bigg)\Bigg\},
\end{aligned} \tag{B6}$$

In (B5) and (B6) $\alpha=\mathrm{Re}\,\gamma_{\Delta\omega}$, and $\beta=\mathrm{Im}\,\gamma_{\Delta\omega}$.

[1] L. P. Gor'kov, A. I. Larkin, and D. E. Khmel'nitskiĭ, JETP Lett. **30**, 228 (1979).
[2] B. L. Al'tshuler, A. G. Aronov, A. I. Larkin, and D. E. Khmel'nitskiĭ, Zh. Éksp. Teor. Fiz. **81**, 768 (1981) [Sov. Phys. JETP **54**, 411 (1981)].
[3] G. Bergmann, Phys. Rev. B **28**, 2914 (1983); Phys. Rep. **107**, 1 (1984).
[4] B. L. Al'tshuler, JETP Lett. **41**, 648 (1985).
[5] Yu. A. Zyuzin and B. Z. Spivak, Zh. Éksp. Teor. Fiz. **93**, 994 (1987) [Sov. Phys. JETP **66**, 560 (1987)].
[6] Yu. N. Barabanenkov, Yu. A. Kravtsov, V. D. Ozrin, and A. I. Saichev, Prog. Opt. **29**, 67 (1991).
[7] V. L. Kuz'min and V. P. Romanov, Usp. Fiz. Nauk **166**, 247 (1996) [Phys. Usp. **39**, 231 (1996)].
[8] V. V. Sobolev, *Light Scattering in Planetary Atmospheres*, Pergamon Press, Oxford-New York (1975).
[9] V. V. Ivanov, Zh. **53**, 589 (1976) [Sov. Astron. **20**, 332 (1976)].
[10] J. X. Zhu, D. J. Pine, and D. A. Weitz, Phys. Rev. A **44**, 3948 (1991).
[11] T. M. Nieuwenhuizen and J. M. Luck, Phys. Rev. E **48**, 569 (1993).
[12] R. Berkovits and Sh. Feng, Phys. Rev. B **45**, 97 (1992).
[13] A. Lagendjik, R. Vreeker, and P. de Vries, Phys. Lett. A **136**, 81 (1989).
[14] I. Freund and R. Berkovits, Phys. Rev. B **41**, 496 (1990).
[15] N. Garcia, A. Z. Genack, and A. A. Lisyansky, Phys. Rev. B **46**, 14 475 (1992).
[16] A. A. Lisyansky and D. Livdan, Phys. Lett. A **170**, 53 (1992); Phys. Rev. B **47**, 14 157 (1993).
[17] M. C. W. van Rossum and Th. M. Nieuwenhuizen, Phys. Lett. A **177**, 452 (1993).
[18] I. Freund, J. Opt. Soc. Am. A **41**, 3274 (1994).
[19] J. H. Li and A. Z. Genack, Phys. Rev. E **49**, 4530 (1994).
[20] M. Ospeck and S. Fraden, Phys. Rev. E **49**, 4578 (1994).
[21] V. S. Potanov, Teor. Mat. Fiz. **100**, 287 (1994); Teor. Mat. Fiz. **100**, 424 (1994).
[22] Yu. N. Barabanenkov and V. D. Ozrin, Zh. Éksp. Teor. Fiz. **94**(6), 56 (1988) [Sov. Phys. JETP **67**, 1117 (1988)].
[23] D. B. Rogozkin and M. Yu. Cherkasov, JETP Lett. **58**, 585 (1993); Phys. Rev. B **51**, 12 256 (1995).
[24] K. M. Case and P. F. Zweifel, *Linear Transport Theory*, Addison-Wesley, Reading, Mass. (1967).
[25] H. C. van de Hulst, *Multiple Light Scattering*, Academic Press, New York (1980).
[26] M. B. van der Mark, M. P. van Albada, and A. Lagendjik, Phys. Rev. B **37**, 3575 (1988).
[27] R. Pnini and B. Shapiro, Phys. Rev. B **39**, 6986 (1989); Phys. Lett. A **157**, 265 (1991).
[28] A. F. Molich, B. P. Oekry, W. Schupita, B. Sumetsbergen, and G. Magerl, Phys. Rev. A **50**, 1581 (1994).
[29] E. E. Gorodnichev, S. L. Dudarev, and D. B. Rogozkin, Zh. Éksp. Teor.

Fiz. **96**, 847 (1989) [Sov. Phys. JETP **69**, 481 (1989)]; Phys. Lett. A **144**, 48 (1990).
[30] Z. D. Genchev, Physica B **175**, 405 (1991).
[31] D. B. Rogozkin, Zh. Éksp. Teor. Fiz. **111**, 1674 (1997) [JETP **84**, 916 (1997)].
[32] M. J. Stephen and G. Cwilich, Phys. Rev. Lett. **59**, 285 (1987).
[33] S. Feng, C. Kane, P. A. Lee, and A. D. Stone, Phys. Rev. Lett. **61**, 834 (1988).
[34] S. Hikami, Phys. Rev. B **24**, 2672 (1981).

Translated by P. Shelnitz

Crystal-field induced mixing of electron states in C_{60} crystals at high pressure

K. P. Meletov*) and V. K. Dolganov

Institute of Solid State Physics, 142432 Chernogolovka, Moscow Region, Russia
(Submitted 9 July 1997)
Zh. Éksp. Teor. Fiz. **113**, 313–322 (January 1998)

Optical absorption spectra of thin fullerene (C_{60}) crystals in the range 1.7 to 3.8 eV have been measured at $T=300$ K and at pressures up to 2.5 GPa. The spectrum shifts toward the red with pressure, and the electron absorption intensity is redistributed among its bands. The intensity of the band associated with the lowest direct electron interband transition monotonically increases with pressure, whereas the intensity of the upper interband feature decreases. Bands related to weak edge absorption in the range between 1.7 and 2.2 eV gradually merge with the band associated with the lowest interband transition, whose intensity rises with pressure. A similar redistribution of intensity among electron transition bands has been observed when comparing the spectrum of an isolated C_{60} molecule and that of a C_{60} crystal. The results indicate that the crystal-field induced mixing of electron states is present in solid C_{60}, and they can be discussed in terms of the Craig–McClure model, which was suggested to describe crystal-field induced mixing of electron states in anthracene and naphthalene molecular crystals. [S1063-7761(98)02101-5]

1. INTRODUCTION

Electronic spectra of C_{60} molecules and C_{60} crystals have been discussed in numerous theoretical and experimental works. The first calculations of C_{60} molecule electron spectra were reported soon after the discovery of stable large all-carbon molecules in products of graphite laser ablation.[1–3] Then followed more accurate quantum mechanical models of the C_{60} molecular structure, electron and phonon spectra.[4–8] Experimental study of the C_{60} electron spectrum became a real possibility after the efficient technique for synthesizing fullerenes in a graphite arc had come into being.[9–12] Detailed measurements of C_{60} molecular absorption spectra and their vibrational analysis were performed for solid solutions of C_{60} in argon and hexane matrices by means of high-resolution Fourier-transform spectroscopy.[13,14] This research indicated that the range of weak edge absorption in the C_{60} molecular spectrum, 1.78 to 3.04 eV, is due to electron–phonon replicas of dipole-forbidden electron transitions. These transitions, assisted by non-totally-symmetric, phonons are due to the Hertzberg–Teller and Jahn–Teller effects, and also show up in absorption spectra of C_{60} films and C_{60} thin crystals.[15,16] Dipole-allowed electron transitions have been detected above 3 eV, and the most intense of them generate spectral features at 4.84 and 5.88 eV.[14]

The recorded absorption spectrum of the C_{60} molecule is in very good agreement with theoretical calculations of its electron spectrum, although certain differences in numerical parameters are present.[8,14] Optical absorption spectra of C_{60} crystals are similar to the C_{60} molecule absorption spectrum, overall, which is quite common in molecular crystals. Essentially all features of the C_{60} crystal absorption spectrum can be also seen in the molecule spectrum, but shifted to the low-energy side. This shift in molecular crystal spectra is called a crystalline shift, and its value, which is proportional to the transition intensity, can be up to hundreds of millielectronvolts.

Meanwhile, intensities of absorption bands in spectra of C_{60} crystals and C_{60} molecules in both gas-phase and solid solutions are notably different.[17] The major difference is a broad absorption band between 2.2 and 3.1 eV in the crystal spectrum. Its intensity is notably higher than those of weak edge features, but, it is weaker than the dipole-allowed absorption features on the high-energy side.

Attention has been focused on the origin of this band in several studies, and the most common view is that the band is due to direct transitions between the valence and conduction bands. The considerable width of the band is thought to be due to splitting of degenerate electron states of both valence and conduction bands of the crystal. Note that presently we have no convincing evidence that would allow us to attribute the band to a specific transition in the molecular spectrum. Since there are no transitions of such intensity in this range of the electron spectrum, the band is thought to be connected with forbidden molecular transitions that are allowed in the crystal due to a change in molecular positional symmetry.

We suggest, however, that this band is due to a weak dipole-allowed transition in the molecular spectrum whose intensity increases considerably in a crystal owing to its mixing with a stronger electron transition on the high-energy side. Similar mixing of the molecule electron states due to the crystal field was detected in exciton absorption spectra of anthracene and naphthalene molecular crystals, and described by Craig[18] and McClure and Schnepp.[19] Later research demonstrated[20] that the band intensity in this case is also changed considerably by hydrostatic pressure applied to a crystal, since it intensifies the interaction between molecules and the effect of the crystal field.

In view of this, we have investigated in detail the band intensities in electron absorption spectra of C_{60} crystals under hydrostatic pressure. We have recorded absorption spectra of C_{60} thin crystals in the 1.7–3.8 eV range at $T=300$ K and at pressures up to 2.5 GPa. Alongside the red shift of

 1063-7761/98/010177-05$15.00

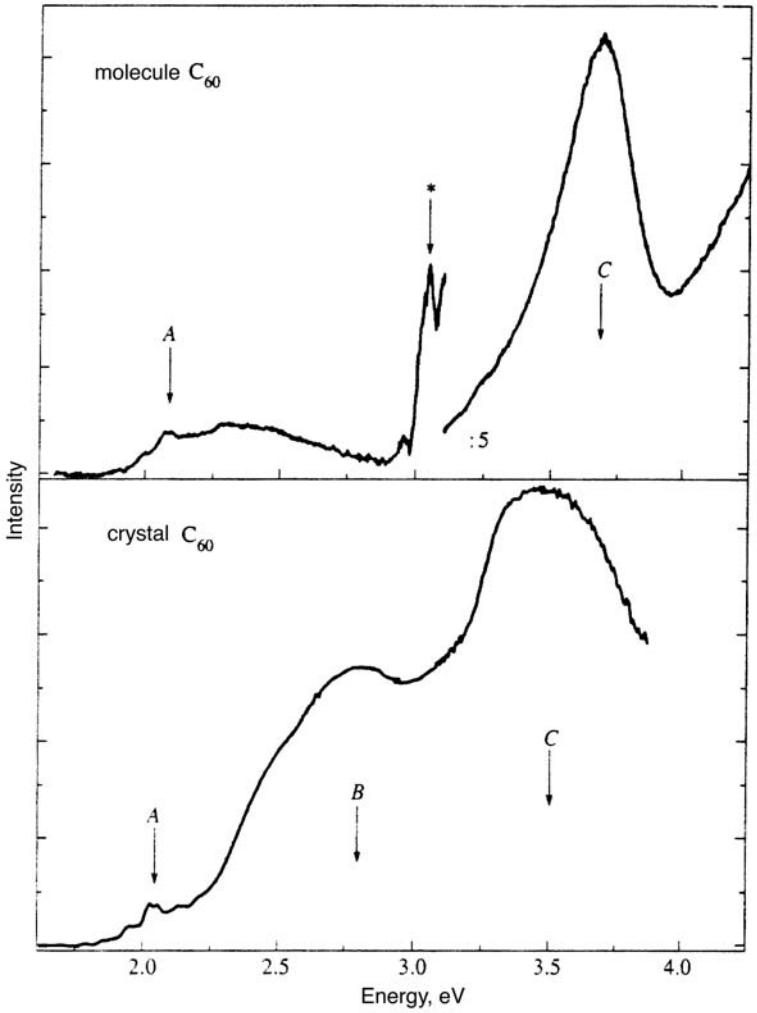


FIG. 1. Absorption spectra of a C_{60} crystal with a thickness of less than 1 μm and of C_{60} solution in toluene at $T=77$ K and standard pressure.

spectral features due to pressure, which was detected earlier, notable redistribution of their intensities has been observed. The intensity of the band corresponding to the lowest direct interband transition rises monotonically with the pressure, whereas the intensity of the interband transition of the higher energy decreases. Similar changes in the intensities of these features were detected in the absorption spectrum of the C_{60} molecules dissolved in toluene in comparison with that of the C_{60} crystal. These results will be discussed in terms of the crystal-field induced mixing of molecular electron states suggested by Craig and McClure.

2. EXPERIMENT

Absorption spectra were measured using C_{60} thin crystals grown from either a toluene solution or a gaseous phase. The initial material purity was higher than 99%. We selected for measurements crystals shaped as thin plates with lateral dimensions of about 100×100 μm and a thickness less than 1 μm. The crystal thickness was derived from the optical density obtained by comparison to the optical density of C_{60} thin films.[15]

Absorption spectra were recorded on a computer-controlled spectrometer built around an MDR-23 grating monochromator. Measurements were taken in a broad spectral region in the visible and ultraviolet light using high-quality achromatic quartz optics. The image of a crystal placed into a high-pressure cell was magnified by a factor of about 80, processed to an intermediate plane defined by two crossed 200×200 μm optical slits, and projected to the monochromator input slit. The crystal optical density was derived from transmission spectra normalized to the transmission spectra of the diamond anvils, which were measured at all pressures. Thus the contribution from the edge absorption of the diamond anvils, which notably changed with the pressure, could be eliminated.

High-pressure experiments were performed using a high-pressure diamond anvil cell of the Merril–Basset type.[21] The working area of the anvils had a diameter of 600 μm, and the aperture diameter of the stainless steel gasket was 250 μm. The pressure-transmitting medium was a 1:4 mixture of ethanol and methanol. The pressure gauge was the position of the R_1 line in the ruby luminescence spectrum.[22]

3. RESULTS AND DISCUSSION

The absorption spectrum at standard pressure and $T=77$ K of a C_{60} crystal with a thickness less than 1 μm is given in the lower half of Fig. 1. On the low-energy edge, one can see a weak absorption band ranging between 1.78 and 2.2 eV (A-band) with clearly defined fine structure. Next is a more intense absorption band running from 2.2 to 3.0 eV

(*B*-band), and on the high-energy side there is an intense *C*-band peaking at 3.4 eV. The two most intensive bands are in the region of a diamond anvil self-absorption, its lower bound being at ~ 4 eV.

The upper half of Fig. 1 shows an absorption spectrum of C_{60} dissolved in toluene plotted on the same scale. The graph clearly shows that the two spectra are, by and large, similar, although there are notable differences between them. Above all, the molecular spectrum is shifted to the high-energy side with respect to the crystal spectrum. The shift is about 0.04 eV for the fine structure of the relatively weak *A*-band, and about 0.4 eV for the strong *C*-band. Moreover, the *C*-band is considerably broader in the crystal spectrum than in the molecule spectrum. The most striking difference between the two spectra is that the crystal spectrum contains the *B*-band, which peaks at 2.8 eV, whereas no such feature can be seen in the molecule spectrum.

In this spectral range, the molecule shows the weak and narrow feature marked by an asterisk in Fig. 1. The intensity distributions are also notably different: the *C*-band intensity in the crystal spectrum is much lower than in the molecule spectrum. Thus, the intensity is substantially redistributed among the absorption bands in the transition from isolated molecules to the crystal: that of the *C*-band drops considerably, and the *B*-band intensity increases. Note that this fact was discussed in earlier studies of C_{60} optical absorption spectra. Several interpretations have been suggested, and the most popular of them is the hypothesis that electron transitions forbidden by molecular symmetry are partially allowed in the crystal, since the molecule is deformed by the local crystal field. Unsuccessful attempts were also made to simulate this effect in absorption spectra of C_{60} solutions by taking solvents with a higher polarity.

Detailed measurements of optical absorption spectra of C_{60} thin crystals in the visible and UV at high pressure show that a redistribution of intensity among the electron transition bands also occurs when crystals are compressed hydrostatically. Figure 2 shows absorption spectra of a C_{60} thin crystal at $T=300$ K, at pressures ranging from one atmosphere in the lower graph to 2.5 GPa in the upper diagram. The *C*-band intensity gradually decreases with pressure, and the *B*-band intensity increases. The spectrum is simultaneously shifted toward lower energies, and the bands become appreciably wider. These changes are reversible, and when the pressure is lifted, all typcial features of the initial spectrum come back. Note that the red shift of the C_{60} absorption bands due to high pressure was previously investigated in several experiments.[16,23,24] In those experiments, a notable increase in the *B*-band intensity in absorption spectra of C_{60} thin crystals was detected,[16] as in independent measurements of C_{60} thin film spectra at high pressure.[23] Unfortunately, this effect was not given due attention in either of these works.

Figure 3 shows the ratio between the total intensities of the *C*- and *B*-bands, I_C/I_B, due to electron absorption in a C_{60} crystal as a function of pressure. In ascertaining this behavior, we isolated the bands, fitting a Gaussian to each. The *B*-band has a side band on its low-energy edge around 2.5 eV under standard pressure. To obtain a better fit, we approximated the *B*-band as a sum of two Gaussians, and its

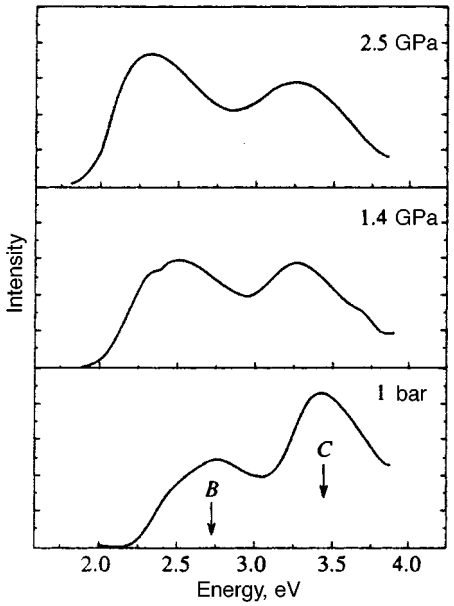


FIG. 2. Absorption spectra of a C_{60} crystal with a thickness of less than 1 μm at $T=300$ K and pressures of up to 2.5 GPa.

total intensity was the sum of their intensities. Empty circles in Fig. 3 represent the ratio of the total intensities, I_C/I_B, measured under increasing pressure, and filled circles show measurements taken as the pressure was decreased. The intensity ratio I_C/I_B drops by a factor of more than two as the pressure grows to 2.5 GPa, but returns to its original value when the pressure is lifted.

The intensity of the *A*-band in the region of weak edge absorption is essentially independent of pressure. This is hard to see in the absorption spectra of the thin crystal shown in Fig. 2, since the *A*-band intensity is very low in this case. Absorption spectra of a crystal with a thickness of 2.8 μm at

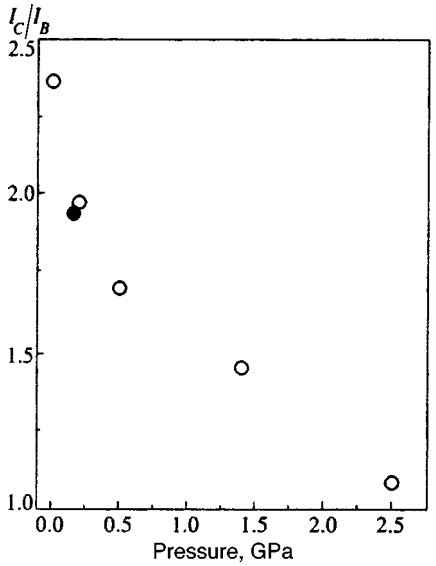


FIG. 3. Ratio of total intensities of the *C*- and *B*-bands, I_C/I_B, in the absorption spectrum of the C_{60} crystal. Empty circles show measurements taken under increasing pressure, and filled circle corresponds to decreasing pressure.

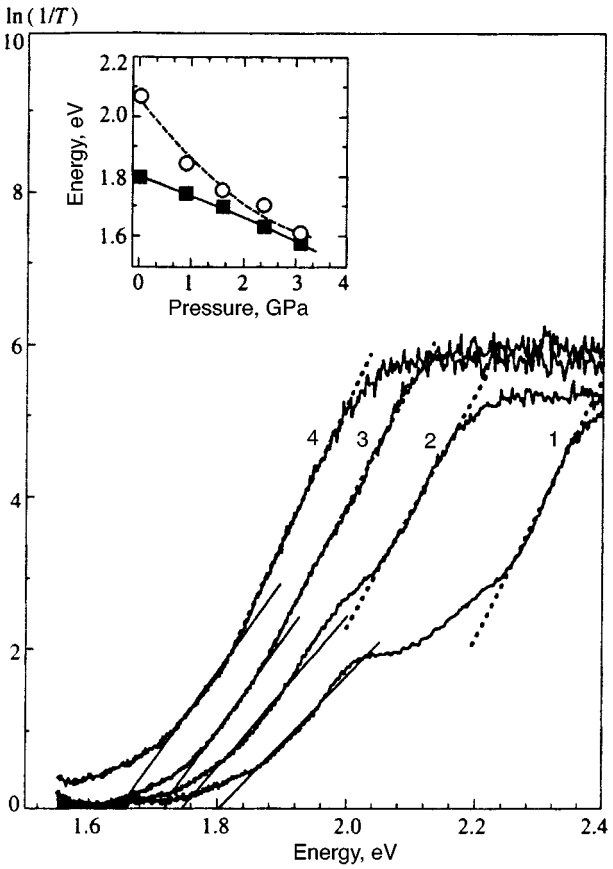


FIG. 4. Absorption spectra of a C_{60} crystal with a thickness of 2.8 μm in the region of the A-band at $T=300$ K and various pressures: 1) 1 bar, 2) 0.9 GPa, 3) 1.6 Gpa, and 4) 2.4 GPa. The insert shows positions of the edges of A-band (filled squares) and B-band (empty circles) vs pressure.

$T=300$ K and pressures up to 2.4 GPa are plotted in Fig. 4. They clearly show the A-band and the low-energy edge of the B-band, whereas the peak of the B-band and the C-band are in the region of total absorption. At room temperature the fine structure of the A-band is essentially invisible, and only a sideband can be seen in this spectral region. The B-band intensity increases with pressure, and its high-energy edge shifts towards the A-band and suppresses it. The high-energy edge of the A-band shifts to the red much more slowly than the corresponding edge of the B-band. Positions of the edges of the A- and B-bands vs pressure are shown in the insert to Fig. 4. The empty symbols correspond to the the B-band edge, defined as the crossing point between the tangent to the spectral profile and the horizontal axis. Filled symbols corresponds to the A-band edge, which is defined similarly. At a pressure of 3 GPa, the A-band is not observable, and edge absorption is due entirely to the B-band. These results indicate different origins of the A- and B-bands in the absorption spectrum of C_{60} crystals. A similar conclusion was derived earlier from analysis of electron–phonon bands in the absorption spectra of C_{60} molecules.[14]

Thus, the distinctive feature of absorption spectra of C_{60} crystals is the broad and intense B-band, which has no analog in the molecular absorption spectrum. At the same time, the C-band intensity in the crystal spectrum is much lower than in the molecular spectrum. Under hydrostatic pressure, the C-band intensity drops further, while the B-band intensity, simultaneously increases; these changes are reversed when the pressure is lifted. The behavior of C_{60} crystal absorption spectra can be interpreted in terms of the crystal-field-induced configurational mixing of electron states in molecular crystals. As shown by Craig and McClure, this effect shows up most clearly when electron states with close energies and substantially different oscillator strengths of the respective optical transitions are mixed.[18,19] This is the case in the electron absorption spectrum of the C_{60} molecule, where the lowest allowed electron transition between states with symmetry $^1A_g \rightarrow {}^1T_u$, whose energy is 3.04 eV and oscillator strength $f \simeq 0.015$ ($*$-band), is much weaker than the higher-energy transition at 3.78 eV and oscillator strength $f \simeq 0.37$.[14] The information concerning classification of electron transitions in the C_{60} molecule given above is based on the analysis of electron–phonon spectra of frozen C_{60} solutions in hexane and trimethylpentane at $T=77$ K,[14] and is in good agreement with the most detailed calculations of the molecule's electron spectrum.[8]

We now examine in more detail Craig's discussion of the effect.[18] The wave function of the lowest excited electron state of the crystal is described in the first order of perturbation theory with respect to the crystal field by the expression

$$\Psi_1^1=\Phi_1^0+\{H^{12}/(H^{11}-H^{22})\}\Phi_2^0+\ldots + \{H^{1r}/(H^{11}-H^{rr})\}\Phi_r^0+\ldots, \quad (1)$$

where $\Phi_1^0 \ldots \Phi_r^0$ are unperturbed wave functions of the various electron states of the same symmetry, $H^{11} \ldots H^{rr}$ are the energies of these states, and $H^{12} \ldots H^{1r}$ are the energies of interaction due to the crystal field. Equation (1) holds if

$$H^{rr}-H^{11} \gg H^{1r}, \quad (2)$$

which is a prerequisite for perturbation theory to be applicable, and holds *a fortiori* for most electron transitions in molecular crystals. In this case the energy of the lowest excited state in first-order perturbation theory is

$$\Delta E=H^{11}+\sum_r{}' (H^{1r})^2/(H^{11}-H^{rr}). \quad (3)$$

The transition matrix element in first-order perturbation theory is

$$M_1^1=M_1^0+\sum_r{}' \{H^{1r}/(H^{11}-H^{rr})\}M_r^0, \quad (4)$$

where $M_1^0 \ldots M_r^0$ are the matrix elements of electron transitions between unperturbed states. It follows from Eq. (4) that the effect of the crystal-field-induced configurational mixing of electron states on the intensity and energy of electron transition is the stronger, the higher the interaction energy H^{1r} between electron states, the larger the difference between matrix elements M_1^0 and M_r^0, and the smaller the difference between the energies of the electron transitions, $H^{11}-H^{rr}$. This effect determines the changes in the intensities of electron transitions within a molecule when a molecular crystal is formed, and subsequent changes when the crystal undergoes hydrostatic compression. In the latter case, the

interaction energy H^{1r} increases and $H^{11}-H^{rr}$ drops owing to the difference between the rates of pressure-induced shifts of electron absorption bands.[20]

Note that first-principles calculations of the electron spectrum of the C_{60} crystal indicate that similar changes should occur in electron absorption bands of the molecule.[25] The calculations also predict an intense and wide *B*-band, which has not been detected in the molecule absorption spectra. Its calculated intensity is much higher than the *B*-band intensity in experimental absorption spectra of C_{60} thin films, whereas the *C*-band is absent in the calculations. The intensities of two electron absorption bands above the *C*-band are considerably lower than in experimental absorption spectra. Calculations of the C_{60} crystal absorption spectrum at high pressure by the same authors indicate that the width and amplitude of the *B*-band should increase with pressure,[26] whereas the intensities of electron absorption bands of higher energies should continuously decrease.

Our aggregate results suggest that the lowest interband transition between the top of the valence band and bottom of the conduction band in the C_{60} crystal is related to the *B*-band. The weak edge absorption in the region of the *A*-band has a different nature and is related, as in the case of molecular absorption, to electron–phonon replicas of forbidden electron transitions. This means that the direct band gap in the C_{60} crystal is determined by the position of the long-wave edge of the *B*-band, and is about 2.1 eV at atmospheric pressure. The pressure dependence of the band gap can be derived from the position of the *B*-band long-wave edge, and is shown in the insert to Fig. 4 (empty circles). The rate of the *B*-band pressure shift, dE_g/dP, is -0.15 ± 0.01 eV/GPa at standard pressure, and yields the deformation potential of the C_{60} crystal in accordance with the formula

$$D=dE_g/d\{\ln(V_0/V)\}=-B_0(dE_g/dP), \tag{5}$$

where $B_0=18.1\pm1.8$ GPa is the bulk modulus of the C_{60} crystal derived from x-ray diffraction measurements at high pressure.[27,28] The deformation potential of C_{60} derived from these data is 2.7 ± 0.3 eV. It is about twice the value given in Ref. 28, where the direct gap width was erroneously deduced from the position of the *A*-band edge. Note that our measurement of the deformation potential is in better agreement with the numerical calculation,[26] which yields 3.1 eV.

Thus, our experimental data suggest crystal-field-induced mixing of electron states in the C_{60} crystal. Crystal-field-induced mixing is the primary reason for significant redistribution of intensity among electron absorption bands, due both to formation of a crystal from molecules and hydrostatic compression of the former. The presently available data lead us to claim that the *B*-band in the absorption spectrum of a C_{60} crystal corresponds to the lowest interband transition, and the position of its long-wave edge yields a direct band gap of 2.1 eV. The deformation potential derived from the pressure-induced shift of the *B*-band edge is 2.7 ± 0.3 eV.

In conclusion, we would like to express our gratitude to I. N. Kremenskaya and R. K. Nikolaev for supplying samples of C_{60} crystals and films. We are indebted to the Russian Fund for Fundamental Research (Project 96-02-17489) for partial support of this work, and to the *Fullerenes and Atomic Clusters* state-sponsored program (Project No. 97-016) and NATO Research Committee for support within the *Collaboration Research* program (Grant No. 96-0556).

*)e-mail: mele@issp.ac.ru

[1] H. M. Kroto, J. R. Heath, S. C. O'Brien, R. F. Curl, and R. E. Smalley, Nature **318**, 162 (1985).
[2] R. C. Haddon, L. E. Brus, and K. Raghavachari, Chem. Phys. Lett. **125**, 459 (1986).
[3] M. Ozaki and A. Takahashi, Chem. Phys. Lett. **127**, 242 (1986).
[4] S. Larsson, A. Volosov, and A. Rosen, Chem. Phys. Lett. **137**, 501 (1987).
[5] I. Laszlo and L. Udvardi, Chem. Phys. Lett. **136**, 418 (1987).
[6] F. Negri, C. Orlandi, and F. Zerbetto, Chem. Phys. Lett. **144**, 31 (1988).
[7] J. Feng, J. Li, Z. Wang, and M. C. Zemmer, Int. J. Quantum Chem. **37**, 599 (1990).
[8] M. Braga, S. Larsson, A. Rosen, and A. Volosov, Astron. Astrophys. **245**, 232 (1991).
[9] W. Kratschmer, K. Fostiropoulos, and D. R. Huffman, Chem. Phys. Lett. **170**, 167 (1990).
[10] W. Kratschmer, L. F. Lamb, K. Fostiropoulos, and D. R. Huffman, Nature **347**, 354 (1990).
[11] J. P. Hare, H. W. Kroto, and R. Taylor, Chem. Phys. Lett. **177**, 394 (1991).
[12] H. Ajie, M. M. Alvarez, S. J. Anz, R. D. Beck, K. Fostiropoulos, F. Diederich, D. R. Huffman, M. Kratschmer, Y. Rubin, K. E. Schriver, D. Sensharma, and R. L. Whetten, J. Phys. Chem. **94**, 8630 (1990).
[13] Z. Casyna, P. N. Schatz, J. P. Hare, T. J. Dennis, H. W. Kroto, R. Taylor, and D. R. M. Walton, Chem. Phys. Lett. **183**, 283 (1991).
[14] S. Leach, M. Vervloet, A. Despres, E. Breheret, J. P. Hare, H. W. Kroto, T. J. Dennis, R. Taylor, and D. R. M. Malton, Chem. Phys. **160**, 451 (1992).
[15] C. Reber, L. Yee, J. McKiernan, J. I. Zink, R. S. Williams, W. M. Tong, D. A. A. Ohlberg, R. L. Whetten, and F. Diederich, J. Am. Chem. Soc. **95**, 2127 (1991).
[16] K. P. Meletov, V. K. Dolganov, O. V. Zharikov, I. N. Kremenskaya, and Yu. A. Ossip'yan, J. Phys. I **2**, 2097 (1992).
[17] M. Partini, F. Marabelle, and G. Gruizetti, in *Fullerenes*, K. M. Kadish and R. S. Ruoff (eds.), Electrochemical Society Inc., Pennington USA (1994), p. 639.
[18] D. P. Craig, J. Chem. Soc. , 2302 (1955).
[19] D. S. McClure and O. Schnepp, J. Chem. Phys. **23**, 1575 (1955).
[20] K. P. Meletov, Chem. Phys. Lett. **172**, 147 (1990).
[21] A. Jayaraman, Rev. Sci. Instrum. **44**, 1 (1973).
[22] D. Barnett, S. Block, and G. J. Piermarini, Rev. Sci. Instrum. **57**, 1013 (1986).
[23] F. Moshary, N. Chen, and I. Silvera, Phys. Rev. Lett. **69**, 466 (1992).
[24] D. W. Snoke, K. Syassen, and A. M. Mittelbach, Phys. Rev. B **47**, 4146 (1993).
[25] W. Y. Ching, Ming-Zhu Huang, Yong-Nian Xu, W. C. Harter, and F. T. Chan, Phys. Rev. Lett. **67**, 2045 (1991).
[26] Yong-Nian Xu, Ming-Zhu Huang and W. Y. Ching, Phys. Rev. B **46**, 4241 (1992).
[27] S. J. Duclos, K. Brister, R. C. Haddon *et al.*, Nature **351**, 380 (1991).
[28] K. P. Meletov, G. Kourouklis, D. Christofilos, and S. Ves, Zh. Éksp. Teor. Fiz. **108**, 1456 (1995) [JETP **81**, 798 (1995)].

Translation provided by the Russian Editorial office

Thermopower in quasi-two-dimensional $(BEDT\text{-}TTF)_mX_n$ organic conductors

S. V. Demishev, M. V. Kondrin, V. V. Glushkov, N. E. Sluchanko, and N. A. Samarin

Institute of General Physics, Russian Academy of Sciences, 117942 Moscow, Russia
(Submitted 11 July 1997)
Zh. Éksp. Teor. Fiz. **113**, 323–338 (January 1998)

Thermopower of $(BEDT\text{-}TTF)_mX_n$ organic conductors has been studied using a dedicated measurement technique in the temperature range of 4.2 to 300 K. It turned out that some features of the thermopower in quasi-two-dimensional metals, namely the presence of a peak in the thermopower of α-$(BEDT\text{-}TTF)_2MHg(SCN)_4$ and a plateau in κ-$(BEDT\text{-}TTF)_2Cu(NCS)_2$ in the temperature interval between 10 and 50 K, are probably due to the phonon drag effect. Similar temperature dependences of the Seebeck coefficient can be satisfactorily interpreted in terms of a simple model taking into account the real experimental curve of the phonon heat capacity versus temperature, $C \propto T^2$, which is not described by the Debye formula. One feature distinguishing organic superconductors from magnetically ordered metals is a stronger temperature dependence of the characteristic electron–phonon scattering time $\tau_{e-ph}(T)$. Phonon drag effects also determine the behavior of the thermopower in the $(BEDT\text{-}TTF)_3Cl_2\cdot 2H_2O$ organic conductor, which is characterized by a metal–insulator transition at $T \sim 150$ K. An analysis of measurements of the conductivity and thermopower vs. temperature taken together indicates that the transition in this compound has a complex nature: first (at $T \sim 150$ K) a metal–insulator transition occurs, which produces an energy gap in the band spectrum, then at a lower temperature ($T \sim 20$ K) a transition to a charge-density wave state takes place. © *1998 American Institute of Physics.* [S1063-7761(98)02201-X]

1. INTRODUCTION

In recent years, physical properties of $(BEDT\text{-}TTF)_mX_n$ organic conductors (where BEDT-TTF denotes bis(ethylenedithio)-tetrathiafulvalene, and X is an anion) have been studied intensely using a variety of experimental techniques. One feature of these materials is the presence of the complex-structure layers of the BEDT-TTF molecules separated by anion layers, with electrons confined to BEDT-TTF layers forming an essentially perfect two-dimensional gas (at low temperatures the in-plane conductivity can be several orders of magnitude higher than the conductivity in the perpendicular direction).[1]

Depending on the form of the anion in $(BEDT\text{-}TTF)_mX_n$, various types of ground states occur at low temperatures. Apart from normal metals and superconductors with one- or two-dimensional Fermi surfaces,[2] there are materials in the $(BEDT\text{-}TTF)_mX_n$ group that are characterized by low-dimensional instabilities like charge-density waves, which lead to an insulator ground state,[3–5] or spin-density waves, which modify the metallic ground state owing to magnetic ordering.[6,7]

Until the present time, efforts in the field of electron properties have been largely focused on characterization of Fermi surfaces and studies of interaction in the electron subsystem using quantum oscillation techniques[8–10] or the cyclotron resonance,[11,12] whereas their superconducting properties and metal–insulator transition due to charge-density waves were studied mostly by traditional galvanomagnetic methods.[3–5]

Thermoelectric effects in $(BEDT\text{-}TTF)_mX_n$ have not been studied systematically as yet, and publications on this topic have appeared only occasionally (see, for example, Refs. 13 and 14), and discussion of experimental results was usually limited to determination of the effect sign. Note that the progress in this field is impeded largely by technical difficulties deriving from small dimension and frangibility of single crystals of organic conductors.

At the same time, it is well known that measurements of the Seebeck coefficient as a function of temperature, $S(T)$, yield important information concerning not only parameters of charge carriers, but also the nature of the electron–phonon scattering in the system,[15–17] which is especially interesting in the case of organic conductors, where strong Fermi-liquid effects have been anticipated.[6–12]

Therefore the aims of the reported work were, firstly, to develop a technique of precision thermopower measurements with due account of specific features of organic conductors and, secondly, to measure the Seebeck coefficient vs. temperature in various species of $(BEDT\text{-}TTF)_mX_n$, including superconductors, magnetically ordered metals, and systems that undergo a metal–dielectric transition.

2. EXPERIMENTAL TECHNIQUES

2.1. Samples

For our experiments, we selected $(BEDT\text{-}TTF)_2MHg(SCN)_4$ (where M=Tl, Rb, NH_4), $(BEDT\text{-}TTF)_2Cu(NCS)_2$, and $(BEDT\text{-}TTF)_3Cl_2\cdot 2H_2O$ single crystals. All organic conductors were synthesized by M. Kurmoo at the Royal Institution, UK. We investigated five single crystals of each composition with typical lateral dimensions of 1.5 to 2 mm. Measurements of thermopower and resistivity were conducted along axes of the largest growth (along the b-axis for $(BEDT\text{-}TTF)_2Cu(NCS)_2$ and c-axis for $(BEDT\text{-}TTF)_2MHg(SCN)_4$). The samples of

 1063-7761/98/010182-08$15.00

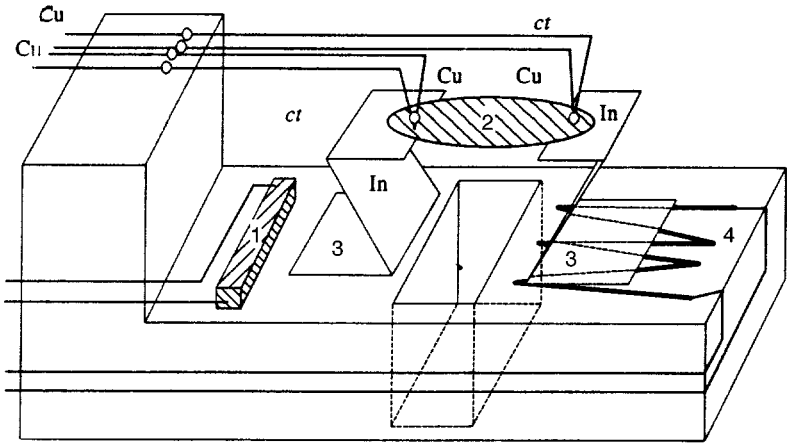


FIG. 1. Diagram of the low-temperature section of the facility for thermopower measurements.

$(BEDT\text{-}TTF)_2MHg(SCN)_4$ with M=Tl, Rb are organic metals whose Fermi surfaces contain both one- and two-dimensional sections,[2] and the presence of one-dimensional sections leads to instabilities like spin-density waves,[6–10] so that at $T<10$ K these metals are magnetically ordered. The samples of $(BEDT\text{-}TTF)_2NH_4Hg(SCN)_4$ and $(BEDT\text{-}TTF)_2Cu(NCS)_2$ are superconductors with $T_c\approx 1$ K[18] and $T_c\approx 11$ K,[14,19] respectively, the latter being a record for quasi-two-dimensional organic conductors. Note that the $(BEDT\text{-}TTF)_2MHg(SCN)_4$ compounds belong to a single structure class (so-called α-phases) with identical molecule packing patterns in the BEDT-TTF layers, whereas the superconductor $(BEDT\text{-}TTF)_2Cu(NCS)_2$ belongs to κ-phases. Note also that essentially all organic conductors undergo structural changes around $T\sim 100$ K associated with changes in BEDT-TTF molecule packing patterns in the corresponding layers. For this reason, a metallic state often comes at $T<100$ K,[1,2] and in discussing the metallic systems we will focus our attention on this temperature range.

One feature of the $(BEDT\text{-}TTF)_3Cl_2\cdot 2H_2O$ organic conductor is that metallic conductivity at $T>150$ K is replaced by semiconducting conductivity at lower temperatures; reducing the temperature to $T\approx 10$ K increases the resistivity by a factor of about 10^5.[3] It was suggested in early publications that the reason for the metal–insulator transition in this material is a change in the energy band pattern in the original semimetallic state, leading to formation of a band gap at $T\sim 150$ K.[20] A detailed analysis of the temperature dependence of the conductivity and magnetoresistance,[3] however, allowed the researchers to attribute the metal–insulator transition to formation of charge-density waves, which was later supported by structural analysis.[4,5]

2.2. Thermopower measuring technique

From the technical viewpoint, the major problem in measurements of thermopower in organic conductors is fabrication of good thermal contacts with two opposite sample faces, which is a *sine qua non* for accurate measurements of the temperature gradient. Since samples of organic conductors are very brittle, we had to design a dedicated fixture for supporting samples that would allow us to cool samples down without destruction due to thermal contraction of the sample holder. Note that single crystals of organic conductors have very small dimensions, so we had to select single crystals that were as long as possible—about 1.5–2.5 mm.

Another serious difficulty is conducting measurements strictly in the range of linear response, which can be difficult in the case of a metal-like sample, whose thermopower decreases with temperature. When the conventional technique is used, one has to increase the temperature difference ΔT across the sample in order to achieve the required voltage sensitivity.

We used in the thermopower measurements a low-temperature helium cryostat in which a sample was mounted on a cold plate in the vacuum jacket. By varying the liquid-helium flow rate, we could tune the temperature in the range between 1.5 and 300 K, and an electronic control system driving a heating system stabilized the prescribed temperature to within 0.05 K. A diagram showing the design of the cell for measuring thermopower is given in Fig. 1.

The cell temperature was measured by resistance thermometer 1. Organic metal sample 2 was glued to electrically insulated Z-shaped stripes 3 of thin indium foil, which maintained the sample in a strain-free state during alteration of the cell linear dimensions due to temperature variations. A temperature gradient across the sample was generated by a small heater 4. In order to reduce the heat flow, a large opening was milled in the copper cold plate supporting the assembly (Fig. 1). The thermopower generated in the sample was picked up using two copper and two constantan contacts. The sites where copper wires were soldered to the contact pads were attached to the sample holder by a thermally conducting glue and electrically insulated by thin capacitor paper. This allowed us to minimize the temperature difference between the constantan/copper junctions and thereby get rid of parasitic thermopower. In order to obtain equal temperature gradients between copper and constantan contacts on the sample, the free ends of the wires were soldered pairwise (one copper wire to one constantan wire), pressed to the ''hot'' and ''cold'' edges of the sample, and cemented by an electrically conducting glue. This technique yielded highly reliable electric and thermal contacts to the sample, and in testing experiments the assembly withstood no less than ten cooling/heating cycles without a notable degradation of sample properties and parameters of electric contacts. Mea-

surements on different single crystals of the same materials demonstrated that the thermopower could vary by 5–10%, which is also an indication of fairly good reproducibility and reliability of thermal and electric contacts fabricated by this method.

The electric signals were picked up and experimental parameters were driven by a dedicated facility build around a computer. All electric voltages were measured by Shch31 digital voltmeters with an input impedance of at least 100 MΩ, and after data processing the signal amplitude sensitivity was about 10 nV. The thermopower was measured as follows. Initially the plate temperature was stabilized at a prescribed value. Then a constant current, whose magnitude could be varied by the computer code, was fed to the heater. When a steady temperature gradient was obtained, thermopower signals from the copper, U_{Cu}, and constantan, U_{ct}, contacts were simultaneously recorded, after which a new heater current amplitude was set, and the measurement cycle repeated. In the linear range of the thermoelectric response, the measurement data plotted in coordinates U_{Cu} vs. U_{ct} formed a straight line, whose slope was calculated using the least squares method. The number of measurement cycles and maximum heater current were selected so that, firstly, the response was linear in the temperature gradient, $U \sim \Delta T$, and, secondly, the number of points was sufficient to keep the maximum absolute error in the least squares method within 3–5%. Under these conditions, the temperature difference across the sample was 1.5–2 K and decreased considerably in the range close to the liquid-helium temperature. Note that the time required to accumulate data needed for plotting the curve of $U=f(\Delta T)$ around the liquid-helium temperature was several tens of minutes. The total time of the heating/cooling cycle during which an $S(T)$ curve was plotted was about fifteen hours.

Since the slope is given by

$$\tan\,\alpha=\frac{S_{\mathrm{Cu}}-S_{\mathrm{sample}}}{S_{\mathrm{ct}}-S_{\mathrm{sample}}},$$

the Seebeck coefficient of the sample, S_{sample}, with respect to one contact (in this specific case, copper) can be calculated, given the calibration curve of the difference between the thermopowers of two contact materials (in fact, the calibration of the copper/constantan thermocouple). The copper contacts were fabricated from high-purity wire, whose thermopower at $T<6$ K was additionally tested using a lead sample as a reference (we used materials manufactured by Goodfellows, Great Britain).

Note that this experimental technique has certain advantages over the traditional differential method, since it allows one to get round the unknown temperature difference between the temperature gradient sensor and the sample-wire contact and to check that measurements are performed in the linear signal range. A similar technique was suggested by Laurent *et al.*,[21] but they did not vary the temperature gradient, as was done in the reported paper.

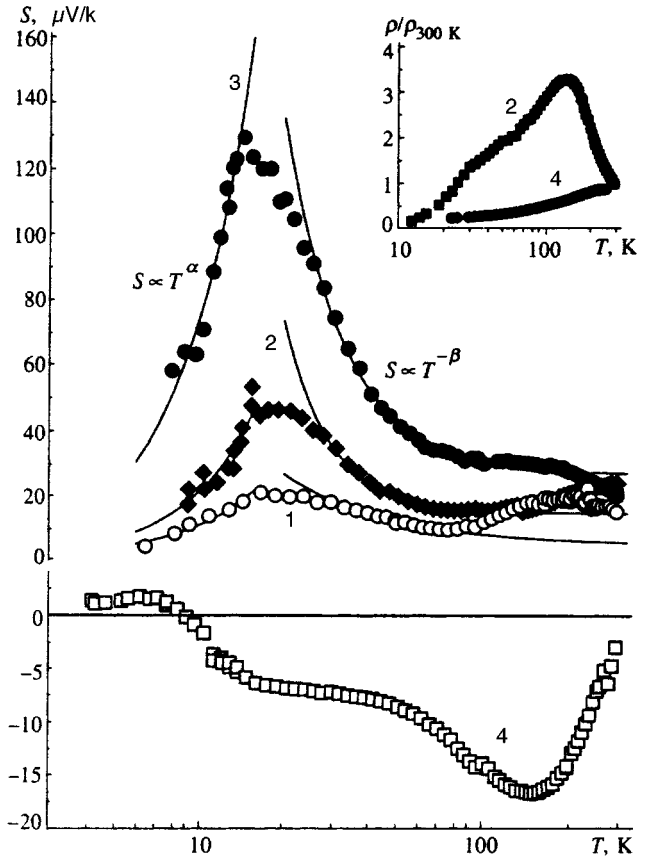


FIG. 2. Thermopower of various quasi-two-dimensional organic metals vs. temperature: 1) $(BEDT\text{-}TTF)_2NH_4Hg(SCN)_4$; 2) $(BEDT\text{-}TTF)_2RbHg(SCN)_4$; 3) $(BEDT\text{-}TTF)_2TlHg(SCN)_4$; 4) $(BEDT\text{-}TTF)_2Cu(NCS)_2$. The insert shows typical curves of resistivity vs. temperature, labeled with the same numbers as in the main graph.

3. EXPERIMENTAL RESULTS

3.1. Organic conductors with metallic conductivity

Thermopower measurements for organic metal samples with metallic conductivity are shown in Fig. 2. It is noteworthy that α-phase samples demonstrate peaks of $S(T)$ in the low-temperature region about 20 K, the peak amplitude being small for $(BEDT\text{-}TTF)_2NH_4Hg(SCN)_4$ (curve *1*), and most easily seen in curves *2* and *3* for $(BEDT\text{-}TTF)_2RbHg(SCN)_4$ and $(BEDT\text{-}TTF)_2TlHg(SCN)_4$, respectively.

The sample conductivity vs. temperature was measured using the traditional four-terminal configuration along the same crystal axis as the thermopower. The curve of $\rho(T)$ for α-phases has the typical shape shown in the insert to Fig. 2 (curve *2*). In the high-temperature region, ρ increases with decreasing temperature, and only below 150 K the conductivity has clearly metallic nature since ρ drops as the temperature is lowered. Our data are in agreement with previously published results.[2,6–10]

Unlike α-phases, $(BEDT\text{-}TTF)_2Cu(NCS)_2$ exhibits metallic properties much more explicitly: the high-temperature peak of its resistivity is considerably lower, its $S(T)$ curve has no peak in the low-temperature region, and $S(T)\approx$ const around $T\sim 20$ K (see curves *4* in the main diagram of Fig. 2

and the insert). It was reported in literature[17–19,22] that a superconducting transition to the state with $\rho=0$ was detected at $T=11$ K in κ-(BEDT-TTF)$_2$Cu(NCS)$_2$. Note also that throughout the studied temperature interval $S(T)>0$ for α-phase samples, whereas $S(T)<0$ in κ-(BEDT-TTF)$_2$Cu(NCS)$_2$ in the range $T>11$ K.

Near the superconducting transition at $T\sim 10$ K, the thermopower of (BEDT-TTF)$_2$Cu(NCS)$_2$ abruptly decreases, passes through zero at $T=10$ K, and saturates at a level of $+1.5$ μV/K (Fig. 2, curve *4*). It is generally recognized that in the superconducting state $S\equiv 0$,[15] but this assertion applies only to an isotropic superconductor. In the anisotropic case, which occurs in organic superconductors, theory predicts[22] that values $S\neq 0$ are possible, which was observed in experiments (Fig. 2). Note that the absolute value of the copper thermopower for $T<10$ K is at most of order 0.5 μV/K (in accordance with published data for pure copper[23]), so the positive $S(T)$ cannot be attributed to superconducting short-circuiting via the organic conductor. Thus, the behavior of thermopower in κ-(BEDT-TTF)$_2$Cu(NCS)$_2$ is natural for an anisotropic metal with a superconductor transition around $T\sim 10$ K.

It is of interest to compare our thermopower measurements in κ-(BEDT-TTF)$_2$Cu(NCS)$_2$ with measurements[14] of crystals of the same composition fabricated at the Institute of Chemical Physics in Chernogolovka, Russian Academy of Sciences. The overall shape of the $S(T)$ curve in the region above the superconducting transition (Fig. 2) is similar to that given in Ref. 14, but the minimum value of S in Ref. 14 is -27 μV/K, in contrast to our measurement of -17 μV/K. This discrepancy can be attributed either to specific features of organic conductors fabricated in different laboratories or to possible methodological errors: the standard thermopower measurement technique employed in Ref. 14 can lead to significant measurement errors in the temperature gradient across a small sample, hence to systematic errors in the Seebeck coefficient. According to our data, the thermopower changes sign below the superconducting transition, $T<11$ K, and settles at $S=+1.5\rangle\mu$V/K, whereas in Ref. 14 the observed value $S=0$ is reported. The reason for this discrepancy remains unclear and requires further investigation.

It is clear that unlike κ-(BEDT-TTF)$_2$Cu(NCS)$_2$, α-phases show an anomalous peak in $S(T)$ (Fig. 2). In order to analyze this effect in quantitative terms, let us consider the asymptotic behavior

$$S(T)\propto T^{\alpha}, \tag{1}$$

which is valid at temperatures below the peak, and

$$S(T)\propto T^{-\beta}, \tag{2}$$

which describes the curve in the high-temperature region ($T>30$ K). Solid lines in Fig. 2 show fits of (1) and (2) to experimental data, and values of the indices α and β are listed in Table 1. The uncertainty in parameters obtained by the nonlinear analysis is 6–10%.

TABLE I. Parameters of power functions in Eqs. (1) and (2) approximating thermopower maximum in organic conductors.

Material	α	β
(BEDT-TTF)$_2$NH$_4$Hg(SCN)$_4$	1.55	1.2
(BEDT-TTF)$_2$RbHg(SCN)$_4$	1.75	2.7
(BEDT-TTF)$_2$TlHg(SCN)$_4$	1.76	2.1
(BEDT-TTF)$_3$Cl$_2\cdot$2H$_2$O	3.9	2.7

It is clear that the values of α for all tested materials are close and range between approximately 1.5 and 1.8, whereas β is spread over a wider interval between 1.2 for M=NH$_4$ and 2.7 for M=Rb.

3.2. Organic conductor undergoing a metal–insulator transition

Measurements of the thermopower and resistivity of a (BEDT-TTF)$_3$Cl$_2\cdot$2H$_2$O sample versus temperature are plotted in Fig. 3. In the range $T>150$ K the temperature dependence of the conductivity has a metallic nature, and $S(T)>0$.

At about $T\sim 150$ K a metal–insulator transition occurs in (BEDT-TTF)$_3$Cl$_2\cdot$2H$_2$O, accompanied by reversal of the sign of the thermopower. At lower temperatures the absolute value of $S(T)$ grows, and this trend persists down to 40 K (Fig. 3). As the temperature is lowered to 20–30 K, the thermopower drops sharply in absolute value, and $S(T)\approx 0$ for $T<20$ K to within the experimental accuracy (Fig. 3). Note that in this temperature range (BEDT-TTF)$_3$Cl$_2\cdot$2H$_2$O samples have extremely high resistivity, about four orders of magnitude higher than $\rho(T\sim 150$ K), and the uncertainty in $S(T)$ increases to 3–5 μV/K.

It is notable that there is a correlation between features of the curves of $\rho(T)$ and $S(T)$: the bend $\rho(T)$ at $T\sim 60$ K

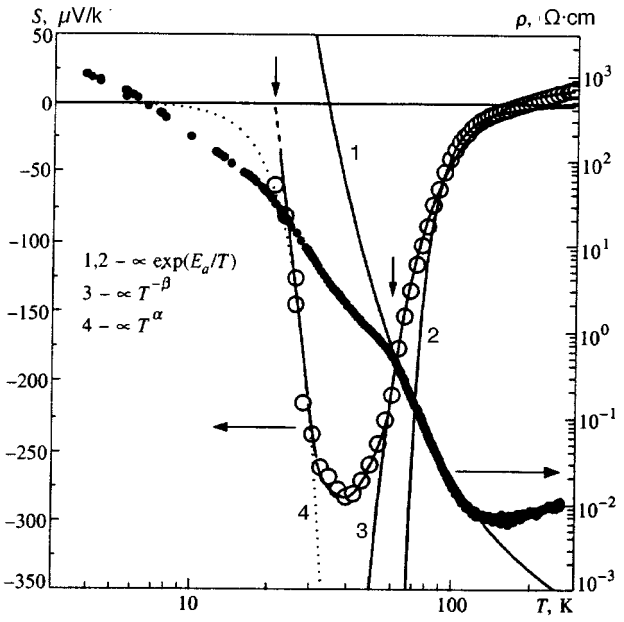


FIG. 3. Resistivity and thermopower of (BEDT-TTF)$_3$Cl$_2\cdot$2H$_2$O versus temperature. Curves 1–4 are plots of different theoretical functions.

corresponds to the point of inflection of $S(T)$, and the low-temperature bend in $\rho(T)$ at $T\sim 20$ K to $S(T)\approx 0$ (Fig. 3, cited temperatures marked by arrows).

For comparison with measurements of samples with metallic conductivity, we processed measurements of $S(T)$ in the regions both below and above the peak using Eqs. (1) and (2) (curves *3* and *4* in Fig. 3). The resulting indices α and β are listed in Table 1. Whereas index β is about 2.7 and close to the corresponding values for $(BEDT\text{-}TTF)_2RbHg(SCN)_4$ and $(BEDT\text{-}TTF)_2TlHg(SCN)_4$, the value of α is larger, $\alpha\sim 3.9$, i.e., the drop in $S(T)$ for $(BEDT\text{-}TTF)_3Cl_2\cdot 2H_2O$ is significantly steeper (Figs. 2 and 3). The temperature dependence of $S(T)$ will be discussed in detail in the next section.

4. DISCUSSION OF RESULTS

4.1. Sign of thermopower in quasi-two-dimensional organic conductors

For $(BEDT\text{-}TTF)_2MHg(SCN)_4$ α-phases and $(BEDT\text{-}TTF)_2Cu(NCS)_2$, the thermopower signs are opposite (Fig. 2), although, according to published data,[2,6–10,19,24] their transport properties (relating, for example, to quantum oscillations) are controlled in either case by hole sections of their Fermi surfaces. Recall that the thermopower of copper is positive and for $T<300$ K its magnitude is less than 1.5 μV/K,[23] so the thermopower features illustrated by Figs. 2 and 3 must be associated specifically with organic conductors. The data in Fig. 2 indicate that in $(BEDT\text{-}TTF)_2Cu(NCS)_2$ the contribution of electron sections of the Fermi surface is larger than generally accepted, however, further research and more accurate measurements of the band structure parameters are needed in order to have an ultimate judgement about the thermopower sign in tested organic conductors.

4.2. Features of thermopower in the region of the metal–insulator transition in $(BEDT\text{-}TTF)_3Cl_2\cdot 2H_2O$

Data on $(BEDT\text{-}TTF)_3Cl_2\cdot 2H_2O$ thermopower vs. temperature yield additional information about the nature of the metal–insulator transition in this material. It is known (Sec. 2.1) that there are two alternative interpretations of this phenomenon. The first is based on a transformation of the original semimetallic spectrum to a dielectric spectrum. The other model relates the metal–insulator transition to formation of a charge-density wave. According to Gruner,[25] the behavior of the resistivity as a function of temperature (Fig. 3) derives from the fact that in the highly anisotropic two-dimensional case, partial ordering first takes place along an easy axis at $T\sim 150$ K, while a charge-density wave for the entire conducting plane is finally generated at $T<20$ K.[3–5]

One can see (Fig. 3) that in fact $S(T)\approx 0$ at $T<20$ K, as it should be for a collective mode like a charge-density or spin-density wave, or a superconductor.[15] Reversal of the thermopower sign near the metal–insulator transition is more easily ascribed to opening of a gap in the electron spectrum. For example, in the simplest model with isotropic bands and equal number densities of electrons and holes, $n=p$, one can easily derive from the standard expression for the thermopower due to two carrier groups[15] that $S>0$ in a degenerate semimetal corresponds to the condition $\tau_h>\tau_e$, where τ_e and τ_h are relaxation times for electrons and holes, respectively. In the dielectric case, for which transport is due to activation processes, $S<0$ if the effective masses and relaxation times satisfy the condition $\tau_h<\tau_e m_h/m_e$. Combining these two conditions, we obtain the condition for sign reversal in the form $\tau_e<\tau_h<\tau_e m_h/m_e$, which can be easily satisfied if $m_e\ll m_h$.

It follows from the discussion above that the model of the metal–insulator transition in $(BEDT\text{-}TTF)_3Cl_2\cdot 2H_2O$ can be more complicated than that suggested in Refs. 3 and 5, which is based solely on the charge-density wave concept. At $T\sim 150$ K, the band structure changes, and a gap is probably opened in the $(BEDT\text{-}TTF)_3Cl_2\cdot 2H_2O$ spectrum, which leads to higher resistivity and a larger absolute value of thermopower; at $T\sim 20$ K, the system then forms a charge-density wave, and as a result the Seebeck coefficient vanishes and $\rho(T)$ acquires a break in the low-temperature region (Fig. 3). Note that the suggested model is in full agreement with available structural data.[4]

At the same time, the observed temperature dependence of the thermopower in the temperature range below the transition at $T\sim 150$ K is stronger (with asymptotic behavior $S\propto 1/T^{2.7}$) than $S\propto 1/T$ which corresponds to thermal activation across a band gap.[14] For this reason, we have to assume that some features of the $(BEDT\text{-}TTF)_3Cl_2\cdot 2H_2O$ thermopower are due to phonon drag.[15,16]

An accurate calculation of phonon-drag effects is a rather complex problem and requires detailed information about the variance of relaxation times.[16] Usually theoretical calculations for the degenerate case predict dependences like $S\propto 1/T^4$ or $S\propto 1/T^3$. The latter corresponds to the Herring relaxation mechanism[16] and is close to the experimental data for some organic conductors (Table 1). In the case of carrier activation, a simple estimate of the phonon drag effect as a function of temperature can be obtained using the expression[15]

$$S(T)=\frac{C(T)}{n(T)e}\left(\frac{1}{1+\tau_{e-ph}(T)/\tau(T)}\right), \tag{3}$$

where $C(T)$ is the phonon heat capacity, $n(T)$ and e are the concentration and charge of carriers, $\tau_{e-ph}(T)$ is the electron–phonon relaxation time, and $\tau(T)$ is the relaxation time of the phonon gas.

It follows from Eq. (3) that in the case of thermal activation of carriers, the temperature dependence of the thermopower is determined, to a first approximation, by the exponential temperature dependence of the concentration, whereas all other parameters in Eq. (3) are described by power laws, i.e., one should expect that at $T<150$ K the $(BEDT\text{-}TTF)_3Cl_2\cdot 2H_2O$ resistivity should be described by the formula $\rho\propto S\propto\exp(E_a/T)$. We have derived from the resistivity measurements the activation energy $E_a\approx 500$ K (Fig. 3, curve *1*), and from the approximation of the initial section of the $S(T)$ curve (Fig. 3, curve *2*) $E_a\approx 490$ K, which is the same to within the experimental accuracy of ~ 15 K. One can see in Fig. 3, however, that the exponential asymptotic behavior describes thermopower data only up to

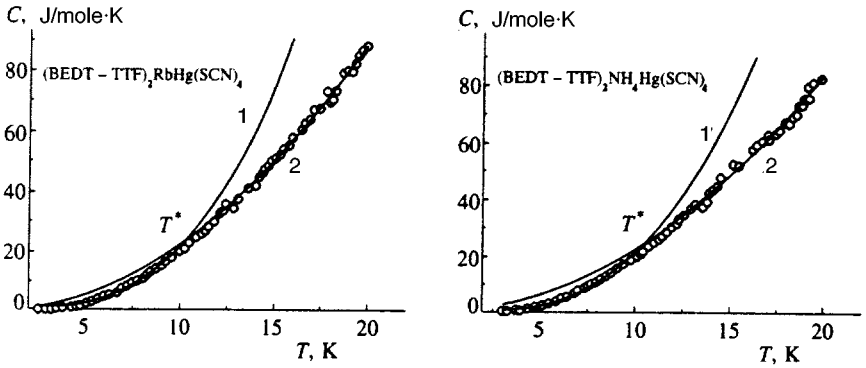


FIG. 4. Switching between two different theoretical temperature dependences of heat capacity in quasi-two-dimensional organic metals: (1) Debye formula $C \propto T^3$; (2) $C \propto T^{\delta}$, where $\delta = 2$. Experimental data are taken from Ref. 27.

$T \sim 80$ K, i.e., on the upper half of the $\rho(T)$ activation section, and at lower temperatures the thermopower magnitude is lower than expected. This discrepancy can be ascribed, for example, to the temperature dependence of the phonon heat capacity, since $C(T)$ drops with decreasing temperature.

Note that the fit of the high-temperature section of the $S(T)$ curve ($T > 80$ K) by Eq. (2) yields $S(T) \propto 1/T^{4.1}$, which corroborates the stronger temperature dependence in this region. Moreover, this result, as well as the approximation $S(T) \propto 1/T^{2.7}$, which is the best over a wider temperature interval (Fig. 3, curve *3*), is in disagreement with the theoretical dependence $S \propto 1/T$, which describes activation transport in the absence of phonon drag. Thus, the data in Fig. 3 indicate that the increase in the (BEDT-TTF)$_3$Cl$_2\cdot$2H$_2$O thermopower magnitude at low temperatures is due to phonon drag. Further decrease in the temperature leads to formation of a charge-density wave and, as a result, $S \approx 0$. The peak in the $S(T)$ curve in Fig. 3 is due to this change of the transport regime.

4.3. Low-temperature peak in thermopower of quasi-two-dimensional organic metals

We now consider possible mechanisms leading to formation of $S(T)$ peaks in (BEDT-TTF)$_2$MHg(SCN)$_4$ samples. Since these samples are in a metallic state at least at $T < 100$ K,[2,6–12,24] the most probable cause of the thermopower growth with decreasing temperature, as in the case of (BEDT-TTF)$_3$Cl$_2\cdot$2H$_2$O, is the phonon-drag effect, although the measured index β (Table 1) is less than the theoretical values $\beta = 3-4$ for degenerate systems.[15] It is noteworthy that the values of α for all tested samples are close, $\alpha \sim 1.5-1.75$, and considerably less than in the case of (BEDT-TTF)$_3$Cl$_2\cdot$2H$_2$O, where $\alpha \sim 3.9$ (Table 1). Thus, the drop in $S(T)$ at $T < 15$ K (Fig. 2) can hardly be ascribed to a transition to a state with a charge-density wave since as was shown for (BEDT-TTF)$_3$Cl$_2\cdot$2H$_2$O, the thermopower should drop more rapidly in this case (Fig. 3). We suppose that this drop can also be described in terms of the phonon-drag model.[15]

In the degenerate case, $n = \text{const}$ and the thermopower as a function of temperature is determined by temperature dependences of the heat capacity and relaxation times. In the three-dimensional case $\tau \propto T^{-5}$, irrespective of the features of the dispersion curve,[26] and the electron–phonon scattering time is a slower function of temperature, for example, $\tau_{e-ph} \propto T^{-3/2}$.[16] Therefore the ratio τ_{e-ph}/τ drops with decreasing temperature, and ''high-temperature'' asymptotes, when $\tau_{e-ph}/\tau \gg 1$, can be derived from Eq. (3),

$$S(T) \propto C(T)\frac{\tau(T)}{\tau_{e\text{-}ph}(T)} \tag{4}$$

as well as ''low-temperature'' asymptotes, $\tau_{e-ph}/\tau \ll 1$:

$$S(T) \propto C(T). \tag{5}$$

The traditional interpretation[15] of the $S(T)$ peak is that $C(T) = \text{const}$ in the high-temperature region and, given the relaxation times selected above, $S(T) \propto 1/T^{7/2}$, whereas in the low-temperature range $S(T) \propto C(T) \propto T^3$. Thus the interpretation of thermopower as a function of temperature in terms of the phonon-drag model requires information about the temperature dependence of specific heat in real samples.

Published data on $C(T)$ of quasi-two-dimensional organic conductors[27–30] suggest that for all organic conductors studied, the heat capacity curve plotted in the standard manner, $C/T = f(T^2)$, is highly nonlinear at $T > 10$ K.[27–30] This means that temperature dependence of the phonon component of heat capacity is significantly different from that prescribed by the Debye law, $C \propto T^3$, and the change in this temperature dependence should occur at about $T^* \sim 10$ K.

We have calculated the exponent δ in the heat capacity versus temperature,

$$C(T) \propto T^{\delta}, \tag{6}$$

separately in each temperature range, $T < T^*$ and $T > T^*$, using the data from Refs. 27–29. Typical curves are given in Fig. 4, and they clearly indicate that at $T^* \sim 10$ K the asymptotic $C(T)$ power index actually does change for organic conductors. The resulting exponents δ are listed in Table 2, and their uncertainty is about 0.05.

It follows from Table 2 that the temperature dependence of the heat capacity is universal for quasi-two-dimensional organic conductors. At $T < T^*$ the heat capacity is described by the Debye formula to within the experimental errors, and $\delta \approx 3$, while at $T > T^*$ δ drops to $\delta \approx 1.7-2.0$. Thus, the lattice specific heat dominates the electronic contribution in the temperature range $T > 2$ K, which was studied in Refs. 27–29, and measurements at lower temperatures are needed to properly isolate the electronic heat capacity. As a result, measurements[27–29] of $C(T)$ can be used in analysis of the thermopower due to phonon drag using Eqs. (3)–(5).

TABLE II. Parameters derived from measurements of heat capacity as a function of temperature in quasi-two-dimensional organic conductors given in Refs. 27–29 and calculations by the phonon-drag model.

Material	δ		γ	τ_{e-ph}/τ
	$T<T^*$	$T>T^*$		($T<20$ K)
$(BEDT\text{-}TTF)_2NH_4Hg(SCN)_4$	3.00	1.91	3.11	0.13
$(BEDT\text{-}TTF)_2RbHg(SCN)_4$	3.05	2.03	4.73	0.06
$(BEDT\text{-}TTF)_2KHg(SCN)_4$	3.09	1.84	-	-
$(BEDT\text{-}TTF)_2TlHg(SCN)_4$	-	-	3.94	0.02
$(TMTSF)_2PF_6$	2.94	1.70	-	-

Unfortunately, the relatively high superconducting transition temperature of κ-$(BEDT\text{-}TTF)_2Cu(NCS)_2$ and the ensuing additional contribution to $C(T)$ make an accurate determination of δ difficult in this case. Our estimate of δ based on the curve[30] of $C(T)$ in a magnetic field of 5 T yields 2.45 ± 0.05. Since the curve of $C(T)$ in Ref. 30 is plotted over the temperature range 6–12 K, this value of δ most likely corresponds to the intermediate temperature range ($T\sim T^*$), and additional information is needed for accurate determination of high- and low-temperature asymptotic behavior.

It is interesting that a similar change of the regime was theoretically analyzed by Kosevich[26] for the heat capacity of layered crystals with tight bonds within one layer and loose bonds between layers. According to Ref. 26, the anisotropy of elastic constants is negligible at low temperatures, and the traditional formula $C\propto T^3$ applies. As the temperature increases, the contribution of high-frequency bending modes to $C(T)$ becomes important, and their dispersion curve, although three-dimensional, contains a dominant term quadratic in the quasimomentum, and as a result the Debye formula should be replaced with $C\propto T^2$.[26] According to the standard model of quasi-two-dimensional organic conductors, they are composed of sets of organic molecule layers (BEDT-TTF in our specific case) separated by anion polymer layers.[1,2] It is obvious that the anisotropy of elastic moduli should be strong and the heat capacity vs. temperature should be an anomalous function. To the best of our knowledge, however, the applicability of the theoretical results[26] to the specific heat of organic conductors as a function of temperature has not been discussed as yet.[27–30]

Comparison between Figs. 2 and 4 demonstrates that the thermopower of α-phases peaks in the temperature range where $\delta\approx2$. Given Eqs. (3) and (6), one can easily show that in the ''low-temperature'' limit ($\tau_{e-ph}/\tau\ll1$)

$$\alpha\approx\delta-\gamma\frac{\tau_{e\text{-}ph}}{\tau}, \tag{7}$$

and in the ''high-temperature'' range, where $\tau_{e-ph}/\tau\gg1$,

$$\beta\approx\gamma\left(1-\frac{\tau}{\tau_{e-ph}}\right)-\delta, \tag{8}$$

where the index γ determines the power-law dependence:

$$\frac{\tau_{e-ph}}{\tau}\propto T^{\gamma}. \tag{9}$$

In deriving Eqs. (7) and (8) we took into account corrections of the first order in τ_{e-ph}/τ and τ/τ_{e-ph} in the respective temperature ranges. Comparison of the data in Tables 1 and 2 shows that condition $\alpha\approx\delta$, which is usually included in the standard approach (Eq. (5)), is satisfied only to low accuracy, so the correction to α is substantial.

We now estimate parameters of the phonon-drag model in Eqs. (7) and (8) for samples of $(BEDT\text{-}TTF)_2MHg(SCN)_4$. Since no data on the specific heat for M=Tl in the temperature range of interest are available, we have used results for its closest isostructural analogue M=K, whose electronic properties and magnetic order pattern are essentially identical.[6,7,12] In addition, we assume that the phonon-drag correction in Eq. (8) is negligible; then $\gamma\approx\delta+\beta$, and in the temperature range below the thermopower peak ($T<20$ K in Fig. 2) $\tau_{e-ph}/\tau\approx(\delta-\alpha)/(\alpha+\beta)$. The parameters γ and τ_{e-ph}/τ calculated in this way are listed in Table 2. Using these data to extrapolate τ_{e-ph}/τ to the temperature range above the peak, we can show that the correction to γ is at most 1–8%, which justifies our estimation procedure.

Thus, it follows from our analysis that the electron–phonon interaction time divided by the phonon relaxation time in superconducting $(BEDT\text{-}TTF)_2NH_4Hg(SCN)_4$ is greater than in magnetically ordered $(BEDT\text{-}TTF)_2MHg(SCN)_4$ (M=Rb, Tl) samples, and the parameter γ, on the contrary, is smaller. It is known that the maximum possible value of γ is 5, which corresponds to $\tau_{e-ph}=$const.[24] Standard scattering theory yields $\tau_{e-ph}\propto T^{-3/2}$ or $\tau_{e-ph}\propto T^{-1/2}$,[31] and in the presence of a strong random potential the phonon scattering time it should be $\tau_{e-ph}\propto T^{-1}$, so it is feasible that real systems are characterized by γ ranging between 3.5 and 5, which is in satisfactory agreement with Table 2. Moreover, the stronger the temperature dependence of τ_{e-ph} at low temperatures, the higher the index γ and, as follows from experimental data reported in this paper, the feature distinguishing superconducting organic metals from magnetically ordered ones is probably the steeper function $\tau_{e-ph}(T)$.

If we ''extrapolate'' this trend and assume that the higher the superconducting transition temperature, the stronger the temperature dependence of τ_{e-ph}, then we can give a qualitative interpretation of $S(T)$ for κ-$(BEDT\text{-}TTF)_2Cu(NCS)_2$ (Fig. 2). If the temperature dependence of τ_{e-ph} is described by a power law with an exponent of up to 3, we have $\tau_{e-ph}/\tau\propto T^{-2}$ and, as follows from Eq. (8), the phonon contribution is, in a way, compensated for, so that $\beta\approx0$ and the thermopower has a plateau as a function of temperature. Precisely this behavior is observed in κ-$(BEDT\text{-}TTF)_2Cu(NCS)_2$ in the range $15\text{ K}<T<50$ K, i.e., near the peak position in other materials (Fig. 2, curve *4*).

5. CONCLUSIONS

In this paper, we have shown that some features of thermopower in the studied samples of quasi-two-dimensional organic conductors, namely the peaks for α-$(BEDT\text{-}TTF)_2MHg(SCN)_4$ and the plateau for κ-$(BEDT\text{-}TTF)_2Cu(NCS)_2$ in the temperature range 10–50

K are probably due to phonon-drag effects. Measurements of the Seebeck coefficient versus temperature can be satisfactorily interpreted in terms of a simple model that takes the actual temperature dependence of the phonon heat capacity into account, which is not described by the Debye formula.

The phonon-drag effects probably also determine the behavior of the thermopower in the $(BEDT\text{-}TTF)_3Cl_2\cdot 2H_2O$ organic conductor, characterized by the metal–insulator transition at a temperature around $T\sim 150$ K. The combined analysis of conductivity and thermopower measurements as functions of temperature has allowed us to improve upon the results of previous research[3–5] and establish that the transition in this material is of a complex nature: first the electron spectrum turns ''dielectric,'' with a gap emerging at $T\sim 150$ K, then a transition to the state with a charge-density wave occurs at $T\sim 20$ K.

We are indebted to J. Singleton, S. J. Blundell, and M. Kurmoo, who supplied samples of organic conductors. This work was supported by INTAS grants 93-2400 and 94-4435, and by the Russian Fund for Fundamental Research (Project 96-16176).

[1] T. Ishiguro and K. Yamaji, *Organic Superconductors*, Springer-Verlag, Berlin-Heidelberg-New York (1990).
[2] J. Wosnitza, in *Springer Tracts in Modern Physics*, Springer-Verlag, New York (1996), Vol. 134.
[3] S. V. Demishev, N. E. Sluchanko, N. A. Samarin, M. V. Kondrin, J. Singleton, M. S. L. Du Croo de Jongh, W. Hayes, M. Kurmoo, and P. Day, Ferroelectrics **176**, 329 (1996).
[4] C. J. Kepert, *PhD Thesis*, University of London (1996).
[5] W. Lubczynski, S. V. Demishev, J. Singleton, J. M. Caulfield, L. Du Croo de Jongh, C. J. Kepert, S. J. Blundell, W. Hayes, M. Kurmoo, and P. Day, J. Phys.: Condens. Matter **8**, 6005 (1996).
[6] T. Sasaki and N. Toyota, Phys. Rev. B **49**, 10120 (1994).
[7] M. V. Kartsovnik, A. E. Kovalev, and N. D. Kushch, J. Phys. I **3**, 1187 (1993).
[8] F. L. Pratt, J. Singleton, M. Doporto, A. J. Fisher, T. J. B. M. Janssen, J. A. A. J. Perenboom, M. Kurmoo, W. Hayes, and P. Day, Phys. Rev. B **45**, 13904 (1992).
[9] N. Harrison, A. House, I. Deckers, J. Caulfield, J. Singleton, F. Herlach, W. Hayes, M. Kurmoo, and P. Day, Phys. Rev. B **52**, 5584 (1995).
[10] J. Caulfield, S. J. Blundell, M. S. L. du Croo de Jongh, P. T. J. Hendriks, J. Singleton, M. Doporto, F. L. Pratt, A. House, J. A. A. J. Perenboom, W. Hayes, M. Kurmoo, and P. Day, Phys. Rev. B **52**, 8325 (1995).
[11] S. V. Demishev, N. E. Sluchanko, A. V. Semeno, and N. A. Samarin, JETP Lett. **61**, 313 (1995).
[12] S. V. Demishev, A. V. Semeno, N. E. Sluchanko, N. A. Samarin, I. B. Voskoboinikov, V. V. Glushkov, J. Singleton, S. J. Blundell, S. O. Hill, W. Hayes, M. V. Kartsovnik, A. E. Kovalev, M. Kurmoo, P. Day, and N. D. Kushch, Phys. Rev. B **53**, 12794 (1996).
[13] T. Komatsu, H. Sato, N. Matsukawa, T. Nakamura, H. Yamochi, G. Saito, M. Kusunoki, K. Sakasguchi, and S. Kagoshima, Synth. Met. **70**, 779 (1995).
[14] L. I. Buravov, A. V. Zavarykina, N. D. Kushch, V. N. Laukhin, V. A. Merzhanov, A. G. Khomenko, and É. B. Yagubskii, Zh. Éksp. Teor. Fiz. **95**, 322 (1989) [Sov. Phys. JETP **68**, 182 (1989)].
[15] P. M. Chaikin, in *Organic Superconductivity*, V. Z. Kresin and W. A. Little (eds.), Plenum Press, New York (1990), p. 101.
[16] B. M. Askerov, *Electron Transport in Semiconductors* [in Russian], Nauka, Moscow (1985), p. 197.
[17] N. F. Mott and E. A. Davis, *Electron Processes in Non-Crystalline Materials*, Clarendon Press, Oxford (1979).
[18] H. Mori, I. Hirabayashi, S. Tanaka, T. Mori, H. Inokuchi, K. Oshima, and G. Saito, Synth. Met. **55–57**, 2443 (1993).
[19] N. Toyota and T. Sasaki, Solid State Commun. **74**, 361 (1990).
[20] M.-H. Whangbo, J. Ren, D. B. Kang, and J. M. Williams, Mol. Cryst. Liq. Cryst. **181**, 17 (1990).
[21] Ch. Laurent, S. K. Patapis, M. Laguesse, H. W. Vanderschueren, A. Rulmont, P. Tarte, and M. Ausloos, Solid State Commun. **66**, 445 (1988).
[22] B. T. Geilikman, *Research in Low-Temperature Physics* [in Russian], Atomizdat, Moscow (1979), p. 36.
[23] *Physical Constants* [in Russian], I. S. Grigor'ev and E. Z. Meilikhov (eds.), Énergoatomizdat, Moscow (1991), p. 560.
[24] J. Singleton, J. Caulfield, S. Hill, S. Blundell, W. Lubczynski, A. House, W. Hayes, J. Perenboom, M. Kurmoo, and P. Day, Physica B **211**, 275 (1995).
[25] G. Gruner, Rev. Mod. Phys. **60**, 1129 (1988).
[26] A. M. Kosevich, *Principle of Crystal Lattice Mechanics* [in Russian], Nauka, Moscow (1972), p. 122.
[27] Y. Nakazawa, H. Sato, A. Kawamoto, and K. Kanoda, Synth. Met. **70**, 943 (1995).
[28] J. Odin, J. C. Lasjaunias, K. Biljakovic, P. Monceau, and K. Bechgaard, Solid State Commun. **91**, 523 (1994).
[29] R. G. Goodrich and J. Xu, Physica B **165&166**, 889 (1990).
[30] J. E. Graebner, R. C. Haddon, S. V. Chichester, and S. H. Glarum, Phys. Rev. B **41**, 4808 (1990).
[31] V. F. Gantmakher and I. B. Levinson, *Scattering of Current Carriers in Metals and Semiconductors* [in Russian], Nauka, Moscow (1984), p. 91.

Translation provided by the Russian Editorial office

Thermopower of $Al_{1-x}Si_x$ solid solutions in vicinity of lattice instability

N. E. Sluchanko, V. V. Glushkov, S. V. Demishev, M. V. Kondrin, N. A. Samarin, and V. V. Moshchalkov*)

Institute of General Physics, Russian Academy of Sciences, 117942 Moscow, Russia

V. V. Brazhkin

Institute of High-Pressure Physics, Russian Academy of Sciences, 142092 Troitsk, Moscow Region, Russia
(Submitted 18 July 1997)
Zh. Éksp. Teor. Fiz. **113**, 339–351 (January 1998)

The thermopower coefficient as a function of temperature, $S(T)$, has been measured in nonequilibrium superconductors, such as $Al_{1-x}Si_x$ substitutional solid solutions and Al–Si alloys on various decay stages. When aluminum is substituted with silicon, the contribution to the thermopower due to phonon-drag effects, which are dominant in pure aluminum at low temperatures, is suppressed, and low-temperature anomalies in $S(T)$ detected in compositions near lattice instability limit are determined by the diffusion component of the thermopower. The low-temperature anomalies in the thermopower and the notable increase in the coefficient in front of the linear term in $S(T)$ are attributed to effects of thermopower renormalization due to the electron–phonon interaction enhancement with ''soft modes'' in the face-centered cubic (FCC) lattice of $Al_{1-x}Si_x$ solid solutions. The nature of these anomalies in $S(T)$ is analyzed in terms of the Kaiser and Reizer–Sergeev models. [S1063-7761(98)02301-4]

1. INTRODUCTION

In recent years, a lot of attention has been attracted by theoretical and experimental studies of the thermopower coefficient as a function of temperature, $S(T)$, in various metal-like systems, such as glasses,[1] high-temperature superconductors,[2] fullerenes,[3] etc. This interest is caused by high ''susceptibility'' of this parameter to the electron spectrum of a material and the character of quasiparticle scattering, in particular, to features of electron–phonon interaction in the normal state. According to some researchers,[2–4] features of the electron–phonon interaction and their effect on the thermopower in the temperature range above the superconducting transition may supply a clue to which of the superconductivity mechanisms is realized in a specific material and also yield information about the Eliashberg function of studied materials. Moreover, the enhancement factor of electron–phonon interaction can be derived directly from thermopower measurements in metals with substitutional impurities.[4,5]

At the same time, important properties of most oxide high-temperature superconductors, fullerenes, and organic superconductors are determined by proximity to the structural lattice instability. This is, on one hand, one of the factors of superconductivity enhancement,[6] on the other hand, it makes difficult an investigation of these multicomponent materials.

$Al_{1-x}Si_x$ substitutional solid solutions synthesized under high pressure[7] are among the most simple metal-like systems from the viewpoint of both lattice structure, electron spectrum, and chemical composition. Substitution of Al with Si without breaking the FCC lattice configuration in the solid solution leads to an increase in the superconducting transition temperature T_c up to ~ 11 K at $x \approx 0.2$. Note that $Al_{1-x}Si_x$ and $Al_{1-x}Ge_x$ compounds are metastable,[8] which, together with the enhancement of the superconductivity by one order of magnitude (since $T_c(\mathrm{Al}) \approx 1.18$ K) and persistence of the FCC lattice structure, makes these materials convenient models for studying superconductivity.

Attention was focused on the nature of the superconductivity enhancement in these nonequilibrium materials in our previous research.[9,10] The stability region and decay kinetics of supersaturated $Al_{1-x}Si_x$ solid solutions on early and late stages of their phase transformations have also been studied.[11,12] Our previous research[9–12] indicated that the origin of the increase in T_c of $Al_{1-x}Si_x$ compounds is the enhancement of the electron–phonon interaction due to cluster modes generated in the Al-based FCC lattice in approaching the lattice instability limit or spinodal. To the best of our knowledge, no measurements of the thermopower coefficient of $Al_{1-x}Si_x$ metastable solid solutions with $x \geqslant 0.01$ have been performed as yet. Since a considerable enhancement of the electron–phonon interaction in $Al_{1-x}Si_x$ has been detected in this range of Si concentrations (the electron–phonon interaction constant $\lambda(Al_{0.92}Si_{0.08}) = 0.9$,[10] whereas $\lambda(\mathrm{Al}) = 0.38$), it is of interest to study $S(T)$ in compositions approaching the lattice instability.

Thus, the aim of the reported research was to investigate in detail the thermopower coefficient versus temperature and establish its relation to the enhancement of electron–phonon interaction and charge carrier scattering in nonequilibrium $Al_{1-x}Si_x$ model solid solutions over wide ranges of the silicon content x and temperature.

2. EXPERIMENTAL TECHNIQUE

1. From the technical viewpoint, the major problem in measurements of thermopower of small metallic samples

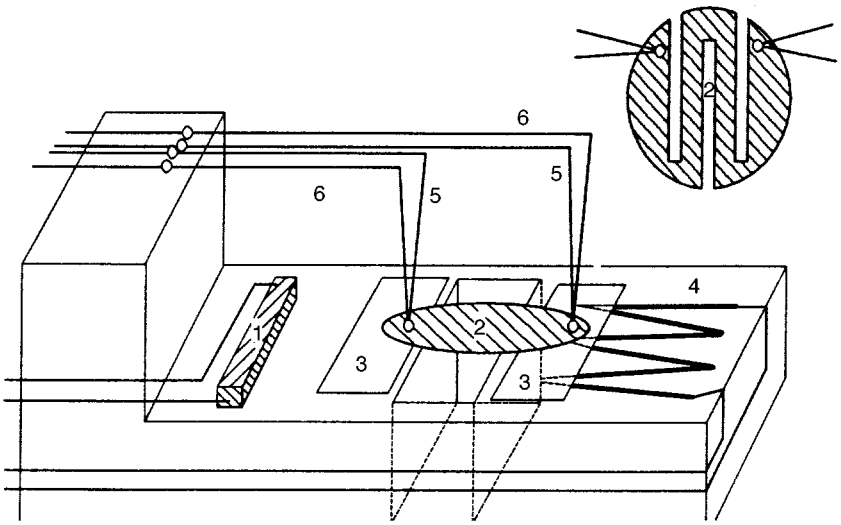


FIG. 1. Diagram of the low-temperature section of the facility for measuring thermopower (see text).

with relatively low resistivities is fabrication of reliable thermal contacts on two opposite sample faces in an evacuated volume, without which accurate measurements of the temperature gradient are impossible. In addition, owing to the high thermal conductivity of such materials, there is a problem of detecting weak signals in measuring both the temperature gradient and thermopower in the sample. In order to reduce the shunting effect of the sample on the temperature gradient, we polished disk-shaped samples with a diameter no larger than 3.5 mm to a thickness of about 100 μm, after which comb-shaped patterns of slits were made on both sides of the disk using an electric-erosion facility in order to increase the effective distance between the two potential contacts on the sample (see the insert to Fig. 1).

Measurements were performed in a low-temperature sealed-off tube with a cold plate supporting the sample in the evacuated volume, the whole assembly being inserted into an STG-25 (40) helium transport Dewar with a volume of 25 or 40 liters. By pumping liquid helium from the transport Dewar through a sintered copper throttle to the cold plate, we could control the sample holder temperature throughout an interval of 1.5 to 300 K, and an electronic circuit driving the heaters maintained the temperature at a prescribed value to within 0.05 K. A schematic drawing of the assembly designed for measuring thermopower is given in Fig. 1.

The plate temperature was measured by resistance thermometer 1. Sample 2 was glued to two sapphire substrates 3, which electrically insulated the disk sample mounted for measurements from the copper support. A temperature gradient across the sample was generated by small electric heater 4. In order to reduce the heat flow between the sample ends via the holder, an opening was milled in the copper plate (Fig. 1). The thermopower signal was picked up from two copper (5) and two constantan (6) contacts. Junctions between copper signal wires and the contacts made good thermal contact with the copper plate through thin capacitor paper and heat-conducting glue, which made it possible to equalize the copper–constantan junctions and thereby eliminate spurious thermopower measurements. In order to have equal temperature differences between the copper and constantan contacts, they were soldered to one another pairwise (copper to constantan), each pair was pressed to either the ''hot'' or ''cold'' end of the sample and cemented by an electrically conducting glue. This assembly assured good thermal and electric contact with the sample, and in quality-assurance measurements the samples withstood more than ten heating/cooling cycles without notable degradation of their properties and parameters of electric contacts.

Electric signals were picked up and the experiment was controlled by a dedicated electronic unit. At the initial moment, the holder temperature was set at a fixed value. Then a current of a fixed amplitude, which was determined by the computer code, was fed to the small heater. When the temperature difference across the sample settled, thermopower signals U_{Cu} and U_{ct} from the copper and constantan contacts, respectively, were recorded simultaneously. A new heating current amplitude was fixed thereafter, and the measurement cycle was repeated. Steady-state thermopower measurements plotted in U_{Cu} vs U_{ct} coordinates were on a straight line, whose slope was calculated using the least squares method. The number of measurement cycles and the maximum heating current amplitude were selected so that the uncertainty of the least squares method was at most 3–5%. Since the slope is

$$\tan\,\alpha=\frac{S_{\mathrm{Cu}}-S_{\mathrm{sample}}}{S_{\mathrm{ct}}-S_{\mathrm{sample}}}, \tag{1}$$

and the calibration curve of the thermopower for the contact materials is known (this is, in fact, a calibration curve of the copper/constantan thermocouple), the sample thermopower coefficient S_{sample} with respect to one contact (in this specific case, copper) can be calculated. Note that this technique has an advantage over the standard differential method,[13] because in our method the unknown temperature difference between the temperature pickup and the sample-wire contact cancels. A similar technique was described by Laurent *et al.*,[14] but they did not vary the temperature gradient at a fixed temperature, as we did in our experiments.

In some cases thermopower measurements were checked using the standard differential technique for measuring $S(T)$,

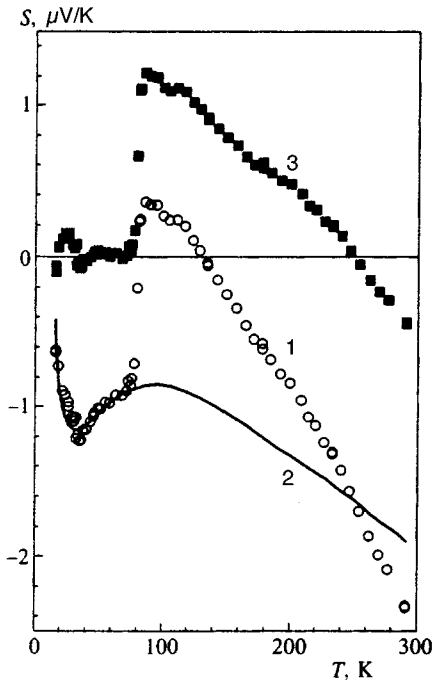


FIG. 2. Test measurements of $La_{2-x}(BaSr)_xCuO_4$ thermopower with respect to gold (see text).

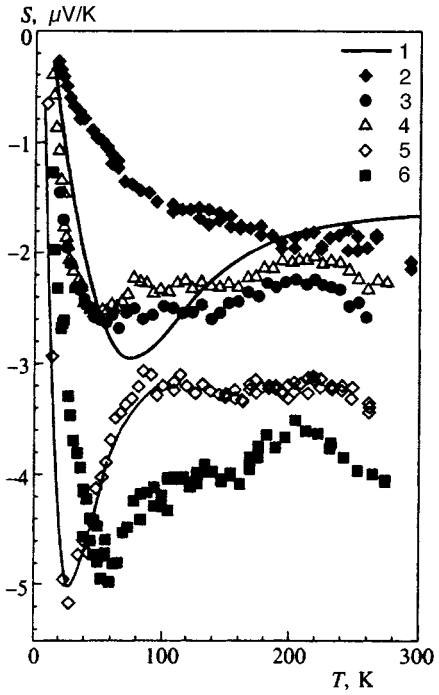


FIG. 3. Thermopower of $Al_{1-x}Si_x$ versus temperature for the following compounds: 1) pure aluminum[16]; 2) aluminum thermally processed at high pressure; 3) $x=2$ at. %; 4) $x=4$ at. %; 5) $x=6$ at. %; 6) $x=8$ at. %.

the temperature gradient across the sample being measured by a small Cu–Cu:Fe thermocouple. Measurements based on various experimental techniques performed on different samples of the same composition demonstrated that absolute value of $S(T)$ varied by 5 to 10%, which is a clear indication of good reproducibility and reliability of the experimental method.

2. Samples of $Al_{1-x}Si_x$ supersaturated substitutional solid solution with $x \leqslant 0.08$ were fabricated by quenching under a high pressure of 8 GPa in a Toroid chamber.[15] The synthesis scheme, methods of $Al_{1-x}Si_x$ sample preparation and characterization were similar to those described in Refs. 7 and 9–12.

3. The experimental facility shown in Fig. 1 was calibrated and tested using a sample of $Bi_{1.1}SrCa_{0.4}CuO_x$ oxide high-temperature superconductor, whose parameters were measured by the differential method over a temperature interval of 4.2 to 300 K. Measurements of thermopower with respect to gold are plotted in Fig. 2, which, in addition to $\Delta S = S_{BiSrCaCuO} - S_{Au}$ (curve 1), also shows the negative gold thermopower (curve 2) and $S(T)$ for the studied ceramic sample directly derived from these two sets of measurements. Note that $T_c \approx 80$ K obtained in this experiment is in good agreement with resistance and magnetic susceptibility measurements performed on the same sample. The data on $S_{BiSrCaCuO}(T)$ (curve 3) in the low-temperature range, where the ceramic is in the superconducting state, allow us to estimate the absolute error of measurements, which is at most 0.2 μV/K in this case.

3. EXPERIMENTAL RESULTS

Measurements of thermopower of $Al_{1-x}Si_x$ solid solutions using samples of four different compositions with silicon content below the point of absolute instability of the aluminum-based FCC lattice are given in Fig. 3. The graph also shows for comparison curves of $S(T)$ for pure aluminum (99.9%) thermally processed at high pressure in conditions similar to those of the high-pressure synthesis and the $S(T)$ curve for aluminum taken from Ref. 16.

The comparison between curves 1 and 2 in Fig. 3 clearly shows that thermal treatment of aluminum at high pressure leads to a notable reduction of the down-peak in the thermopower at $T_{min} \approx 80$ K (curve 1 in Fig. 3), which is ascribed to phonon-drag effects in aluminum. As the silicon content in the $Al_{1-x}Si_x$ solid solution with the FCC lattice increases in the range of up to 4 at. %, the down-peak in the $S(T)$ curve is restored (compare curves 2, 3, and 4 in Fig. 3). At the same time, this feature in the $S(T)$ curves is broadened, and the peak position for $x=0.02$ and $x=0.04$ is notably shifted to lower temperatures.

These changes in the $S(T)$ curve caused by thermal processing and substitution of aluminum with silicon in $Al_{1-x}Si_x$ are in good agreement with the results of Refs. 17 and 18, which reported on disordering of the aluminum lattice through either substitution of Al with Ag, Zn, and Mg at low impurity concentrations, $x \leqslant 0.005$,[17] or generation of various types of vacancies under a flow of neutrons with energies higher than 0.1 MeV.[18] Tendencies characteristic of changes in the thermopower of aluminum-based alloys were qualitatively analyzed in Ref. 19, and it turned out that the increase in scattering due to structural defects was accompanied not only by a suppression of the down-peak in the aluminum thermopower at $T_{min} \approx 80$ K, which is due to phonon-drag effects, but also by the emergence of an additional negative diffusion component in the thermopower.

A further increase in the silicon content in nonequilibrium $Al_{1-x}Si_x$ solid solutions near the FCC lattice instability limit leads to a notable increase in the amplitude of the low-temperature feature in the $S(T)$ curve (curves 5 and 6 in Fig. 3). Note that, apart from an increase in the down-peak amplitude in alloys with $x=0.06$ and 0.08, the absolute thermopower at room temperature also increases significantly (Fig. 3).

With a view to elucidating the nature of these anomalies in the $Al_{1-x}Si_x$ thermopower, it seems interesting to study changes in the thermopower coefficient $S(T)$ due to a step-by-step isothermal annealing, which initiates decay of supersaturated solid solutions. Given specific features of phase transitions in $Al_{1-x}Si_x$,[10–12] it is possible to stabilize successive intermediate metastable states by fast quenching from the annealing temperature and to measure thermopower as a function of temperature in Al–Si alloys with various silicon contents in the solid-solution phase.

In terms of the phase composition, Al–Si metastable alloys in intermediate stages of decay are metal matrices of $Al_{1-x}Si_x$ solid solutions with submicron silicon inclusions. Measurements of small-angle neutron scattering in $Al_{1-x}Ge_x$ demonstrated[20] that the decay of an aluminum-based solid solution leads to formation of dielectric inclusions with shapes close to spherical in the metallic matrix, and their dimension in late stages (coalescence) is about 40–60 Å. Changes in the thermopower of an aging alloy containing either noncoherent or spherical coherent inclusions are largely controlled by the depletion of the solid-solution matrix,[21,22] whereas formation and growth of inclusions have not any noticeable effect on the thermopower coefficient $S(T)$ magnitude in such a multiphase system.

Thus, the results of earlier research indicate[20–22] that one should expect that submicron semiconducting (Si, Ge) inclusions generated in the process of $Al_{1-x}Si_x$ and $Al_{1-x}Ge_x$ decay in the Al-based solid solution matrix have little influence on the changes in the alloy thermopower due to a phase transformation. As a result, the evolution of $S(T)$ should be largely controlled by the thermopower of the solid-solution matrix, i.e., the Si content in $Al_{1-x}Si_x$.

We selected for measurement with the step-by-step isothermal annealing a $Al_{0.92}Si_{0.08}$ sample, whose composition was very close to the region of the FCC lattice instability limit. The procedure of step-by-step annealing is described elsewhere.[11] When a new content of the $Al_{1-x}Si_x$ solid solution was obtained in the sample, it was cooled down to room temperature, after which $S(T)$ was measured over a temperature interval of 4.2 to 300 K. The measurement data are given in Fig. 4 (the numbers of the curves correspond to the real-time sequence of studied intermediate states).

In each intermediate state, the resistivity $\rho(T)$ and Hall constant R_H were measured independently at room and liquid-helium temperatures. These data allow us to ascribe to each sample state both a superconducting transition temperature T_c and silicon content in the $Al_{1-x}Si_x$ solid solution.[7]

In considering the set of $S(T)$ curves in Fig. 4, note the similarity between curve 3 in Fig. 4 and curve 5 in Fig. 3, which correspond to states with close $T_c \approx 4.1$ K and the same silicon content $x=0.06$. Similarly, $S(T)$ curves of

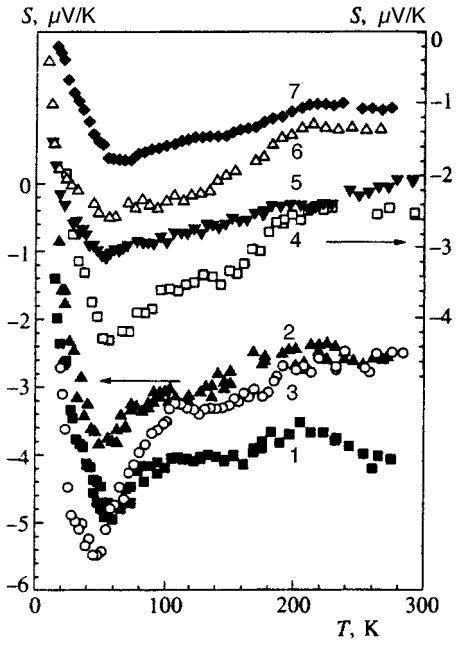


FIG. 4. Curves of thermopower of $Al_{0.92}Si_{0.08}$ versus temperature obtained in different intermediate states of the sample annealing: 1) $T_c=5.4$ K; 2) $T_c=5.0$ K; 3) $T_c=4.05$ K; 4) $T_c=3.75$ K; 5) $T_c=3.18$ K; 6) $T_c=1.87$ K; 7) $T_c=1.2$ K.

samples with $T_c \approx 2.5-3$ K ($x \approx 0.04-0.05$, curve 4 in Fig. 3 and curve 5 in Fig. 4) show close values of thermopower and have similar behavior. At the same time, changes in the amplitude and shape of the $S(T)$ curve for $Al_{1-x}Si_x$ and aging Al–Si alloys in the late stages of decay, when the impurity content in the solid solution and number and dimensions of inclusions increase, are quite considerable. These differences must probably be attributed to effects of boundary regions around submicron Si inclusions in the Al-based matrix, where the FCC structure of the solid solution is notably distorted.

4. DISCUSSION OF RESULTS

One of the key factors in the analysis of the thermopower coefficient in the low-temperature range is selection of the proper procedure for separating contributions from the phonon-drag effects, $S_{ph}(T)$, and diffusion component, $S_d(T)$. The diffusion thermopower is usually estimated by the Mott formula

$$S_d=\frac{\pi^2 k_B T}{3e}\left[\frac{\partial \ln \sigma(E)}{\partial E}\right]_{E=E_F}, \qquad (2)$$

where k_B is Boltzmann's constant, e is the electron charge, σ is the conductivity, and E_F is the Fermi energy.

Theoretical studies of the diffusion thermopower at high and low temperatures have led to the conclusion[23] that in these limiting cases the corrections to S_d due to effects of electron–phonon interaction, including nonelastic electron–phonon scattering, are inessential, and Eq. (2) is asymptotically exact. The most difficult region for interpretation of thermopower measurements in metallic systems is the range of intermediate temperatures, where, along with phonon-drag

effects[13] in materials with strong electron–phonon interaction, nonmonotonic behavior of the diffusion thermopower has been detected.[4] Among the widely known materials in which the relation between electron–phonon interaction and anomalies in $S_d(T)$ has been well established, the most notable examples are metallic glasses,[24–26] Chevrel's phases,[5,27] fullerenes,[28] and some oxide high-temperature superconductors.[4,29] An important point is that most of these materials are superconductors in which the proximity to a structural instability, hence the enhancement of the electron–phonon interaction, is a major cause of emergence and strengthening of superconductivity.[6]

Using Kaiser's approach, we can describe the anomalies in the diffusion thermopower at low and intermediate temperatures by the expression[4,5]

$$S_d^{\text{e-ph}} = S_d - S_B = a\lambda\lambda_S(T), \tag{3}$$

where S_B is the diffusion thermopower without taking into account an electron–phonon interaction, which is a linear function of temperature, λ is the dimensionless constant of electron–phonon interaction:

$$\lambda \equiv 2\int \frac{d\omega}{\omega}\alpha^2(\omega)F(\omega), \tag{4}$$

λ_S is the normalization factor describing the enhancement of electron–phonon interaction, which is a function of temperature:

$$\lambda_S = \int \frac{d\omega}{\omega}\alpha^2(\omega)F(\omega)G_S\frac{\hbar\omega}{k_BT}\Big/\int \frac{d\omega}{\omega}\alpha^2(\omega)F(\omega), \tag{5}$$

$G_S(\hbar\omega/k_BT)$ is a known universal function,[30] $\alpha^2(\omega)F(\omega)$ is the Eliashberg function.

In this connection, a correct incorporation of $S_B(T)$ and $S_{\text{ph}}(T)$ is sufficient for separating $S_d^{\text{e-ph}}$. Relying on the arguments about the effect of structural scattering on the phonon-drag thermopower in Al and results of Refs. 1 and 5, we will use measurements of $S(T)$ in aluminum thermally processed under high pressure (curve 2 in Fig. 3) as estimates of the sum of $S_B(T)$ and $S_{\text{ph}}(T)$ in separating $S_d^{\text{e-ph}}$. Really, the thermopower due to phonon drag (the down-peak at $T_{\min}=80$ K in Fig. 3) is suppressed by the structural disorder in Al due to structural defects generated by thermal treatment under pressure. When Al is substituted with Si, additional disorder is introduced into the FCC lattice of the solid solution because of this doping. In this case, one should expect a further drop in $S_{\text{ph}}(T)$ for $Al_{1-x}Si_x$ owing to a decrease in the carrier mean free path.[9]

Moreover, spectra of $L_{\text{II–III}}$ X-ray emission from aluminum, NMR Knight shift measurements, and data on the low-temperature specific heat of Al[10] indicate that the Fermi energy E_F and electron density of states $N(E_F)$ change only slightly if the FCC structure persists in $Al_{1-x}Si_x$. Given that Eq. (2) holds, this means that changes in $S_B(T)$ should be small in $Al_{1-x}Si_x$.

Thus, the suggested estimate of the sum of $S_{\text{ph}}(T)$ and $S_B(T)$ in the thermopower of $Al_{1-x}Si_x$ based on measurements of $S(T)$ (curve 2 in Fig. 3) yields the upper limit for this sum.

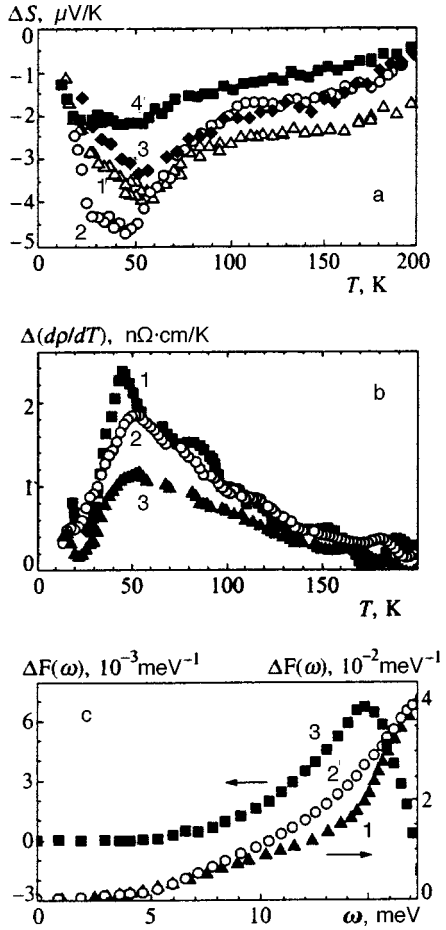


FIG. 5. Parameters of intermediate states of $Al_{0.92}Si_{0.08}$ vs. temperature: (a) $\Delta S(T)=S_d^{\text{e-ph}}(T)$ at (1) $T_c=5.4$ K, (2) 4.05 K, (3) 3.75 K, (4) 3.18 K; (b) $\Delta(d\rho/dT)$ at (1) $T_c=5.4$ K, (2) 3.75 K, (3) 1.87 K; (c) phonon density of states: (1) pure aluminum; (2) $Al_{0.094}Si_{0.06}$; (3) changes in the density of states $\Delta F(\omega)$ taken from Ref. 31.

The parameter $S_d^{\text{e-ph}}$ as a function of temperature obtained using the separation procedure described above is plotted in Fig. 5a for several $Al_{1-x}Si_x$ compounds. Figure 5b shows $\Delta(d\rho/dT)$ vs. temperature, which is a change in the resistivity temperature coefficient for several intermediate states of $Al_{0.92}Si_{0.08}$ obtained in the process of annealing with respect to $d\rho/dT$ for the ultimate, equilibrium state of the Al–Si alloy. We emphasize that the curves of $\Delta(d\rho/dT)$ in Fig. 5b show low-temperature anomalies in the diffusion component of the resistivity (see also Ref. 10), whose position and amplitude correlate with features in $S_d^{\text{e-ph}}$ in Fig. 5a.

Within the framework of the model based on Eqs. (3)–(5) the emergence of the down-peak in $S_d^{\text{e-ph}}$ must probably be attributed to features of the parameter $\lambda_S(T)$, and in particular, of the Eliashberg function $\alpha^2(\omega)F(\omega)$. On the other hand, studies of inelastic neutron scattering in $Al_{1-x}Si_x$ revealed[31] that substitution of aluminum with silicon leads to softening of phonon modes in the solid solution, but no radical changes in the phonon density of states were detected (see Fig. 5c). In this case, the emergence of the low-temperature feature in $S_d^{\text{e-ph}}$ of $Al_{1-x}Si_x$ (Fig. 5a) can be ascribed to the enhancement of the electron–phonon interac-

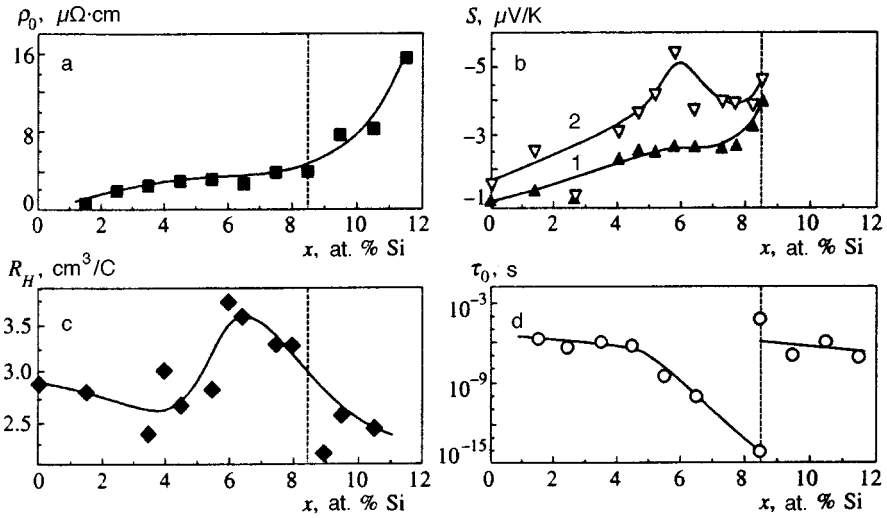


FIG. 6. Parameters of $Al_{1-x}Si_x$ solid solutions vs. concentration: (a) residual resistivity $\rho_0(T)$ taken from Ref. 10; (b) thermopower $S(T)$ at (1) room temperature and (2) in the vicinity of the $S(T)$ minimum in Fig. 3; (c) Hall coefficient $R_H(x)$ at room temperature; (d) reciprocal frequency of atomic motion.[11]

tion with ''soft'' modes generated in the FCC lattice owing to substitution effects.

We stress that a higher density $F(\omega)$ of phonon states in the low-frequency band of the spectrum is not a decisive factor in superconductivity enhancement of aluminum-based solid solutions. For example, the emergence of a localized mode in $Al_{1-x}Ge_x$ caused by substitution of aluminum with germanium results in a considerably smaller growth in T_c than in the case of $Al_{1-x}Si_x$ at an equal content of silicon in the FCC matrix of the solid solution.[32]

An estimate of the change in λ in $Al_{1-x}Si_x$ from $\lambda(Al)=0.38$ to $\lambda(Al_{0.92}Si_{0.08})\approx 0.9$ given in Ref. 10 allows us to interpret the increase in the magnitude of the negative thermopower of $Al_{1-x}Si_x$ at room temperature in terms of Eqs. (3)–(5) (Figs. 3–5). Since both the dimensionless factor λ of the electron–phonon interaction defined by Eq. (4) and the temperature-dependent parameter $\lambda_S(T)$ given by Eq. (5) change considerably when the concentration x in $Al_{1-x}Si_x$ is varied, it seems reasonable to relate not only the enhancement of superconductivity, but also the emergence and growth of low-temperature features of the diffusion thermopower to changes in the electron–phonon interaction parameter $\alpha^2(\omega)$. At the same time, integral equations (3)–(5) show little promise for an accurate quantitative description of the relation between the low-temperature anomalies of the diffusion thermopower and parameter $\alpha^2(\omega,T)$. It is probable that these difficulties are caused by the limits of the approximation of weakly nonequilibrium states of metallic systems used in deriving Eqs. (3)–(5).

Interestingly enough, similar anomalies of $\alpha^2(\omega)$ were predicted by Weber[33] for oxide high-temperature superconductors $La_{2-x}(Ba,Sr)_xCuO_4$ using the model of nonorthogonal tight binding in lattice dynamics. In this case,[33] the development of a strong Peierls instability in the $La_{2-x}(Ba,Sr)_xCuO_4$ lattice due to changes in its composition leads to emergence of a relatively small number of phonon modes in the spectrum $F(\omega)$, which are characterized by anomalously strong interaction with conduction electrons.

A different approach to the thermopower in disordered metallic systems with strong electron–phonon interaction based on solutions of the quantum kinetic equation has been proposed recently.[34–36] It was shown,[34] in particular, that the effect of electron–phonon–impurity interference must be taken into account, along with the thermopower renormalization due to electron–phonon interaction in all orders of the perturbation theory. As a result, both the magnitude and sign of $S(T)$ strongly depend on the impurity scattering amplitude of electrons; moreover, contributions to the impurity thermopower due to the electron–phonon interaction and electron–phonon–impurity interference have similar structure and cannot be separated by processing experimental data.[34,35] The calculations described in Refs. 34 and 35 have been applied recently to $S(T)$ in oxide high-temperature superconductors and solid substitutional solutions based on tin and lead.[36]

On the other hand, our preliminary analysis based on the approach of Refs. 34–36 indicates that, if integral equations (1)–(5) of Ref. 36 are applied to the quantitative description of the experimental data plotted in Figs. 3 and 4, there is a problem discussed previously in connection with the interpretation of $S(T)$ in terms of Kaiser's model. If the Debye phonon spectrum, which was used in Ref. 36, is replaced with the real phonon density of states $F(\omega)$ for $Al_{1-x}Si_x$ solid solution,[31] it has little impact on the result. As a result, the issue of accurate quantitative interpretation of low-temperature anomalies in the diffusion thermopower in nonequilibrium $Al_{1-x}Si_x$ interstitial solid solutions in close proximity to the lattice instability has remained unsettled.

Note that, in addition to the anomalous behavior of the $Al_{1-x}Si_x$ thermopower near the lattice instability (Fig. 6b), a significant nonlinearity in the residual resistivity $\rho_0(T)$ (Fig. 6a) has been detected in the range of concentration $x=5-8$ at. %, as well as a maximum in the Hall coefficient $R_H(x)$ (Fig. 6c). Thus far, all the anomalies of kinetic coefficients listed above, which are caused by the proximity to the spinodal, and hence, related to high atomic mobility (Fig. 6d) in the FCC lattice of model $Al_{1-x}Si_x$ solid solutions, have not found their adequate theoretical interpretation.

5. CONCLUSIONS

In this paper, we have shown that some features of thermopower in nonequilibrium $Al_{1-x}Si_x$ substitutional solid so-

lutions near the lattice instability, namely the low-temperature down-peak in $S(T)$ and substantial increase in the coefficient of its linear term, must be attributed to anomalies in the diffusion component of thermopower. Structural disordering caused by thermal processing of aluminum samples under high pressure and substitution of aluminum with silicon in the FCC lattice leads to suppression of the thermopower due to phonon drag, which dominates in pure aluminum. Our analysis has allowed us to relate the thermopower anomalies near the lattice instability to the enhancement of electron–phonon interaction with ''soft modes'' in the FCC lattice of $Al_{1-x}Si_x$ solid solutions.

Thus, the experimental data of Refs. 9–12 combined with the results of the reported study has allowed us to establish a relationship between parameters of the superconducting and normal states in the $Al_{1-x}Si_x$ model system, where changes in x lead to growth in the superconducting transition temperature by an order of magnitude, although the crystal lattice and electron spectrum are essentially the same.

We are indebted to A. K. Savchenko, J. Singleton, W. Hayes, A. O. Orlov, and G. Snider for many helpful discussions. The work was supported by the Russian Fund for Fundamental Research (Grant 96-16176), by the *Fullerenes and Atomic Clusters* program sponsored by the Russian Ministry of Science and Technology, and by INTAS (Grants 93-2400 and 94-4435).

*Lab. voor Vaste-Stoffysica en Magnetisme, K. U. Leuven, Celestijnenlaan 200 D, B-3001 Leuven, Belgium.

[1] C. Shearwood and D. Greig, Solid State Commun. **71**, 207 (1989).
[2] E. C. Jones, D. P. Norton, B. C. Sales *et al.*, Phys. Rev. B **52**, R743 (1995).
[3] I. Inabe, H. Ogata, Y. Maruyama *et al.*, Phys. Rev. Lett. **69**, 3797 (1992).
[4] A. B. Kaiser and G. Mountjoy, Phys. Rev. B **43**, 6266 (1991).
[5] A. B. Kaiser, Phys. Rev. B **35**, 4677 (1987).
[6] *The Problem of High-Temperature Superconductivity* [in Russian], V. L. Ginzburg and D. A. Kirzhnits (eds.), Nauka, Moscow (1977).
[7] V. F. Degtyareva, G. V. Chipenko, I. T. Belash *et al.*, Phys. Status Solidi A **89**, K127 (1985).
[8] J. Chevrier, D. Pavuna, F. Cyrot-Lackmann, Phys. Rev. B **36**, 9115 (1987).
[9] V. V. Brazhkin, V. V. Glushkov, S. V. Demishev *et al.*, J. Phys.: Condens. Matter **5**, 5933 (1993).
[10] N. E. Sluchanko, V. V. Glushkov, S. V. Demishev *et al.*, Phys. Rev. B **51**, 1112 (1995).
[11] N. E. Sluchanko, V. V. Glushkov, S. V. Demishev *et al.*, J. Phys.: Condens. Matter **6**, 9079 (1994).
[12] N. E. Sluchanko, V. V. Glushkov, S. V. Demishev *et al.*, Phys. Rev. B **53**, 11304 (1996).
[13] F. J. Blatt, P. A. Schröder, C. A. Foils, and D. Greig, *Thermoelectromotive Force of Metals* [Russian translation], Metallurgiya, Moscow (1980).
[14] Ch. Laurent, S. K. Patapis, M. Laguesse *et al.*, Solid State Commun. **66**, 445, (1988).
[15] L. G. Khvostantsev, L. F. Vereschagin, and A. P. Novikov, High Temp.-High Press. **9**, 16 (1977).
[16] R. P. Huebener, Phys. Rev. **171**, 634 (1968).
[17] A. W. Dudenhoeffer and R. R. Bourassa, Phys. Rev. B **5**, 1651 (1972).
[18] G. Sieber, G. Wehr, and K. Boning, J. Phys. F **7**, 2503 (1977).
[19] T. Ribka and R. R. Bourassa, Phys. Rev. B **8**, 4449 (1973).
[20] O. I. Barkalov, I. T. Belash, A. I. Kolesnikov *et al.*, J. Phys.: Condens. Matter **192–193**, 486 (1995).
[21] R. Borrelly, J. M. Pelletier, and P. F. Gobin, Acta Metall. **26**, 1863 (1978).
[22] F. Sachslehner, C. Wutte, and F. Stangler, Phys. Status Solidi A **120**, 581 (1990).
[23] M. Johnson and G. D. Mahan, Phys. Rev. B **42**, 9350 (1990).
[24] B. L. Gallagher and D. Greig, J. Phys. F **12**, 1721 (1982).
[25] A. B. Kaiser, J. Phys. F **12**, L223 (1982).
[26] A. B. Kaiser and G. E. Steadman, Solid State Commun. **54**, 91 (1985).
[27] V. Vasudeva Rao, G. Sangarajan, and R. Srinivasan, J. Phys. F **14**, 973 (1984).
[28] D. T. Morelli, Phys. Rev. B **49**, 655 (1994).
[29] A. B. Kaiser, Phys. Rev. B **37**, 5924 (1988).
[30] A. B. Kaiser, Phys. Rev. B **29**, 7088 (1984).
[31] J. Chevrier, J. B. Suck, J. J. Capponi *et al.*, Phys. Rev. B **49**, 961 (1994).
[32] A. I. Kolesnikov, O. I. Barkalov, I. T. Belash *et al.*, J. Phys.: Condens. Matter **5**, 4737 (1993).
[33] W. Weber, Phys. Rev. Lett. **58**, 1371 (1987).
[34] M. Yu. Reizer and A. V. Sergeev, Zh. Éksp. Teor. Fiz. **93**, 2191 (1987) [Sov. Phys. JETP **93**, (1987)].
[35] A. V. Sergeev, M. Yu. Reizer, and D. V. Livanov, Phys. Rev. B **50**, 18694 (1994).
[36] D. V. Livanov and A. V. Sergeev, Phys. Rev. B **48**, 13137 (1993).

Translation provided by the Russian Editorial office

NMR in $^{55}Mn^{2+}$ nuclei in the quasi-one-dimensional antiferromagnetic $CsMnBr_3$

A. S. Borovik-Romanov,*) S. V. Petrov, and A. M. Tikhonov

Kapitsa Institute of Physics Problems, Russian Academy of Sciences, 117334 Moscow, Russia

B. S. Dumesh

Institute of Spectroscopy, Russian Academy of Sciences, 142092 Troitsk, Moscow Region, Russia
(Submitted 23 July 1997)
Zh. Éksp. Teor. Fiz. **113**, 352–368 (January 1998)

The NMR spectrum of the quasi-one-dimensional easy-plane antiferromagnetic $CsMnBr_3$, which has trigonal spin lattice, is investigated in detail. The measurements were performed on a wide-band NMR decimeter microwave-band spectrometer over a wide range of magnetic fields at temperatures 1.3–4.2 K. All three branches of the NMR spectrum previously found by us [JETP Lett. **64**, 225 (1996)] are severely distorted because of the dynamic interaction with the Goldstone mode in the antiferromagnetic resonance spectrum. The experimental results in fields up to 40 kOe are described satisfactorily by an equation obtained by Zaliznyak *et al.* [JETP Lett. **64**, 473 (1996)]. Formulas are obtained in our work that agree very well with experiment at all fields up to the ''collapse'' field H_c of all sublattices. The unbiased NMR frequency in $CsMnBr_3$ is determined to be $\nu_{n0}=416$ MHz ($T=1.3$ K) in zero external magnetic field, and in this way the reduction in the spontaneous moment due to the quasi-one-dimensional nature of the system of Mn^{2+} spins, which according to our data amounts to 28%, is determined more accurately. The field dependences of the directions of the magnetic sublattices with respect to the magnetic field are obtained from the NMR spectra, confirming the equations of Chubukov [J. Phys. Condens. Matter **21**, 441 (1988)]. The results on the field dependence of the width and intensities of the NMR lines are discussed, along with three observed anomalies: 1) a strong increase in the NMR frequency for nuclei in sublattices that are perpendicular to the magnetic field; 2) the nonmonotonic temperature dependence of the resonance field for the lower branch of the spectrum; 3) the presence of two branches of the NMR spectrum in large H_c fields, in which the $CsMnBr_3$ must be a quasi-one-dimensional antiferromagnetic. [S1063-7761(98)02401-9]

1. INTRODUCTION

The quasi-one-dimensional antiferromagnetic $CsMnBr_3$ ($T_N=8.3$ K) has been vigorously investigated in recent years, both theoretically and experimentally. A number of extremely interesting properties have been found in the electron spin system of this material: quasi-one-dimensional exchange interaction and trigonal magnetic structure,[1] a phase transition to a quasicollinear structure in a comparatively weak magnetic field[2] H_c, and the electron susceptibility anisotropy[3,4] associated with the latter, a large reduction in the electron spin moment,[1,4] and the presence of the Goldstone mode in the antiferromagnetic resonance (AFMR) spectrum,[5,6] the frequency of which is proportional to the cube of the magnetic field. This paper is devoted to an investigation of the unique features of the NMR spectrum of the nuclei of $^{55}Mn^{2+}$ magnetic ions, a brief observational description of which has been given elsewhere.[7]

There are several distinctive features of NMR in the nuclei of magnetic ions in antiferromagnetics.[8] The NMR frequency of the nuclei of magnetic ions is determined primarily by the magnitude of the hyperfine field, which for the $3d$ elements is proportional to the average electron spin $\langle S\rangle$. In particular, for $^{55}Mn^{2+}$ ions this field can amount to 600–700 kOe. Accordingly, NMR is observed in $^{55}Mn^{2+}$ nuclei in relatively weak fields at very high frequencies (up to 700 MHz).

A strong anisotropy of the exchange interaction (the ratio of the exchange integrals along the chains to their value perpendicular to the chains ≈ 500) is preserved in $CsMnBr_3$ despite the establishment of three-dimensional ordering. Therefore, a considerable reduction must occur in the mean spin $\langle S\rangle$ of the Mn^{2+} magnetic ion because of the presence of large quantum fluctuations induced by the quasi-one-dimensional nature of the exchange interaction. The NMR frequency at the nuclei of the magnetic ions in zero magnetic field, as we have pointed out, is determined by $\langle S\rangle$. Therefore, an investigation of the NMR spectrum is one of the most accurate methods of determining $\langle S\rangle$ and, accordingly, the reduction in the spin moment. Determination of the electron spin reduction in $CsMnBr_3$ was one of the goals of this work.

$CsMnBr_3$ is an easy-plane antiferromagnetic with trigonal magnetic structure. When an external magnetic field is applied in any direction in the easy plane, the NMR spectrum must split into three branches. By investigating the dependence of the resonance frequencies of these branches on the applied field, one can trace the transition of the trigonal antiferromagnetic structure into quasi-collinear. The second goal of our work was to investigate this process and to compare the data obtained with the prediction of the Chubukov

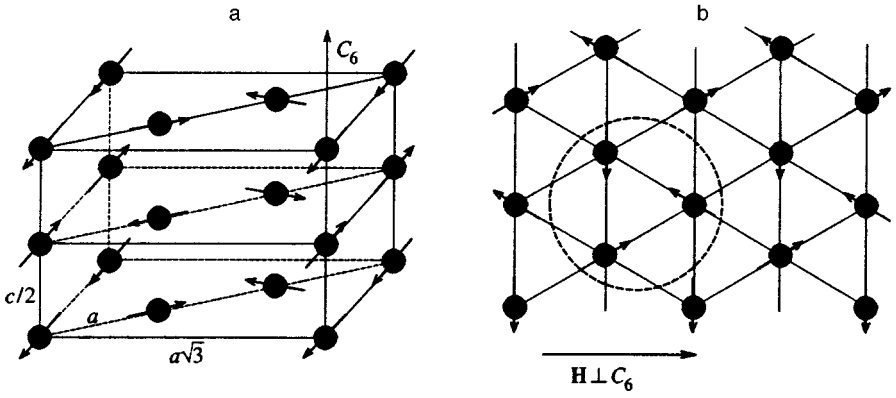


FIG. 1. Magnetic structure of the $CsMnBr_3$ antiferromagnetic (a—unit magnetic cell; b—orientation of spins in hexagonal plane with respect to the applied field **H**).

theory.[5] It must be pointed out that $CsMnBr_3$ is the first antiferromagnetic in which NMR could be observed in the nuclei of all sublattices forming the trigonal magnetic structure.

Another unique feature of the resonance properties of antiferromagnetics is associated with the strong correlation of the vibrations of the nuclei with the vibrations of the electron system. The effect of this correlation on the AFMR was first established by Heeger, Portis, Teaney and Witt,[9] who found a strong temperature dependence for the location of the AFMR line in the $KMnF_3$ antiferromagnetic and observed a double electron-nuclear resonance. A thorough theoretical analysis was performed by de Gennes *et al.*[10] and it was shown that the strong dependence of the AFMR on the temperature of the nuclear magnetic system is direct evidence of the interaction of the nuclear and ionic spins, leading to a dynamic frequency shift (DFS): to a pulling of their spectra.

Since $CsMnBr_3$ is an easy-plane antiferromagnetic, one could expect the appearance of a pulling effect in this material too. After we had found the strong distortions of the NMR spectrum due to the dynamic interaction of the electron and nuclear spins (we reported this elsewhere[7]), Zaliznyak and Zorin[11] found that in the AFMR spectrum there is a gap, which is also caused by a pulling of the spectral branches. In this paper we compare the results obtained by the NMR and AFMR methods.

Greater absorption of a radio-frequency field, associated with an amplification of the amplitude of the pumping field at the nuclei of the magnetic ions because of the dynamic component H_{hf} of the hyperfine field,[8] is observed in magnetically ordered crystals. The magnitude of this amplification depends on the form of the AFMR spectrum.

2. CRYSTALLINE AND MAGNETIC STRUCTURE OF THE ANTIFERROMAGNETIC $CsMnBr_3$

$CsMnBr_3$ belongs to the large family of binary halides of type ABX_3, where A is an alkali metal, B is a $3d$ metal, and X is a halogen. The crystal structure is described by the D_{6h}^4 spatial symmetry group with lattice parameters $a=7.61$ Å and $c=6.52$ Å.[12] The Mn atoms in the plane perpendicular to the C_6 axis form a hexagonal lattice.

The crystal lattice determines the specific magnetic structure of this substance, which is determined in experiments on the elastic scattering of neutrons.[1] Because of the fact that the distance between adjacent Mn atoms along the C_6 axis is one-half that in the plane, the exchange interaction integral $J=214$ GHz determined in neutron scattering experiments,[13] which characterizes the energy of the antiferromagnetic exchange interaction along the hexagonal axis, is several hundred times J', which characterizes the antiferromagnetic interaction in the perpendicular plane. The value $J'=0.46$ GHz is obtained from neutron diffraction data[13] and $J'=0.5$ GHz from AFMR data.[6] At $T_N=8.3$ K, three-dimensional antiferromagnetic ordering occurs in the Mn^{2+} spin system. The anisotropy energy, characterized by the constant D ($D=2.9$ GHz[2] and 2.4 GHz[6]), establishes the directions of all spins in the chains perpendicular to the sixth-order symmetry axis. The weak exchange interaction J' between spins lying in one plane leads to the appearance of trigonal 120-degree magnetic structure. Thus, the magnetic structure of $CsMnBr_3$ can be considered a set of one-dimensional antiferromagnetic chains, elongated along the C_6 axis and interacting weakly among themselves. The unit magnetic cell is shown in Fig. 1a.

Six sublattices form the magnetic structure. Since anisotropy is essentially absent in the basis plane, all of the Mn^{2+} magnetic moments are expanded even in a small constant magnetic field H, applied in the plane, so that the magnetic field direction coincides with one of the bisectors of the triangle (see Fig. 1b and Fig. 2a), i.e., the magnetizations M_1, M_4 will be perpendicular to the field while the other two pairs of sublattices ($M_{2,6}$ and $M_{3,5}$) will form angles of $\pm\pi/6$ with the field direction. On the whole, however, the angles between the directions of adjacent Mn^{2+} spins in the plane will remain close to 120°. Increasing the value of H will lead to a decreasing angle α between the sublattices M_2 and M_3 (the same as for M_5 and M_6), as shown by Chubukov:[6]

$$\cos\frac{\alpha}{2}=\frac{1}{2-z},\quad z=\frac{H^2}{H_c^2}, \tag{1}$$

with a vanishing in a field $H_c=(48JJ')^{1/2}S\approx 61$ kOe (experimental value[4] of $H_c\approx 64$ kOe for $T=1.8$ K). Moreover, in each chain a very slight rotation of all spins occurs toward the direction H, which is due to the finiteness of the magnitude of the large exchange interaction field H_E, which is $H_E=8JS\approx 1500$ kOe. Thus, in the magnetic field H_c a col-

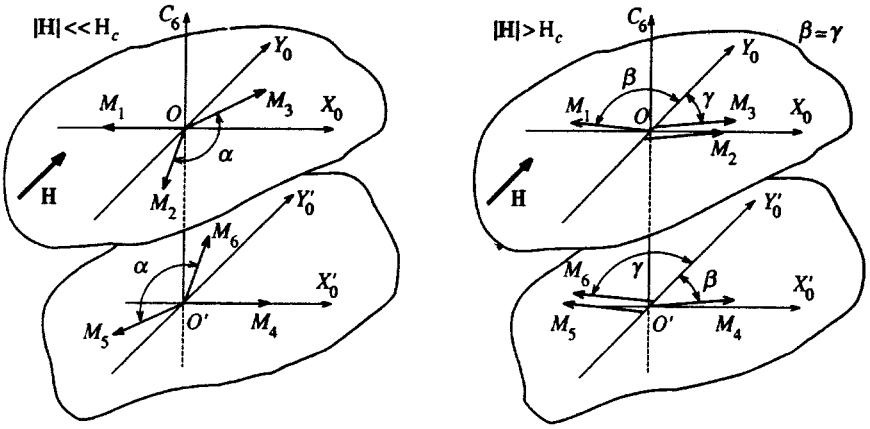


FIG. 2. Schematic representation of the magnetic structure of $CsMnBr_3$ (a—$H \ll H_c$; b—$H > H_c$).

lapse of two pairs of sublattices occurs (a phase transition of the second kind), and the magnetic structure is transformed from six-sublattice to quasi-two-sublattice (Fig. 2b).

The difference between the angles β and γ is negligible. A further increase in the magnetic field should lead to a rotation of all spins toward the direction of the vector **H** until there is a complete breakdown of the antiferromagnetic structure through a spin-flip transition. Measurements of the magnetic moment of $CsMnBr_3$ as a function of the magnetic field (in fields up to 80 kOe) in Refs. 3 and 4 agree for the most part with theory.[5]

The hyperfine interaction energy per $3d$ Mn^{2+} ion in the magnetically ordered state can be written in the form (see Ref. 8)

$$\mathscr{H}_{hf} = A(\langle \mathbf{I} \rangle, \langle \mathbf{S} \rangle), \tag{2}$$

where $\langle \mathbf{S} \rangle$ and $\langle \mathbf{I} \rangle$ are the mean spin of the electron and nucleus of the $^{55}Mn^{2+}$ ions and A is the hyperfine interaction constant ($A<0$ for Mn^{2+}). The static field at the nuclei of the $^{55}Mn^{2+}$ ion (in the absence of an external magnetic field) is

$$\mathbf{H}_{hf} = -\frac{A}{\gamma_n h} \langle \mathbf{S} \rangle, \tag{3}$$

γ_n is the nuclear gyromagnetic ratio ($\gamma_n = 1.06$ MHz/kOe for ^{55}Mn).

Assuming the hyperfine constant A is independent of the sublattice number (and taking into consideration that $A<0$), the magnitudes of the local fields H_{ni0}, which act on the nuclei of the magnetic ions for all six sublattices, have the form

$$H_{ni0} = |\mathbf{H}_{hf} + \mathbf{H}| = H_{hf}\left(1 + \frac{H^2}{H_{hf}^2} - 2\frac{H}{H_{hf}} \cos\,\theta_i\right)^{1/2}, \tag{4}$$

where θ_i is the angle between the external magnetic field vector and the sublattice magnetization. The nuclei of the ions of the sublattices M_1 and M_4 correspond to $i=1$, M_2 and M_5 to $i=2$, and M_3 and M_6 to $i=3$. Let us write the cosines of the angles θ_i in explicit form:

$$\cos\,\theta_1 = \frac{H}{H_E},$$

$$\cos\,\theta_2 = \sin\frac{\alpha}{2} + \frac{H}{H_E}\cos^2\frac{\alpha}{2} + o\left(\frac{H}{H_E}\right),$$

$$\cos\,\theta_3 = -\sin\frac{\alpha}{2} + \frac{H}{H_E}\cos^2\frac{\alpha}{2} + o\left(\frac{H}{H_E}\right). \tag{5}$$

Thus, the NMR frequency in $CsMnBr_3$ in the absence of a dynamic shift is

$$\omega_{ni0} = \gamma_n H_{ni0}. \tag{6}$$

Figure 3 shows the NMR spectrum in $CsMnBr_3$ (4) predicted by theory (ignoring pulling). For the assumptions made previously, the NMR spectrum in fields $H<H_c$ ($\mathbf{H} \perp C_6$) will consist of three branches. In a field $H>H_c$ the NMR spectrum should consist of one branch because of the very small difference between the angles β and γ (the appearance of the dashed curve 1 in place of the theoretical $1'$ will be explained below).

3. SAMPLES

The samples were prepared by the Bridgeman method. Fabrication of the $CsMnBr_3$ crystals has been described in detail elsewhere.[4] The $CsMnBr_3$ crystals grown by this method are transparent and are easily cleaved along planes perpendicular to the binary axes. The C_6 axis uniquely determines the intersection of the cleavage planes. The

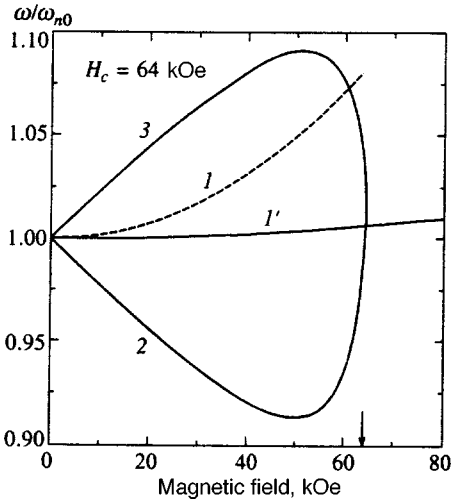


FIG. 3. NMR spectrum of $CsMnBr_3$ ignoring DFS. *1′, 2, 3*—branches of NMR spectrum according to Chubukov's angular dependences,[5] *1*—empirical relationship for unshifted frequency of middle branch (see below).

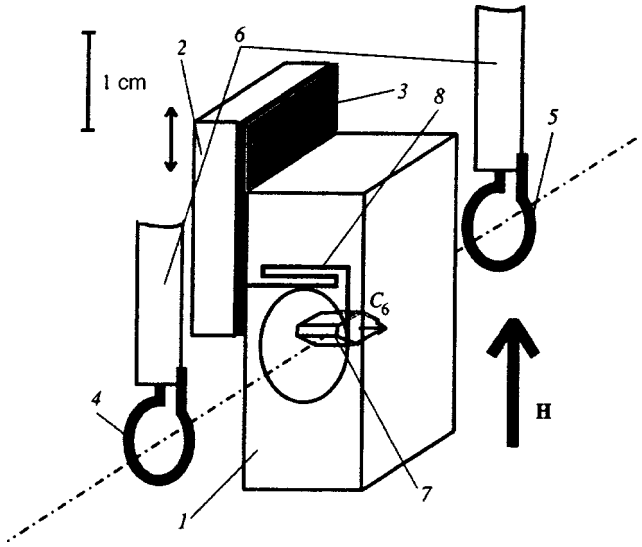


FIG. 4. Resonant circuit: *1*—cavity; *2*—plate; *3*—thin PETE film; *4,5*—coupling loops; *6*—coaxial feed lines; *7*—sample; *8*—narrow slot.

$CsMnBr_3$ crystalline samples are extremely hygroscopic and are hydrated quite rapidly in open air, turning into a white powder that is probably $CsMnBr_3 \cdot 2H_2O$. Therefore, single-crystal samples cleaved from a large crystal were coated with a resin cement. The protective cladding produced in this manner made it possible to work with one $CsMnBr_3$ single crystal for a long period of time.

4. NMR APPARATUS AND MEASUREMENT PROCEDURE

In view of the strong dependence of the NMR signal on the magnitude of the field **H**, a wideband cw decimeter spectrometer of the type described in Ref. 14 with a high-Q cavity was built to investigate nuclear resonance in $CsMnBr_3$. The Q of the resonant section in the frequency range investigated was 300–400.

The copper cavity (*1* in Fig. 4) was a modified version of a tunable cavity of the ''split-ring'' type.[15] Three cavities with differing slot geometry 8 (the size of the gap in the slot is ~0.09 mm), in which mica plates were placed to reduce the frequency, were used to cover the range from 500 to 200 MHz. By moving plate 2 one can change the capacitance between plate 2 and cavity 1 (the insulator 3 is a 5–10 μm thick film of polyethylene terephthalate (trade-named Lavsan), thereby changing the resonant frequency. Microwave power is supplied by the coaxial line 6. Coupling to the resonator is accomplished by means of single-turn coils, one of which is the transmitting loop 4, while the other is the receiving loop 5. The diameter of the coupling loops is 5 mm. The direction of the external magnetic field **H**, produced by a superconducting solenoid, is indicated in Fig. 4. The C_6 axis of the $CsMnBr_3$ single crystal 7 was perpendicular to **H**. The alignment accuracy of the C_6 axis with respect to the solenoid axis was $\sim \pm 3°$, which did not greatly increase the error in determining the resonant field. The sample was mounted on a special fluoroplastic substrate. The entire system was immersed in a helium bath.

Figure 5 is a block diagram of the spectrometer. The frequency of the microwave oscillator G (a Kh1-43 instrument for studying frequency response) was modulated by the low-frequency ($f = 45$ kHz) reference oscillator of the synchronous detector SD2 (PAR 5110 lock-in amplifier). This occurred by mixing the low-frequency signal with the control voltage U_{cont} that sets the frequency of the microwave oscillator. An automatic frequency control (AFC) system, tuned to the first harmonic of the modulation signal, was used to maintain the oscillator frequency at the resonance peak. The AFC system consists of synchronous detector SD1 and a cavity resonance tracker (ETC) with proportional and integral feedback channels.

The output power of the microwave oscillator was less than 3 mW. Absorption in the resonance channel was recorded at the second modulation harmonic U_{s2f} by means of SD2. The depth of the frequency modulation (0.3–3 MHz) was chosen so that it did not greatly broaden the resonant absorption line; it was typically ~1.5 MHz.

Attenuators At1 (10 dB) and At2 (10 or 3 dB) were

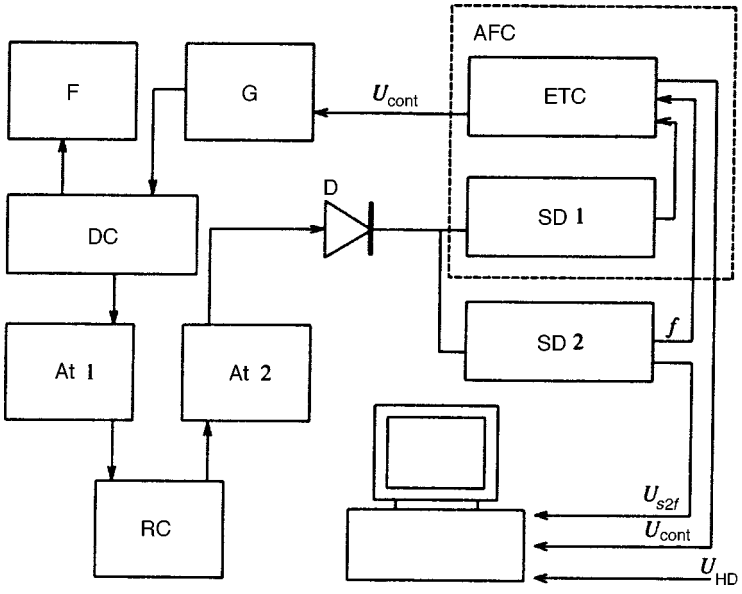


FIG. 5. Block diagram of spectrometer: RC—resonant circuit; G—microwave oscillator; D—detector; SD1—synchronous detector (UNIPAN 232B); SD2—synchronous detector (lock-in PAR 5110); F—Ch3-63/1 frequency meter; At1, At2—attenuators; DC—directional coupler; ETC—electronic tracking circuit; U_{cont}—control voltage; U_{HD}—signal from Hall detector; U_{s2f}—second harmonic signal.

corrected to the input and output of the low-temperature portion of the spectrometer in order to smooth its amplitude–frequency characteristic (AFC). The frequency of the oscillator G was monitored by the frequency meter F (Ch3-63/1), which was connected to the transmitter channel via the directional coupler DC (attenuation of the reflected wave was at least 25 dB).

Thus, the NMR spectrum could be measured in two modes: by scanning the magnetic field at fixed oscillator frequency, or by scanning the frequency at fixed magnetic field. Unfortunately, because of the nonlinearity and poor repeatability of the input line frequency response, this operating mode of the spectrometer could not be fully utilized. Moreover, the sensitivity of the spectrometer turned out to be perfectly adequate in the field-scanning mode. For the first operating mode the error in frequency measurement due to frequency instability of the resonant circuit (Fig. 4) was no greater than 0.1 MHz. The temperature was monitored by means of the resistance of a germanium resistor and the equilibrium saturated vapor pressure of ^{4}He. Its stability was no worse than ± 0.05 K.

The magnetic field intensity was measured with a Hall detector (U_{HD} in Fig. 5), which together with the $CsMnBr_3$ sample were located at equivalent positions inside a superconducting solenoid. The error in determining the magnetic field intensity did not exceed 1%.

To excite resonance and to obtain the maximum NMR amplification, the sample was oriented so that the field **H** of the solenoid and the radio frequency field **h** were mutually perpendicular in the hexagonal plane.

The mass of the $CsMnBr_3$ single crystals investigated ranged from 50 to 100 mg.

5. SPECTRUM OF COUPLED VIBRATIONS IN $CsMnBr_3$

Until recently the effect of dynamic interaction of the electron and nuclear systems of the Mn^{2+} ions on the AFMR spectrum was completely ignored during the investigation of the low-frequency resonance properties of the electron system in small-size noncollinear $CsMnBr_3$, $RbMnBr_3$ and $CsMnI_3$ antiferromagnetics. It was only after the experimental discovery of the pulling of the spectra in $CsMnBr_3$ (severely distorted NMR spectrum[7] and gap in the Goldstone branch of the AFMR spectrum[11]) in Ref. 11 that the spectrum of joint nuclear–electron vibrations was calculated under the assumption that the exchange trigonal structure is not distorted by the field. This condition is satisfied in fields up to ~ 40 kOe. In these fields the formulas of Ref. 11 describe our experimental results satisfactorily.

For a comprehensive description of the results obtained by us over the entire field range up to $H_c = 64$ kOe, we calculated the NMR spectrum using the low-frequency (Goldstone) AFMR mode $\omega_{e1}(H)$ calculated by Chubukov,[5] which is

$$\omega_{e1}(H) \to \gamma_e \sqrt{\frac{3}{4}} \frac{H^3}{H_c^2}. \tag{7}$$

as $H \to 0$.

For an easy-plane two-sublattice antiferromagnetic, it is known that its low-frequency vibration occurs with essentially no escape of spin magnetic moments from the easy plane: the ratio of the amplitude of the magnetization vibrations of the sublattices perpendicular to the plane to the amplitude in the plane is H/H_E. A similar relationship must be satisfied in the case of $CsMnBr_3$. Limiting consideration to only this vibration, the Lagrangian of the electron spins ζ_{e0} can be written in the form[16]

$$\mathscr{L}_{e0} = \frac{1}{2} \frac{\chi_\parallel}{\gamma_e^2} (\dot{\phi}^2 - \omega_{e1}^2(H)\phi^2), \tag{8}$$

where ϕ is the small deviation of each spin in the basis plane from the equilibrium direction, and $\chi_\parallel$ is the susceptibility along the C_6 axis. The quantity $\omega_{e1}(H)$ cannot be given in analytic form. It is the solution of a 6th-degree equation, which can be solved numerically. The form of the given Lagrangian, strictly speaking, is valid as long as no serious deformation of the trigonal magnetic structure occurs. We return to a discussion of this question below.

Typical longitudinal (T_1) and transverse (T_2) relaxation times for the nuclei of the Mn^{2+} ions are $T_1 \sim 1-10$ msec and $T_2 \sim 10$ μsec. These values are clearly several orders of magnitude greater than the period of the natural vibrations of the nuclear magnetization of $^{55}Mn^{2+}$, which is ~ 2 nsec. For this reason the magnetizations of the nuclear sublattices can be considered as conserving the magnetic moments, and the Lagrangian corresponding to their precession in the field $\mathbf{H}_{\text{eff}}$ can be represented in the form proposed for ferromagnetics in the theory of the macroscopic dynamics of magnetic substances.[17] In this case the Lagrangian for the nuclei of one subsystem becomes

$$\mathscr{L}_n = \frac{1}{\gamma_n} (\langle \mathbf{m} \rangle \times \mathbf{\Omega} + \gamma_n \mathbf{H}_{\text{eff}}), \tag{9}$$

where $\langle \mathbf{m} \rangle$ is the paramagnetic moment of the sublattice nuclei, $\mathbf{H}_{\text{eff}} = \mathbf{H} + \mathbf{H}_{hf}$ according to Eq. (4) (the dipole field, which does not exceed 2 kOe as a calculation has shown (see below), is ignored here, and $\mathbf{\Omega}$ is the angular velocity in spin rotation space.

By adding Eq. (8) and the sum of Eq. (9) for all six sublattices, we obtained the Lagrangian of the system of electron and nuclear spins of the antiferromagnetic being considered with the hyperfine interaction taken into account:

$$\mathscr{L}_{en} = \mathscr{L}_{e0} + \sum_{i=1}^{6} \mathscr{L}_{ni}. \tag{10}$$

The characteristic equation of the linearized system (in terms of the small angle of deviation of the spins from equilibrium) of Lagrange equations determines the spectrum of joint vibrations. This equation has the following form for the $CsMnBr_3$ magnetic structure:

$$\omega_{e1}^2 - \omega^2 = \frac{1}{3} \frac{\omega_{T0}^2}{\omega_{n0}^2} \sum_{i=1}^{3} \frac{(\omega^2 + \omega_{ni}^2 - \omega_{ni0}^2)\omega_{ni}^2}{(\omega_{ni0}^2 - \omega^2)},$$

$$\omega_{ni0}^2 = \gamma_n^2 (H_{hf})_i^2 \left(1 + \frac{H^2}{(H_{hf})_i^2} - 2 \frac{H}{(H_{hf})_i} \cos\ \theta_i \right),$$

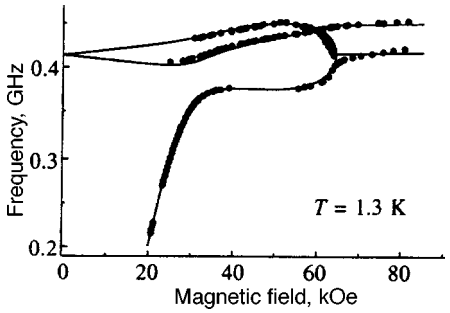


FIG. 6. NMR spectrum of $CsMnBr_3$: points—experimental NMR spectrum at $T=1.3$ K, lines—result of calculation using Eq. (11).

$$\omega_{ni}^2=\gamma_n^2(H_{hf})_i^2\left(1-\frac{H}{(H_{hf})_i}\cos\,\theta_i\right), \tag{11}$$

where the index i of the spin orientation of the Mn^{2+} ion with respect to the applied field runs through values from 1 to 3, $(H_{hf})_i$ is the value of the hyperfine field for nuclei at the ith position, which at some points, as an experiment has shown, depends on the magnitude of the applied magnetic field, $\omega_{n0}=\omega_{ni0}(H)|_{H=0}$ is the unbiased NMR frequency for any $i=1$, 2, 3, and $\omega_{T0}=\gamma_e H_{hf}\sqrt{\chi_n/\chi_\parallel}$ is the gap in the AFMR spectrum (the formula is exactly the same as in Ref. 11).

As already indicated, the Lagrangian of the electrons can apply rigorously only below 40–45 kOe. However, as will be seen from our experimental results, in fields greater than 50 kOe the dynamical interaction of the electron and nuclear systems almost completely vanishes, and it makes no contribution whatsoever to the solution of Eq. (11). Accordingly, the form of the electron Lagrangian exerts no influence whatsoever on these solutions in strong fields. Thus, the expressions for the NMR frequencies calculated from Eqs. (11) are rigorous in fields below 40 kOe and in fields greater than 50 kOe, but between these values they must be considered empirical approximations.

6. EXPERIMENTAL RESULTS AND DISCUSSION

We observed the NMR lines in a $CsMnBr_3$ single crystal over the very wide frequency range $\sim$200–450 MHz, in magnetic fields from 20 to 80 kOe. The basic experimental data were obtained at temperatures of 1.3 and 1.75 K. (Observations of the lower branch were also made at $T=2.5$, 3.0, 3.5 and 4.2 K.) The spectrum at $T=1.3$ K is shown by the points in Figs. 6 and 7.

a) *Shape of NMR spectrum at T=1.3 and 1.7 K.* In the existence region of the trigonal structure the NMR spectrum splits into three branches, which agrees qualitatively with the form of the spectrum assumed above (see Fig. 3). In fields below 45 kOe, the experimental points indicate that all three branches are severely distorted by the dynamical interaction with the low-frequency AFMR mode. At fields above 50 kOe, where the dynamical shift can be ignored, the upper and lower branches of the spectrum convincingly describe the collapse of the trigonal structure, which is completed, according to other authors, at a field of $H_c=64$ kOe.

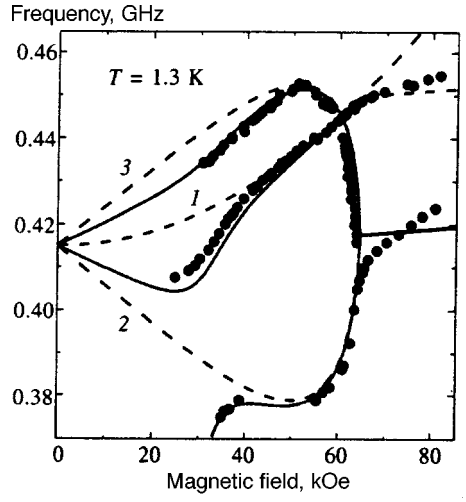


FIG. 7. NMR spectrum of $CsMnBr_3$ (high frequencies): points—experimental NMR spectrum at $T=1.3$ K, solid lines—result of calculation using Eq. (11), dashed lines—unbiased NMR spectrum (6) and (12).

The behavior of the middle branch clearly does not agree with the expression for ω_{n10} (6), represented by curve *1′* in Fig. 3. In order to describe the behavior of this branch it is necessary to assume that H_{hf} for ions of the sublattices perpendicular to the applied field depends on the magnitude of this field. Best agreement with experiment is obtained if this dependence is represented in the form

$$H_{hf}^1=H_{hf}(1+cH^2), \tag{12}$$

where $c=1.9\times10^{-5}$ kOe^{-2}. In Figs. 3 and 7 this empirical dependence is represented by curve *1*. It agrees with experiment only up to $H=H_c$. We return to a discussion of possible reasons for this anomaly below.

Below 45 kOe the NMR spectrum is distorted by strong dynamical interaction. Our experimental data for all three branches of the spectrum are described very well by the theoretical curves calculated from Eqs. (11) and depicted in Figs. 6 and 7 by the solid curves. Equation (11) contains two parameters ω_{n0} and

$$\omega_{T0}=\gamma_e\,\frac{\omega_{n0}}{\gamma_n}\,\sqrt{\frac{\chi_n}{\chi_\parallel}}. \tag{13}$$

In this formula χ_n is easily calculated by using the Curie law for nuclear moments. Using the value $\chi_\parallel=(1.2\pm0.1)\times10^{-2}$ cgs units/mole from Ref. 4, we see that just one unknown, ω_{n0}, remains in the formula for ω_{T0}. This single parameter was determined from the best fit of the theoretical curves for all three branches of the spectrum (11) to the experimental results. As a result, we obtained $\omega_{n0}=416\pm4$ MHz at $T=1.3$ K; this frequency corresponds to $H_{hf}=392\pm4$ kOe. Using the value of the hyperfine constant $A=-(1.53\pm0.04)\times10^{-18}$ erg obtained in Ref. 20 from the value of the hyperfine splitting of the EPR line of the Mn^{2+} ion introduced into $CsMgBr_3$, we determined the mean spin of the Mn^{2+}, $\langle S\rangle=|h\omega_{n0}/A|=1.80\pm0.05$. Uncertainty in the value of A, makes the major contribution to the error. From magnetization measurements[4] at $T=1.8$ K, $\langle S\rangle=1.7\pm0.1$, which agrees well with our data. From neutron scat-

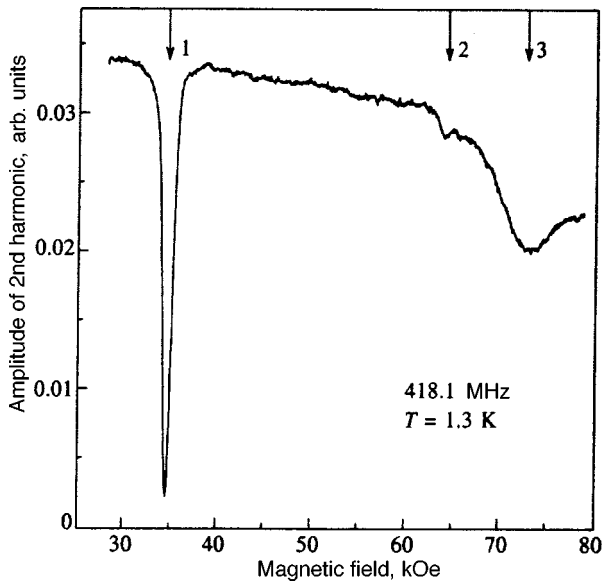


FIG. 8. Example of NMR spectrum at frequency of 418.1 MHz at $T = 1.3$ K. The numbers denote the centers of absorption lines: *1*—middle branch, *2*—upper branch, *3*—signal from collapsing sublattices.

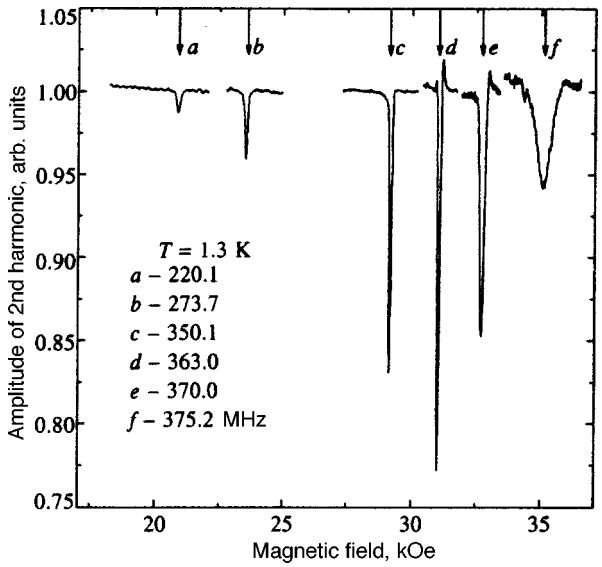


FIG. 9. Experimental low-frequency branch of NMR at $T = 1.3$ K at several frequencies.

tering work,[1] $\langle S \rangle = 1.5 \pm 0.15$ at $T = 4.2$ K; this also agrees satisfactorily with our value if the difference in temperatures is taken into account.

Using Eq. (13) and our value of ω_{n0}, we have the following relationship for the temperature dependence: $\omega_{T0} = (7.9 \pm 0.4)/\sqrt{T}$ GHz (temperature T in K). Note that the value of ω_{T0} calculated in this manner agrees satisfactorily with the experimental AFMR gap for $T = 1.2$ and $T = 2.0$ K.[11]

The lower branch 2 undergoes the greatest distortion due to dynamical coupling. The frequency of this branch in the 20–35 kOe interval of magnetic fields varies drastically (of the order of 10 MHz/kOe). The pulling effect is much less evident on the two upper branches, but it is still quite noticeable. This result agrees qualitatively with the results obtained in Refs. 18 and 19, in which collinear antiferromagnetics with two pairs of Mn^{2+} ions located at crystallographically nonequivalent positions were investigated.

The experimental points for $T = 1.7$ K for all branches of the spectrum are essentially identical to the data for $T = 1.3$ K. Only the lower branch at fields below 30 kOe (up to 0.4 kOe) is shifted slightly toward weaker fields; this is a consequence of the temperature dependence of ω_{T0}.

b) *Width and intensity of NMR lines.* The intensity of the absorption lines depends strongly on the magnetic field for all NMR branches.

The signal of the middle branch falls off abruptly with decreasing magnetic field, and at ~25 kOe it completely disappears; the same thing also occurs with the upper branch, but at fields below ~35 kOe. Such an abrupt falloff in NMR intensity was observed earlier for the upper NMR branch in $CsMnF_3$[18] and $CsMnCl_3$.[19] This can be explained by the fact that the nuclear magnetizations precess in antiphase in the different magnetic sublattices. It is quite possible that the same thing also occurs in $CsMnBr_3$.

Figure 8 shows an example of the spectrum at 418.1 MHz. The left line corresponds to the middle branch of the spectrum near maximum intensity. The middle line is the absorption signal of the upper branch near the phase transition. The right line is the signal from two pairs of collapsed sublattices. Note the abrupt increase in absorption intensity for this branch at a field $\sim H_c$.

The lower limit of observations of the low-frequency branch was 197 MHz (the low-frequency limit of the spectrometer); this corresponds to a resonant field of about 20 kOe ($T = 1.7$ K). Figure 9 shows a set of experimental scans of the low-frequency branch at different frequencies. All measurements were made at a fixed modulation amplitude ≈ 0.4 MHz and $Q \sim 400$. Figure 9 convincingly demonstrates the sharp increase in absorption (by more than a factor of 10) of this branch with frequency. Our analysis of the experimental data with $d\omega/dH$ taken into account shows a strong frequency dependence of the width of the resonance line: at about 210 MHz, the width of the absorption line is ~4 MHz, and at about 360 MHz it drops to approximately 0.5 MHz.

Note that no peculiarities in the line intensities were observed at the intersection point of the upper and middle branches.

7. ANOMALOUS NMR FEATURES IN $CsMnBr_3$

a) *Difference in hyperfine fields for different sublattices.* We have already pointed out that in sublattices oriented perpendicular to the applied field, the behavior of the hyperfine field differs from that in other sublattices, even in relatively weak fields. A similar effect in strong fields is even more surprising. According to the existing theory,[5] in fields greater than H_c $CsMnBr_3$ is transformed into the quasicollinear state, in which all six sublattices are perpendicular to the applied magnetic field, albeit tilted very slightly toward the latter (cosine of the tilt angle $\sim H/H_E$). Therefore, one might expect that the resonant frequency of the nuclei would be the same in all sublattices. Our results show that this is not the

case. We observed two branches of the NMR spectrum in fields above H_c. The lower branch lies in the frequency range in which one should expect it (near ω_{n10}), extending, as it were, the nonexistent branch *1′*, although a tendency toward an excess increase beyond H_c is also evident. The upper branch is located 30 MHz above *1′*. It extends curve *1* but with a much smaller slope. The presence of two branches means that above H_c, the product $A\langle S\rangle$ is 10% higher for ions in sublattices 1, 4 than in sublattices 2, 6 and 3, 5. Two possible reasons for a change in $A\langle S\rangle$ can be considered:

1. $\langle S\rangle$ can increase with the field because of a decrease in spin reduction, as shown elsewhere.[21] It is difficult to explain, however, why the effect of the field on spin reduction is different for different sublattice groups, even though they all lie in the same plane and are perpendicular to the applied field.

2. A change in A may be due to a change in the symmetry of the positions of the magnetic ions of sublattices 1, 4. In fields above H_c in each hexagonal layer the spins of one of these sublattices belong to 1/3 of the magnetic ions and are directed opposite the spins of 2/3 of the magnetic ions, which belong to the other two sublattice pairs. As a result, the spins of the first third of the ions are directed opposite the spins of all adjacent ions. For the other two thirds of the ions, the spin directions of adjacent ions alternate (and sum to zero). Moreover, each of the layers has nonzero magnetization. Thus, the hexagonal symmetry of each layer (and of the crystal as a whole) breaks down, and this should be accompanied by magnetostriction. However, at the present time it is difficult to assess the magnitude of this effect and its influence on the constant A. These symmetry considerations might in principle also explain the anomalous behavior of the frequency of the middle branch at fields less than H_c.

We also estimated the possible effect of a change in the dipole field at the nuclei. As numerical calculations showed, the dipole field at the nuclei is nearly unchanged during the collapse of the sublattices. Its total value amounts to about 1.7 kOe, whereas a field of the order of 30 kOe is required to explain the anomaly being considered.

b) *Temperature dependence.* We also found an extremely unusual change in the behavior of the lower branch of the NMR spectrum with increasing temperature. If, for a temperature increase from 1.3 to 1.7 K, the lower branch is shifted, albeit slightly, but in the ''required'' direction, the lower branch will accordingly be displaced toward weaker fields in accordance with the general prediction that the amount of pulling should decrease with increasing temperature. This effect was observed in all previously investigated collinear antiferromagnetics.

We investigated the behavior of the lower branch of the spectrum at 1.3, 1.7, 2.5, 3.0, 3.5 and 4.2 K. As an example, Fig. 10 shows the experimental data at three temperatures. It is seen that together with the reduction of the NMR frequency at 25 MHz, associated with a 6% decrease in the spontaneous magnetic moment of Mn^{2+}, the lower branch is shifted by about 3 kOe toward larger fields with a temperature increase from 1.7 to 3.5 K. Moreover, the intensity of the NMR signal decreased appreciably, and the resonance line was broadened.

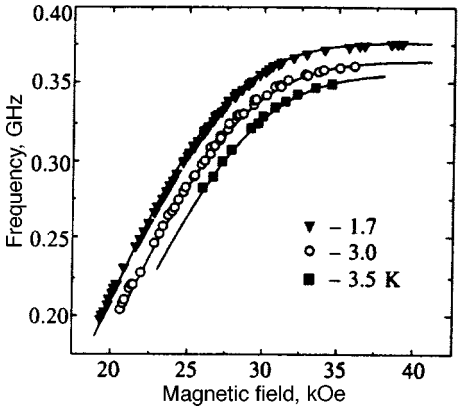


FIG. 10. Temperature dependence of lower NMR branch. Solid lines—result of calculation using Eq. (11) with empirical $F(T)$ function taken into account.

The results indicate that besides the free parameter ω_{n0} (which depends weakly on temperature) and the quantity ω_{T0} (the temperature dependence of which is rigorously defined and shifts the lower NMR branch in the opposite direction), only one quantity—the AFMR frequency ω_{e1}—which should depend on the temperature, remains in Eq. (11), from which the shape of the branches of the NMR spectrum was calculated. We have proposed the following temperature dependence for it:

$$\omega_{e1}(H,T)=\frac{\omega_{e1}(H)}{F(T)}, \tag{14}$$

where $F(T)$ is an empirical function. Then, looking for the best agreement with the experimental curves obtained for the six temperatures stated above using iterative methods, we obtained ω_{n0} and $F(T)$ for these temperatures. Figure 11 shows the function $F(T)$ for various temperatures from 4.2 to 1.3 K. As a result, we have

$$F(T)=1.0+0.1T^2.$$

The equation for $\omega_{e1}(H,T)$, of course, is only valid for fields in which pulling is strong. The temperature at which the NMR line is located at the minimum field is $T_{\min}\approx 1.8$ K. This is thoroughly confirmed by experiment.

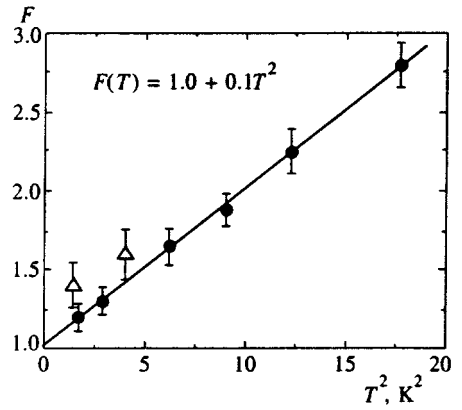


FIG. 11. Empirical $F(T)$ function. Dark points—experimental NMR data, open triangles—fit to AFMR data from Ref. 11.

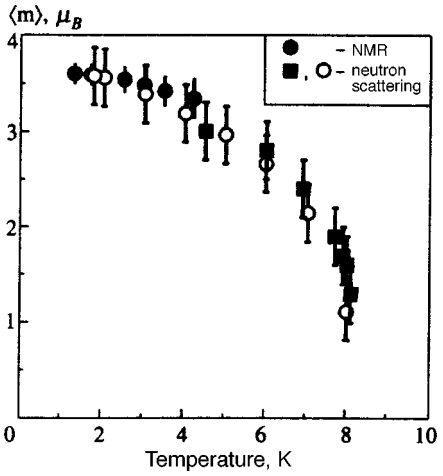


FIG. 12. Temperature dependence of the mean magnetic moment of Mn^{2+} in $CsMnBr_3$. Dark points—NMR; squares—from Ref. 1, open points—from Ref. 22.

Figure 12 shows the behavior of ω_{n0}, demonstrating the temperature dependence that we derived for the magnetization of antiferromagnetic sublattices. The data of two neutron diffraction investigations[1,22] are shown for comparison.

Absolute values are given in Ref. 1, but only relative values in Ref. 22. In the latter case we normalized them to our data at 1.7 K. All results agree to within the error limits.

8. CONCLUSION

In summarizing our work we can point to the following main results:

1. The NMR spectrum in an antiferromagnetic with a trigonal magnetic structure has been investigated for the first time. It has been shown that it is split into three branches and it graphically demonstrates the deformation process of the magnetic triangles with the transition to the quasicollinear structure in a field $H_c = 64$ kOe.

2. The experimental results have made it possible to determine the NMR frequency $\omega_{n0} = 416$ MHz for zero field, and to obtain by means of this value the mean spin of the Mn^{2+} magnetic ion, $\langle S \rangle = 1.80 \pm 0.05$. This means that spin reduction is 28% in quasi-one-dimensional $CsMnBr_3$ because of quantum fluctuations.

3. At fields below 45 kOe, all three branches of the spectrum experience strong frequency shifts (or pulling) due to the dynamical interaction of the electron and nuclear spin systems. The lower branch is displaced by a particularly large amount (nearly 200 MHz).

4. An equation has been obtained that describes the field dependence of all four modes of the spectrum (three quasinuclear and one quasielectronic), in very good agreement with experiment.

5. Despite expectations, in fields above H_c, where $CsMnBr_3$ in a first approximation should behave like a collinear two-sublattice antiferromagnetic with one NMR frequency, we found two NMR branches differing by 30 MHz. This can be explained by a disruption of hexagonal symmetry by the field.

6. It is possible that the unexpectedly large increase in resonant frequency with magnetic field at the nuclei of this pair of sublattices, which is perpendicular to the field, is related to the anomaly described in the previous paragraph.

7. The temperature dependence of the location of the low-frequency branch of the spectrum turned out to be more complicated than in three-dimensional magnetic materials. We were able to explain it only by assuming the existence of a rather strong temperature dependence of the frequency of the low-frequency AFMR branch (cubic in the field).

In conclusion, the authors warmly thank A. F. Andreev for assisting in the theoretical portion of this work, and M. I. Krukin, L. A. Prozorova, A. I. Smirnov, S. S. Sosin, and I. A. Fomin for numerous and very fruitful discussions of the results of this work.

This work was partially supported by the Russian Fund for Fundamental Research (Project 95-02-04569-a) and by the U. S. Civilian Research and Development Foundation Program for Independent States of the Former Soviet Union (CRDF, Grant No. RP1-207). A. M. Tikhonov is also grateful to Forschungszentrum Jülich GmbH.

*)Deceased.

[1] M. Eibischutz, R. C. Sherwood, F. S. L. Hsu, and D. E. Cox, AIP Conf. Proc. **17**, 864 (1972).
[2] B. D. Gaulin, T. E. Mason, M. F. Collins, and J. Z. Lavese, Phys. Rev. Lett. **B62**, 1380 (1989).
[3] B. Ya. Kotyuzhaskiĭ and D. V. Nikiforov, J. Phys.: Condens. Matter **3**, 385 (1991).
[4] S. I. Abarzhi, A. N. Bazhan, L. A. Prozorova, and I. A. Zaliznyak, J. Phys. Condens. Matter **4**, 3307 (1992).
[5] A. V. Chubukov, J. Phys. Condens. Matter **21**, 441 (1988).
[6] N. A. Zaliznyak, L. A. Prozorova, and S. V. Petrov, Zh. Eksp. Teor. Fiz. **97**, 359 (1990) [Sov. Phys. JETP **70**, 203 (1990)].
[7] A. S. Borovik-Romanov, S. V. Petrov, A. M. Tikhonov, and B. S. Dumesh, JETP Lett. **64**, 225 (1996).
[8] E. A. Turov and M. P. Petrov, *NMR in Ferro- and Antiferromagnetics* [in Russian], Nauka, Moscow (1969).
[9] A. J. Heeger, A. M. Portis, D. T. Teaney, and G. Will, Phys. Rev. Lett. **7**, 307 (1961).
[10] P. Pincus, P. G. de Gennes, F. Hartmann-Bourtron, and J. M. Winter, J. Appl. Phys. **34**, 1036 (1963); P. G. de Gennes, P. Pincus, F. Hartmann-Bourtron, and J. M. Winter, Phys. Rev. **129**, 1105 (1963).
[11] I. A. Zaliznyak, N. N. Zorin, and S. V. Petrov, JETP Lett. **64**, 473 (1996).
[12] J. Goodyear and D. J. Kennedy, Acta Crystallogr., Sect. B: **28**, 1640 (1974).
[13] B. D. Gaulin, M. F. Collins, and W. J. L. Buyers, J. Appl. Phys. **61**, 3409 (1987).
[14] B. S. Dumesh, Prib. Tekh. Eksp. **1**, 135 (1986).
[15] W. N. Hardy and L. D. Whitehead, Rev. Sci. Instrum. **52**, 213 (1981).
[16] M. E. Zhitomirsky, O. A. Petrenko, and L. A. Prozorova, Phys. Rev. B **52**, 3511 (1994).
[17] A. F. Andreev and V. I. Marchenko, Usp. Fiz. Nauk **130** (1980) [Sov. Phys. Usp. **23**, 21 (1980)].
[18] L. B. Welsh, Phys. Rev. **156**, 370 (1967).
[19] G. M. Gurevich, B. S. Ḍumesh, S. V. Topalov, A. V. Andrienko, and A. Yu. Yakubovskiĭ, Zh. Ėksp. Teor. Fiz. **84**, 832 (1983) [Sov. Phys. JETP **57**, 483 (1983)].
[20] G. L. McPherson, R. C. Koch, and G. D. Stucky, J. Chem. Phys. **60**, 1424 (1974).
[21] M. E. Zhitomirsky and I. A. Zaliznyak, Phys. Rev. B **53**, 3428 (1996).
[22] X. Xu, K. Okada, M. Fujii, N. Wada, M. Kurisu, and S. Kawano, J. Phys.: Condens. Matter **8**, L371 (1996).

Translated by Eugene R. Heath

Structure of the attraction zones of the final states in the presence of dynamical period doubling bifurcations

O. Ya. Butkovskiĭ*)

Vladimir State University, 600029 Vladimir, Russia

Yu. A. Kravtsov**)

Institute of Space Research, Russian Academy of Sciences, 117810 Moscow, Russia

E. D. Surovyatkina***)

Kamchatka State Marine Academy, 683003 Petropavlovsk-Kamchatskiĭ, Russia
(Submitted 15 May 1997)
Zh. Éksp. Teor. Fiz. **113**, 369–380 (January 1998)

The structure of the attraction zones of the final states associated with dynamical period doubling bifurcations is investigated. It is found that on the ''initial value—transition rate'' plane the attraction zones of the two possible final states alternate with each other and that a subdivision of the attraction regions occurs with a decrease in the transition rate. It is shown that the boundaries of the attraction zones are smeared out because of the effect of noise and in this situation the fine structure of the attraction zones is destroyed. As analytical and numerical calculations have shown, the critical value of the noise variance, corresponding to the boundary between the dynamical (or predictable) and stochastic (or unpredictable) modes, has a power-law dependence on the transition rate with a typical exponent value of one. The existence of ''noise'' invariants is also observed: the integrated (over all initial values) probability of achieving the final state is invariant with respect to the noise level. [S1063-7761(98)02501-3]

1. INTRODUCTION

Bifurcations in systems, whose parameters change at a finite rate, are customarily called dynamic. It is known that in the absence of noise the final state of a nonlinear system experiencing a dynamical bifurcation is completely determined by the initial conditions. This important feature of dynamical ''noise-free'' bifurcations was first pointed out by Shishkova,[1] and subsequently was investigated in detail in many papers, in particular in Refs. 2 and 3, which contain a voluminous bibliography.

The presence of even a small amount of noise radically alters the character of the dynamical bifurcation transition; for a high noise level or, what is equivalent, for a small rate of change of the parameters the transition loses its determinate (or dynamical) character and becomes stochastic. If, as a result of the bifurcation, two energetically equivalent states arise, as occurs in parametrons,[4,5] in systems experiencing period doubling bifurcation,[6] in polarization-unstable lasers,[7] and in a number of other systems, then with an increase in the noise level the system becomes unpredictable, and the probability of coming to one of the two final states approaches 50%. The very same situation also occurs with a decrease in the transition rate, i.e., in the quasistatistical (or adiabatic) limit.

The number of papers that have been devoted to the effect of noise on dynamical bifurcations is relatively small. In particular, it was shown in Refs. 2, 3 and 8 that a delay occurs in the loss of stability after passing through the bifurcation point due to the effect of noise. We have found[9,10] the boundary between the stochastic (or unpredictable) and dynamical (or predictable) modes.

This paper investigates another important aspect of the period-doubling type of dynamical bifurcations: the existence of fine structure of the attraction zones of the final states. The distinguishing feature of these zones is the alternation of the attraction intervals to the first and to the second states not only along the initial value axis but also along the axis representing s, the rate of variation of the control parameter r. The initial model (noisy nonstationary logistic mapping) is described in Sec. 2, and the structure of the attraction zones in Sec. 3.

Another goal of this paper is an analysis of the smearing of the boundaries and destruction of the fine structure of the attraction zones due to the effect of noise (Sec. 4). A secondary but important result of this analysis was a more exact definition of the boundary between the stochastic and dynamical modes (Secs. 5 and 6) compared with our previous papers.[9,10]

Finally, the existence of ''noise'' invariants, i.e., probability characteristics that are invariant with respect to the noise intensity, is established in Sec. 7.

2. INITIAL MODEL: NONSTATIONARY NOISY LOGISTIC MAPPING

The nonstationary noisy logistic mapping

$$x(n+1)=F(x(n))+f(n), \quad R(x(n))=r(n)[1-x(n)]. \quad (1)$$

 1063-7761/98/010206-07$15.00

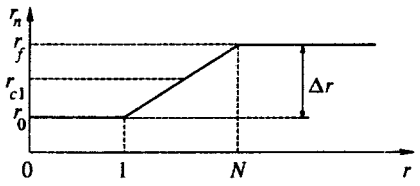


FIG. 1. Piecewise-linear variation of control parameter r with time: $r_n = r_1 + s(n-1)$.

is chosen as the object of our investigation. Here $F(x)$ is the noise-free logistic mapping and $r(n)$ is the control parameter, which varies from the initial value r_0 to the final value r_f in a piecewise-linear manner (Fig. 1)

$$r(n)=\begin{cases} r_0, & n=1, \\ r_0+s(n-1), & 1\leqslant n<N, \\ r_f, & n>N. \end{cases} \tag{2}$$

The discrete time n assumes values of 1, 2, ... The rate of change of the control parameter is denoted by $s=\Delta r/\Delta n$, and the number of steps through which the interval $\Delta r=r_f-r_0$ passes is denoted by N. The quantity N is 1 greater than the integer part of the fraction $\Delta r/s$:

$$N=\Delta r/s+1=(r_f-r_0)/s+1. \tag{3}$$

The initial $(n=1)$ value of the sequence $x(n)$ is denoted by ξ:

$$x_1\equiv\xi. \tag{4}$$

The first period-doubling bifurcation for the logistic mapping occurs at r values exceeding the critical value $r_{c1}=3$, and the second occurs at $r>r_{c2}=3.4$. Therefore, it is advisable to choose the initial value r_1 somewhat less than r_{c1}, and the final value r_f somewhat greater than r_{c1} but smaller than r_{c2}. In the calculations below we will work with the values of $r_1=2.8$ and $r_2=3.2$.

The bifurcation diagram of the logistic mapping for a variation of r within the interval from $r_1=2.8$ to $r_2=3.2$ is shown in Fig. 2a. The first period-doubling bifurcation occurs at the critical value $r=r_{c1}=3.0$. For $r>r_{c1}$ the initial branch $x^*(r)$ becomes unstable (the dotted line in Fig. 2a), and the the system enters one of the two possible stationary states $\bar{x}$ or $\underline{x}$, which are stable fixed double mapping points (Fig. 2b):

$$x(n+2)=F(F(x(n))). \tag{5}$$

The quantity x^* serves as the unstable point of this mapping. For $x>x^*$, the system arrives at the state $\bar{x}$ and for $x<x^*$ at the state $\underline{x}$. A determination of the attraction zones of the states $\bar{x}$ and $\underline{x}$ constitutes the subject of our investigation in this paper.

3. FINE STRUCTURE OF ATTRACTION ZONES IN ABSENCE OF NOISE

Figure 3 shows the ''initial coordinate ξ-rate s'' plane, on which the attraction regions of the state $\bar{x}$ are depicted as black and the attraction regions of the state $\underline{x}$ as white. This figure was obtained by a numerical solution of Eq. (1) in the absence of noise $(f(n)=0)$ for $r_0=2.8$ and $r_f=3.2$, so that $\Delta r=r_f-r_0=0.4$.

The black and white attraction zones alternate with each other in Fig. 3 along both the ξ axis and the s axis. These zones form a cellular structure, which is parametrized by the number of steps N through which the interval $\Delta r=r_f-r_0$ passes. This number is related to the rate s of variation of the control parameter r by Eq. (3). For $0.4<s<\infty$ the transition from r_0 to r_f is accomplished in one step, $N=1$. Initial values of ξ within the interval $0<\xi<4.3$ and $0.57<\xi<1.0$ lead to the state $\bar{x}$ (black regions), while values from the interval $0.43<\xi<0.57$ lead to the state $\underline{x}$ (white regions). Only for $s>0.4$ are there three intervals: two black and one white. For a two-step transition, $N=2$, when the rate s varies within the interval $0.2<s<0.4$, the number of intervals amounts to five. As the number of jumps N increases, the number of alternating black and white intervals increases as $2N+1$. For an N-jump transition the thickness Δs_N of the attraction intervals along the z axis decreases as

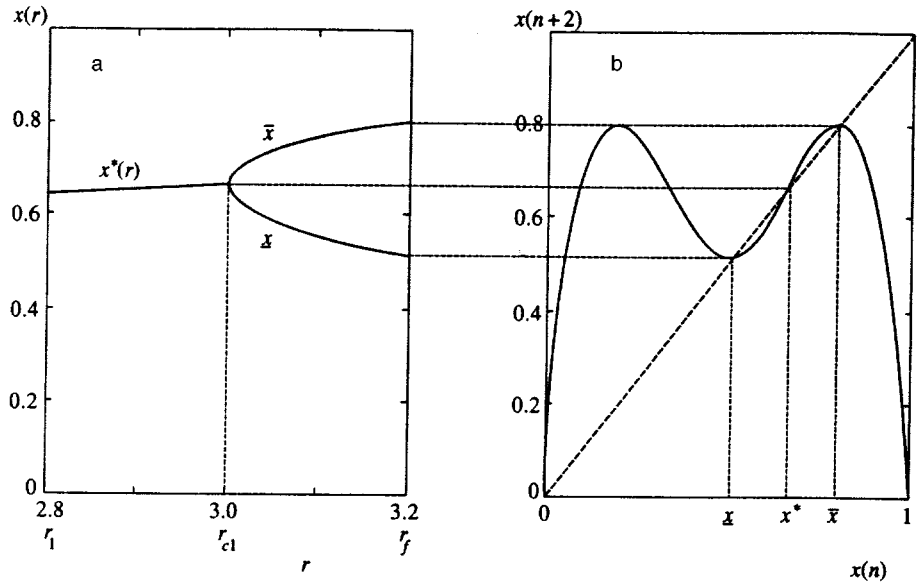


FIG. 2. Bifurcation dynamics of the logistic mapping in the interval $2.8<r<3.2$ (a) and double logistic mapping (b). The points $\bar{x}$ and $\underline{x}$ are stable final states of the system, x^* is unstable state.

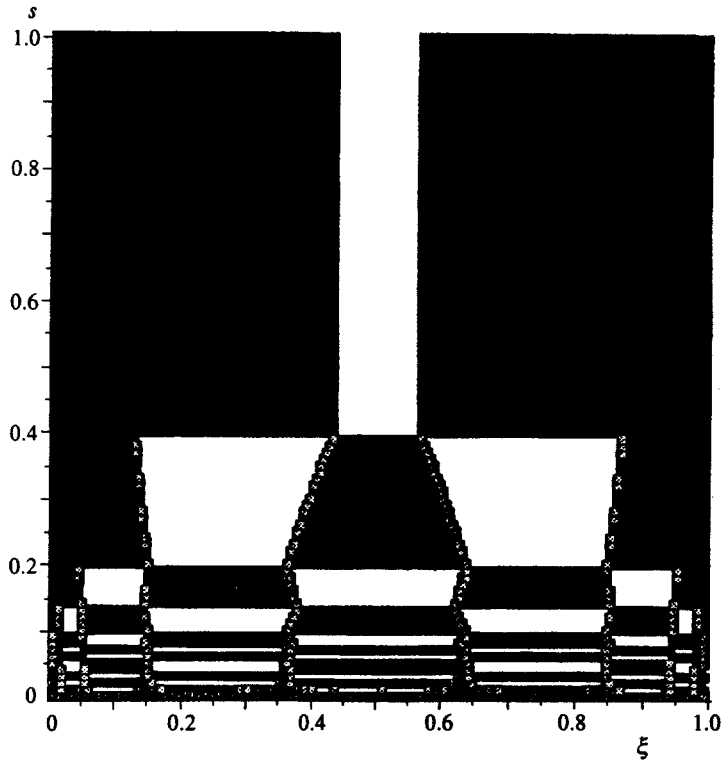


FIG. 3. Division of the ''initial coordinate ξ–rate s'' plane into attraction regions of the states $\bar{x}$ (black cells) and $\underline{x}$ (white cells). The regions subjected to action of noise are gray. These include the destroyed fine cells and smearing of cell boundaries.

$$\Delta s_N = s_N - s_{N+1} = \frac{\Delta r}{N} - \frac{\Delta r}{N+1} = \frac{\Delta r}{N(N+1)}. \tag{6}$$

As a result, the pattern of attraction zones acquires a fine structure with an increase in N, i.e., for a decrease in s. This structure is shown on a larger scale in Fig. 4 than in Fig. 3 for greater clarity.

Our partitioning of the (ξ,s) plane into zones of attraction to the states $\bar{x}$ and $\underline{x}$ represents a modification of the D partitioning in the theory of dynamical systems. In the problem being considered the quantity s characterizes the rate of variation of the control parameter in a nonstationary system, whereas D partitioning is usually done on the plane of the parameters characterizing a stationary state of the system. Generalization of the D partitioning method to a nonstationary system is perfectly natural. The key point in this method is the formal determination of the boundaries of the regions of a nonlinear system with qualitatively different behavior. The physical meaning of the partitioning parameters is important only in the interpretation step.

The unique feature of the D partitioning is that the cell dimensions decrease as the number of steps N increases. The newly formed black and white regions are narrower both in width and thickness than the previous regions and they appear in the lower portion and at the peripheries of Figs. 3 and 4. A reduction in the thickness and width of the regions creates the prerequisites for the destruction of the fine structure of the attraction zones due to the action of noise.

4. SMEARING OF THE ATTRACTION ZONE BOUNDARIES AND DESTRUCTION OF THEIR FINE STRUCTURE DUE TO THE EFFECT OF NOISE

We use very simple assumptions in the noisy logistic mapping (1) with respect to the noise $f(n)$: the fluctuations of $f(n)$ are uniformly distributed within the interval $(-a, +a)$:

$$w(f) = \begin{cases} 1/2a, & |f| \leqslant a, \\ 0, & |f| > a, \end{cases} \tag{7}$$

and the values of $f(n)$ and $f(m)$ at neighboring instants of time are uncorrelated:

$$\langle f(n) f(m) \rangle = \sigma_f^2 \delta(m,n), \quad \sigma_f^2 = a^2/3.$$

Here σ_f^2 is the variance of the fluctuations and $\delta(m,n)$ is the Kronecker symbol. Other assumptions about the distribution function $w(f)$ and the correlation function $(f(n)f(m))$ lead to qualitatively similar results.

Noise primarily smears out the sharp boundaries between the black and white attraction regions. We will arbitrarily depict the transition regions, for which the probabilities $\bar{P}$ and $\underline{P}$ of going into the states $\bar{x}$ and $\underline{x}$ lie within the interval (1/4, 3/4), by gray. The gray regions differ in probability from the purely black and purely white regions by less than 1/4. The interval $1/4 < P < 3/4$ is symmetrical with respect to the point $P = 1/2$, which corresponds to an equal probability of attaining states $\bar{x}$ and $\underline{x}$.

The gray zones in Fig. 3 are calculated for a noise level $\sigma_f^2 = 10^{-6}$. The probabilities $\bar{P}$ and $\underline{P}$ were calculated by averaging the data obtained from 500 realizations of the process $x(n)$ for more than 10,000 points on the (ξ,s) plane. A

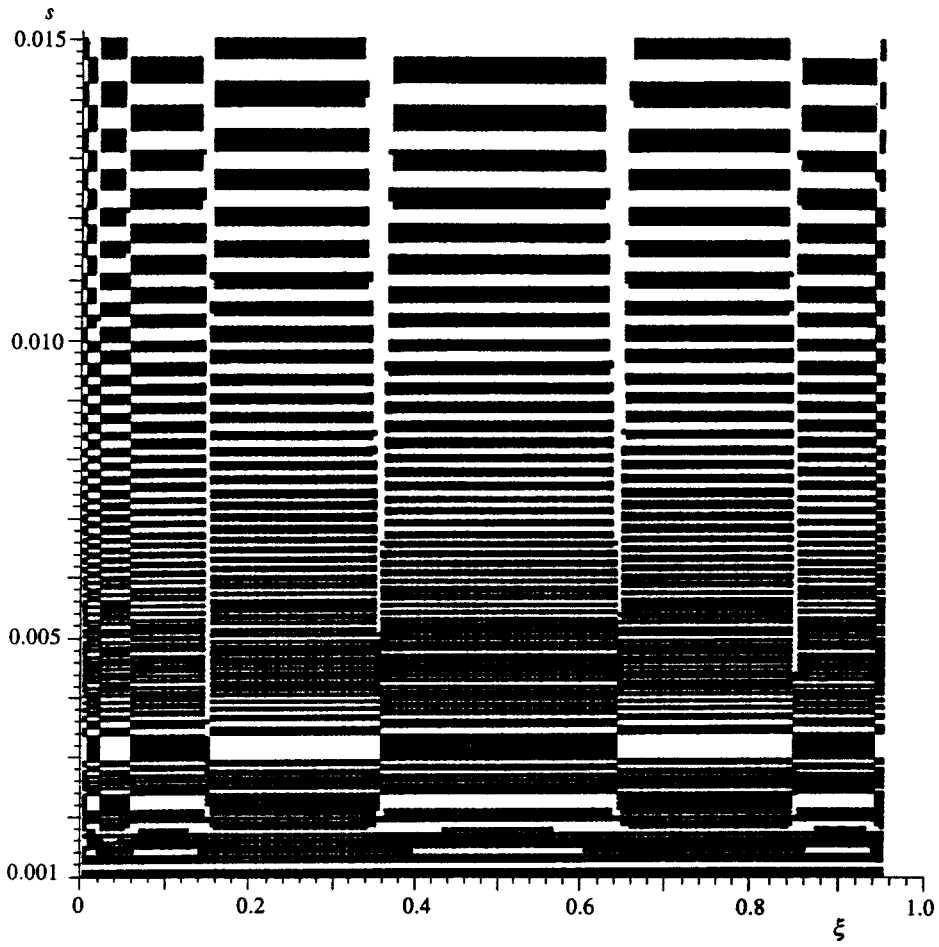


FIG. 4. Fine structure of attraction regions of the states $\overline{x}$ (black cells) and $\underline{x}$ (white cells) for small rates s.

black (or white) color is retained in this figure for the regions for which the probability $\overline{P}$ (or correspondingly $\underline{P}$) differ from unity by less than 1/4: $|1-\overline{P}|\leqslant 1/4$ for the black regions and $|1-\underline{P}|\leqslant 1/4$ for the white regions.

It follows from Fig. 3 that the noise action has completely destroyed the fine structure of the attraction regions at small rates ($s<0.001$ for $\sigma_f^2=10^{-6}$) and has noticeably smeared some boundaries between the white and black regions. The latter effect is only partially evident in Fig. 3 in view of the limited resolution capability of the chosen method of presenting the data.

5. CONDITIONS FOR SMEARING OF BOUNDARIES AND DESTRUCTION OF FINE STRUCTURE OF ATTRACTION ZONES

Due to the action of the external noise $f(n)$ the probability of going into the states $\overline{x}$ and $\underline{x}$ now becomes different from 1 and/or 0, so that the abrupt transition, let us say, from the region $\overline{P}=1$ to the region $\overline{P}=0$ now becomes smooth. Considering the smearing of the attraction zone boundaries due to the effect of noise, let us give the unperturbed (in the absence of noise) distribution of the probability $\overline{P^{(0)}}(\xi|N)$ of going into the state $\overline{x}$ from the initial state ξ after N steps. According to Eq. (1), the noise component shifts the variable $x(n+1)$ by $f(n)$ in each step. Assuming the noise is quite weak, i.e., ignoring nonlinear noise distortion effects, we relate the probability $\overline{P}(\xi|N)$ to the unperturbed distribution $\overline{P^{(0)}}(\xi|N)$ and to the noise distribution probability density $w(f)$. In the absence of nonlinear distortions of the noise component this relationship is given by a convolution transformation. For a single-step transition ($N=1$) we have the obvious relation

$$\overline{P}(\xi|1)=\int_0^1 \overline{P^{(0)}}(\xi'|1)w(\xi-\xi')d\xi', \tag{8}$$

which smooths the distribution $\overline{P^{(0)}}(\xi|1)$ with the weighting function w. For a two-step transition the noise acts twice on the variable x. In the first step the unperturbed probability $\overline{P^{(0)}}(\xi|2)$ is transformed into the intermediate distribution $\overline{P^{(1)}}(\xi|2)$ in accordance with a transformation of the type (8):

$$\overline{P^{(1)}}(\xi|2)=\int_0^1 \overline{P^{(0)}}(\xi'|2)w(\xi-\xi')d\xi', \tag{9}$$

and in the second step $\overline{P^{(1)}}(\xi|2)$ is transformed to

$$\overline{P^{(2)}}(\xi|2)\equiv\overline{P}(\xi|2)=\int_0^1 \overline{P^{(1)}}(\xi|2)w(\xi-\xi')d\xi'. \tag{10}$$

Similarly, for $\overline{P}(\xi|N)$ we have the series of convolutions

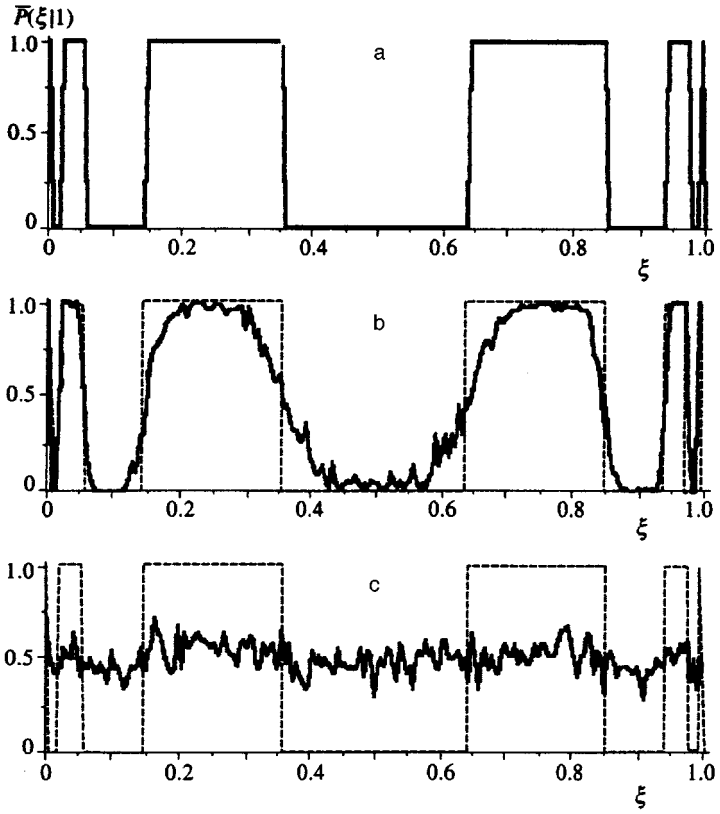


FIG. 5. Evolution of the probability $\bar{P}(\xi|1)$ of single-step transition to the state $\bar{x}$ as the noise level increases: a—unperturbed ''noise-free'' probability distributions ($\sigma_f=0$), b—smearing of boundaries in presence of weak noise ($\sigma_f \ll \sigma_c$), c—smoothing of probability distribution in presence of high noise level ($\sigma_f \gg \sigma_c$).

$$\bar{P}^{(1)}(\xi|N)=\int_0^1 \bar{P}^{(0)}(\xi'|N)w(\xi-\xi')d\xi',$$

.

.

.

$$\bar{P}^{(N)}(\xi|N)\equiv\bar{P}(\xi|N)=\int_0^1 \bar{P}^{(N-1)}(\xi'|N)w(\xi-\xi')d\xi'. \tag{11}$$

Based on these formulas, let us consider the smearing of the sharp boundary located at $\xi=\xi^*$. To do this we assume $\bar{P}^{(0)}(\xi|N)=\theta(\xi-\xi^*)$, where $\theta(\xi-\xi^*)$ is the unit Heaviside function. In view of Eq. (8), for a uniform distribution law (7) the probability $\bar{P}(\xi-\xi^*|1)$ for a one-step transition is given by the piecewise-linear function

$$\bar{P}(\xi-\xi^*|1)=\begin{cases} 0, & \xi<\xi^*-a, \\ (\xi-\xi^*+a)/2a, & \xi^*-a<\xi<\xi^*+a, \\ 1, & \xi>\xi^*+a. \end{cases} \tag{12}$$

According to Eq. (12) the width $\Delta\xi_1$ of the transition zone $(-a,+a)$ is comparable to the rms deviation: $\sigma_f=a/\sqrt{3}$. Each subsequent step broadens the transition zone. By using the arguments on which the derivation of the law of large numbers is based, one can prove that the total width of the transition zone $\Delta\xi_N$ for an N-step transition increases in proportion to $\sqrt{N}$:

$$\Delta\xi_N\approx\sigma_f\sqrt{N}. \tag{13}$$

Increasing in width, the gray transition zones gradually squeeze out the black and white regions and the probabilities of the two final states finally equalize:

$$\bar{P}(\xi-\xi^*|N)\approx\underline{P}(\xi-\xi^*|N)\approx 0.5.$$

This occurs when the width of the transition zone $\Delta\xi_N$ is comparable to the width of the attraction zone $\Delta\xi^*$:

$$\Delta\xi_N=\sigma_f\sqrt{N}\approx\Delta\xi^*. \tag{14}$$

Since $N\approx\Delta r/s$, we obtain the estimate

$$\sigma_c^2\approx(\Delta\xi^*)^2N\approx s(\Delta\xi^*)^2/\Delta r. \tag{15}$$

for the critical noise level σ_c^2, corresponding to equalization of the probabilities $\bar{P}$ and $\underline{P}$ and the destruction of the fine structure of the attraction zones. According to this estimate the critical noise level is proportional to the first power of the rate s: $\sigma_c^2\propto s$.

The process for the smearing of the boundaries and destruction of the fine structure of the attraction zones is illustrated in Fig. 5 by a numerical calculation example. The unperturbed (in the absence of noise) probability distribution $\bar{P}(\xi)$ (Fig. 5a), the probability distribution $\bar{P}(\xi)$ for $\sigma_f<\sigma_c$ corresponding to a slight smearing of the boundaries (Fig. 5b), and the probability distribution $\bar{P}(\xi)$ for $\sigma_f\gg\sigma_c$, which is close to 1/2 (Fig. 5c), are shown in this figure.

For very small rates s (i.e., for a very large number of steps N) the assumption of no nonlinear distortions ceases to be valid sooner or later, and the broadening will occur more rapidly than the power law (15).

6. BOUNDARY BETWEEN STOCHASTIC AND DYNAMICAL MODES OF BIFURCATION TRANSITION

In the purely dynamical mode when noise is absent, the probabilities of the system entering the state $\bar{x}$ or $\underline{x}$ are equal to 1: $\bar{P}(\xi)=1$ or $\underline{P}(\xi)=1$. In other words, for a dynamical transition the final state of the system is completely predictable. In the opposite case of severe noise the stochastic mode is realized, in which the final states are equally probable, $\bar{P}(\xi)=\underline{P}(\xi)=0.5$ and, consequently, unpredictable (more precisely, predictable with a probability of 0.5).

As the arbitrary boundary between the dynamical and stochastic modes we take the intermediate mode, for which the probabilities of going into states $\bar{x}$ and $\underline{x}$ amount to 75%: $\bar{P}(\xi)=\underline{P}(\xi)=0.75$. The boundary between the stochastic and dynamical modes chosen in this manner corresponds to the boundary of the gray regions in Fig. 3.

Let us examine the configuration of this boundary in greater detail. According to Eq. (15), at the middle of the (ξ,s) plane, i.e., for $\xi=1/2$, the critical noise level satisfies a power law of the form

$$\sigma_c^2=As^{\alpha}, \tag{16}$$

where $\alpha=1$ and $A\approx(\Delta\xi_0^*)^2/\Delta r$. For $\Delta\xi_0^*=0.3$ and $\Delta r=0.4$ we have $A\approx 0.25$. At the periphery of Fig. 3, i.e., as $\xi\to 0$ and $\xi\to 1$, the alternating black and white attraction zones narrow: their width $\Delta\xi^*$ decreases approximately by the factor $N=\Delta r/s$. Thus, as $\xi\to 0$ and $\xi\to 1$,

$$\Delta\xi^*\cong\frac{\Delta\xi_0^*}{N}\approx\frac{\Delta\xi_0^*}{\Delta r}\,s,$$

so that we again arrive at a power law of the form (16), but now with the exponent $\alpha=3$ and the coefficient $A_1=A/(\Delta r)^2$:

$$\sigma_c^2=A_1s^{\alpha}=\frac{(\Delta\xi_0)^2s^3}{(\Delta r)^3}. \tag{17}$$

The parameter values adopted above yield $A_1\approx 0.04$.

The overall character of the boundary between the stochastic and dynamical modes on the (ξ,s) plane is shown (for a fixed noise level σ_f^2) in Fig. 6. This boundary rises at the periphery compared with the central portion of the (ξ,s) plane since the fine structure of the attraction zones is destroyed more rapidly at the periphery.

The smooth curve *1* in Fig. 6 refers to the bulk of initial values of ξ and characterizes the lower boundary for the onset of the stochastic mode for most points ξ. Besides the boundary 1, corresponding to the destruction of the fine structure of the attraction zones, the stochastic mode arises at the boundaries of the wide attraction zones where even a small noise can make the final states equally probable. A high sensitivity to noise is typical of all boundaries between the black and white cells. Only the most smeared boundaries, corresponding to the gray regions in Fig. 3, are depicted by the curves *2* in Fig. 6. The rest of the boundaries of the attraction zones are subjected to a somewhat smaller smearing (the smearing of these boundaries is not shown in Fig. 6).

For the smeared boundaries the dependence of the critical noise level σ_c^2 on the rate s is characterized by higher

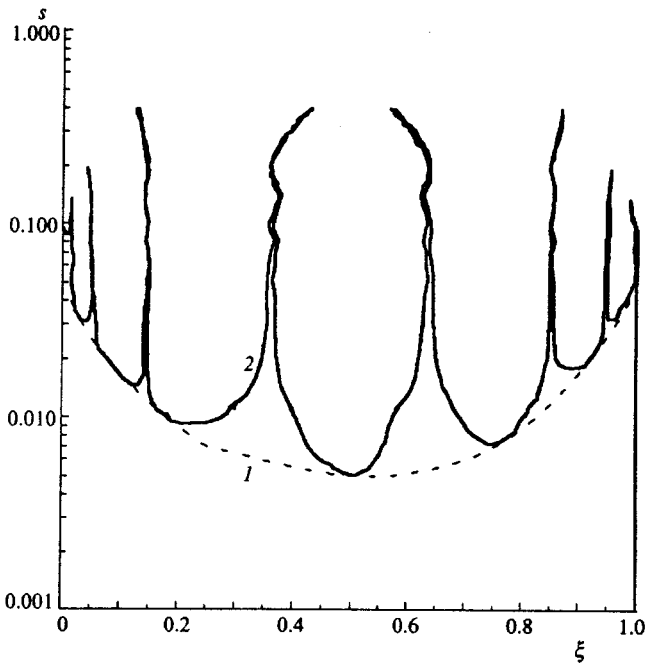


FIG. 6. Schematic representation of boundaries between dynamical and stochastic modes on (ξ,s) plane. The curve *1* corresponds to the destruction of the fine structure of the attraction zones, whereas the curve *2* refers to the smeared boundaries of the black and white cells.

exponents up to $\alpha\approx 5-7$, as was established previously[9,10] for the special initial value $\xi=0.67$, which coincides with the stationary value x^* (Fig. 2).

7. NOISE INVARIANTS

Let us consider the quantity

$$\bar{I}_N=\int_0^1\bar{P}(\xi)d\xi, \tag{18}$$

which represents the probability, integrated over all initial states ξ, of going into the final state $\bar{x}$, i.e., the area bounded by the curve $\bar{P}(\xi|N)$. The unique feature of this quantity is its independence (within the linearized model of the effect of noise) of the noise intensity σ_f^2, and it can therefore be called a noise invariant.

In order to prove the existence of noise invariants, for example, for $N=2$, we integrate Eq. (9) with respect to ξ with the normalization

$$\int_{-\infty}^{\infty}w(f\,)df=1.$$

taken into account. Then we obtain the relation

$$\bar{I}_N^{(0)}=\int_0^1\bar{P}(\xi|2)d\xi=\int_0^1\bar{P}^{(1)}(\xi|2)d\xi$$

$$=\int_0^1\bar{P}^{(0)}(\xi|2)d\xi=\bar{I}_2^{(0)},$$

which attests to the invariance of the quantity $\bar{I}_2$ with respect to the noise intensity. The invariance of the quantity $\bar{I}_N$ is proved similarly.

TABLE I. Values of noise invariants.

N	1	2	3	4	5
$\overline{I}_N$	0.86	0.4	0.64	0.45	0.54

Invariance of the quantity $\overline{I}_N$ means that noise neither increases nor decreases the area beneath the curve $P(\xi|N)$ but only redistributes the probability along the ξ axis. One interesting property of bifurcation transitions of the type being considered follows from this, which has yet to be mentioned in the literature as far as we know. Consider the quantity

$$\overline{I}_N^{(0)} = \int_0^1 \overline{P}^{(0)}(\xi|N)\,d\xi,$$

which characterizes the probability of going into the state $\overline{x}$ in the absence of noise. If this quantity is larger (or smaller) than 0.5, i.e., if the state $\overline{x}$ is reached more (or less) frequently than $\underline{x}$, then this property is also preserved in the presence of noise.

This property can have practical value at relatively small values of N when I_N is still markedly different from 0.5. For example, for $N \leqslant 5$ the difference of $\overline{I}_N$ from 0.5 amounts to at least 0.04 (or 8%), as follows from the table below.

The system considered is an example of a peculiar ''bifurcation'' roulette, in which noise cannot equalize the probabilities of the outcomes $\overline{x}$ and $\underline{x}$. Physically this is explained by the symmetrical action of the noise: the noise, facilitating an increase in transitions from the state $\overline{x}$ to $\underline{x}$, induces the same number of inverse transitions from the state $\underline{x}$ to $\overline{x}$. Of course, as $N \to \infty$, the noise invariants $\overline{I}_N$ and $\underline{I}_N$ approach 0.5, so that the integrated probabilities I_N and $\underline{I}_N$ do not equalize because of an increase in the noise but as a result of a decrease in the rate s (or increase in the number of steps N).

8. CONCLUSION

The above analysis of dynamical period doubling bifurcations has explained several previously unknown features.

First of all, it has been established that the attraction regions of two possible final states $\overline{x}$ and $\underline{x}$ alternate on the ''initial state ξ-transition rate s'' plane, forming fine structure as $s \to 0$.

Secondly, it has been shown that noise smears the boundaries of the attraction zones and thereby destroys their fine structure.

Thirdly and finally, the existence of noise invariants has been discovered, i.e., the integrated (over all initial values) probabilities of going into states $\overline{x}$ and $\underline{x}$ are independent of the noise intensity.

The authors are grateful to I. I. Minakova and M. V. Kapranov for critical comments and valuable advice. This work was performed with the partial financial support of the Ministry of General and Special Education of the Russian Federation under Grant 95-0-8.3-1 (GB 107/96).

*)E-mail: butr@vpti.vladimir.su
**)E-mail: kravtsov@asp.iki.rssi.ru
***)E-mail: elena@marine.kamchatka.su

[1] M. A. Shishkova, Dokl. Akad. Nauk SSSR **209**, 576 (1973).
[2] A. I. Neĭshtad and V. V. Sidorenko, Preprint No. 56 (Institute of Applied Mathematics, Russian Academy of Sciences, Moscow, 1995).
[3] C. Baesens, Physica D **53**, 319 (1991).
[4] E. Goto, Proc. IRE **47**, 1304 (1959).
[5] A. R. Kaplan, Yu. A. Kravtsov, and V. A. Rylov, *Parametric Oscillators and Frequency Dividers* (Sovetskoe Radio, Moscow, 1966).
[6] H. G. Schuster, *Deterministic Chaos* (Physik-Verlag, Weinheim, 1984; Mir, Moscow, 1988).
[7] E. Arimondo, D. Dangolane, C. Grabbanini, E. Menchi, and F. Papoff, J. Opt. Soc. Am. B **4**, 892 (1987).
[8] L. Kapral and P. Mandel, Phys. Rev. A **32**, 1076 (1985).
[9] O. Ya. Butkovskiĭ, J. S. Brush, and Yu. A. Kravtsov, in *Predictability of Complex Dynamical Systems*, ed. by Yu. A. Kravtsov and J. B. Kadtke (Springer Verlag, Heidelberg, 1996), p. 143.
[10] J. S. Brush, O. Ya. Butkovskiĭ, Yu. A. Kravtsov, and E. D. Surovyatkina, Zh. Éksp. Teor. Fiz. **109**, 2201 (1996) [JETP **82**, 1186 (1996)].

Translated by Eugene R. Heath

Nonlinear quantum theory of interaction of charged particles and monochromatic radiation in a medium

G. K. Avetisyan, A. Kh. Bagdasaryan, and G. F. Mkrtchyan

Erevan State University, 375049 Erevan, Republic of Armenia
(Submitted 20 February 1997)
Zh. Éksp. Teor. Fiz. **113**, 43–57 (January 1998)

We study the quantum theory of nonlinear interaction of charged particles and a given field of plane-transverse electromagnetic radiation in a medium. Using the exact solution of the generalized Lamé equation, we find the nonlinear solution of the Mathieu equation to which the relativistic quantum equation of particle motion in the field of a monochromatic wave in the medium reduces if one ignores the spin–spin interaction (the Klein–Gordon equation). We study the stability of solutions of the generalized Lamé equation and find a class of bounded solutions corresponding to the wave function of the particle. On the basis of this solution we establish that the particle states in a stimulated Cherenkov process form bands. Depending on the wave intensity and polarization, such a band structure describes both bound particle–wave states (capture) and states in the continuous spectrum. It is obvious that in a plasma there can be no such bands, since bound states of a particle with a transverse wave whose phase velocity $v_{\rm ph}$ is higher than c are impossible in this case. The method developed in the paper can be applied to a broad class of problems reducible to the solution of the Mathieu equation. © *1998 American Institute of Physics.* [S1063-7761(98)00401-6]

1. INTRODUCTION

While in spontaneous processes of emission of radiation by charged particles the medium acts as the third body needed for the conservation of energy and momentum, in processes where particles interact with an external field of electromagnetic radiation new phenomena that are essentially nonlinear manifest themselves.[1–5] For instance, in an insulator, where the refractive index $n=\sqrt{\varepsilon\mu}>1$, there is the Vavilov–Cherenkov effect (here ε and μ are the dielectric constant and permeability of the medium). In a plasma medium, where $n<1$, such processes are impossible (which is also true of a vacuum, where $n=1$), but in this case the plasma can act as the third body for energy and momentum conservation in the creation and annihilation of electron–positron pairs by the photon field.[6–8]

The interaction of electrons and electromagnetic radiation in a dielectric medium, i.e., a stimulated Cherenkov process, has its own special features due to the threshold nature of spontaneous Cherenkov radiation (the electron velocity v must be higher than the phase velocity c/n of the wave emitted in the given medium[1)]) or to the requirement that a certain coherence condition be met, $v\cos\theta=c/n$, where θ is the Cherenkov angle (c is the speed of light in vacuum). The existence of a threshold velocity in the spontaneous process leads to a threshold value of the field strength in the stimulated process, and, depending on whether the strength of the external field is greater or smaller than this critical value, the interaction of charged particles and the radiation field manifests itself differently (see Ref. 1).[2)] If we introduce a dimensionless Lorentz invariant parameter of the intensity of the given radiation field,

$$\xi^2=-\frac{e^2a_0^2}{m^2c^4},\quad a_0^2=a_{0i}a_0^i \tag{1.1}$$

(a_0 is the amplitude of the four-dimensional vector potential of the wave), then according to the results of Ref. 1, for $\xi>\xi_{\rm cr}$, where

$$\xi_{\rm cr}^2=\left(1-n\frac{v_0}{c}\right)^2\left[(n^2-1)\left(\frac{mc^2}{E_0}\right)^2\right]^{-1}, \tag{1.2}$$

the wave field becomes a potential barrier from which a particle is (inelastically) ''reflected.'' Here the expression for the critical-intensity parameter has been written for the case where the initial electron velocity $\mathbf{v}_0$ (electron energy E_0) is directed along the field's wave vector $\mathbf{k}$ ($|\mathbf{k}|=n\omega/c$), i.e., when the initial Cherenkov angle with the external wave is zero. If $\theta\neq0$ (wave polarization is unimportant in this case) and the interaction angles are not too small, $|\sin\theta|E_0/mc^2\gg\xi$ (see Ref. 5), we have

$$\xi_{\rm cr}^2(\theta)=\left(1-n\frac{v_0}{c}\cos\theta\right)^2\left[(n^2-1)\left(\frac{mc^2}{E_0}\right)^2\left|\sin\theta\right|\right]^{-1}. \tag{1.3}$$

In this case stimulated Cherenkov interaction is nonlinear in the field strength itself, and ''reflection'' occurs from the corresponding phase planes, while the electrons that initially were in the wave can be captured by it. This nonlinear ''reflection'' or capture of electrons by the radiation field has a straightforward physical interpretation in the reference frame associated with the wave (since $n>1$, we have $v_{\rm ph}<c$, and such a reference frame R does indeed exist[3)]). In this reference frame there can only be a (nonuniform) magnetic field which, beginning with a value $\xi>\xi_{\rm cr}$, reverses the particle's motion ($\mathbf{p}\to-\mathbf{p}$), i.e., the particle is elastically reflected.

This phenomenon, i.e., the existence a critical field strength, dramatically changes the behavior of electromagnetic processes in a medium. In particular, we can speak of the Compton effect in a dielectric medium only when the

field strength of the wave does not exceed this value.[4] This value determines the width of the Cherenkov resonance, i.e., the closeness to the Cherenkov cone, which in linear theory is limited only approximately by the condition for the applicability of perturbation-theory techniques. The exact solution of the classical problem shows[1,5] that there is a minimum width of the Cherenkov resonance determined by the strength of the field,

$$\left| v_0 \cos\theta - \frac{c}{n} \right|_{\min} = \Delta(\xi),$$

and that linear theory can be applied only if

$$\left| v_0 \cos\theta - \frac{c}{n} \right|_{\min} \gg \Delta(\xi),$$

or in other words, $\xi \ll \xi_{cr}$. But $\xi_{\rm cr}$, as Eqs. (1.2) and (1.3) show, can be arbitrarily small near the Cherenkov cone, which means that the Cherenkov process is highly nonlinear, no matter how weak the field is. In this connection it must be said that the results of studies in the stimulated Cherenkov effect obtained by perturbation techniques and reported in the reviews in Refs. 11 and 12 were erroneous. These results have been analyzed in detail in Ref. 13.

The nonlinear dynamics of the Cherenkov process and the features of this process mentioned earlier have been studied largely in classical terms. The solution of the quantum problem, which could, at least in principle, describe the quantum dynamics of the nonlinear interaction of an electron and an electromagnetic field in the medium, has been found only in the special case in which $\theta = 0$ and the wave is circularly polarized.[3,14] And, as noted earlier, when the field is turned on and off adiabatically, the electron can only be ''reflected'' from the wavefront (from the envelope of the wave pulse) due to the intensity effect ($\xi^2 > \xi_{\rm cr}^2$). As result of such (classical) ''reflection'' and because of the wave properties of the particle, there occurs an essentially quantum phenomenon: modulation of the electron probability density at hard x-ray frequencies due to the superposition of the incident and ''reflected'' electron waves.[3] It is still unclear what role the quantum phenomena play in the general case and how important the formation of bound electronic states in the capture mode is. Here we are forced to deal with a qualitative manifestation of the quantum nature of the particle–wave system, when the expected discrete spectrum of bound states of a particle are influenced in a quantum manner by the other potential wells, which are infinite in number (a situation resembling the pattern of electronic states in a solid).

In this paper we study the quantum theory of the interaction of a charged particle and coherent electromagnetic radiation in a medium. Our goal is to establish an overall picture of the nonlinear quantum dynamics of a stimulated Cherenkov process and to fill, in this way, the gaps mentioned earlier.

In Sec. 2 we discuss the nonlinear interaction of charged particles and the given field of a wave on the basis of the relativistic Klein–Gordon equation. We classify the cases in which such equations can be reduced to the Mathieu equation. Since the Mathieu equation cannot be solved exactly and the above phenomena cannot be studied by perturbation-theory techniques[15] or the eikonal approximation,[16] in Sec. 3 we develop a new approach: under certain conditions the Mathieu equation can be replaced by the generalized Lamé equation, which can be solved exactly. We study the class of bounded solutions. In Sec. 4 we study the stability of the wave functions and show that particle states form a system of allowed and forbidden bands (in the reference frame R associated with the wave this system corresponds to the band structure of the particle energy spectrum). Finally, in Sec. 5 we discuss the results.

2. KLEIN–GORDON EQUATION FOR A PARTICLE IN THE FIELD OF A PLANE MONOCHROMATIC WAVE IN A MEDIUM, REDUCIBLE TO THE MATHIEU EQUATION

If we ignore spin–spin interaction, the Klein–Gordon wave equation describes the interaction of a relativistic charged particle and a given field of electromagnetic radiation:

$$-\hbar^2 \frac{\partial^2 \Psi}{\partial t^2} = \left[c^2 \left(\hat{\mathbf{p}} - \frac{e}{c}\mathbf{A} \right)^2 + m^2 c^4 \right] \Psi, \tag{2.1}$$

where e and m are the particle's charge and mass, $\hat{\mathbf{p}} = -i\hbar\nabla$ is the particle's generalized momentum operator, and $\mathbf{A} = \mathbf{A}(t - nx/c)$ is the vector potential of a plane wave propagating along the x axis. We assume that the wave is monochromatic with frequency ω and polarization

$$\mathbf{A} = \left\{ 0,\ A_0 \sin\left[\omega\left(t - n\frac{x}{c} \right) \right],\quad gA_0 \cos\left[\omega\left(t - n\frac{x}{c} \right) \right] \right\}. \tag{2.2}$$

Here $g = 0$ corresponds to linear polarization and $g = \pm 1$ to right- and left-hand polarization, respectively.

In solving Eq. (2.1) it is convenient to replace the variables x and t by the wave coordinates $\tau = t - nx/c$ and $\eta = t + nx/c$. Then, as (2.2) shows, the variables η, y, and z are cyclic, so that the eigenvalues of the operators

$$\hat{\Lambda} = -i\hbar\frac{\partial}{\partial\eta},\quad \hat{p}_y = -i\hbar\frac{\partial}{\partial y},\quad \hat{p}_z = -i\hbar\frac{\partial}{\partial z}$$

are conserved: $\Lambda = \text{const}$, $p_y = \text{const}$, and $p_z = \text{const}$. Note that the two components p_y and p_z of the generalized momentum are conserved because the presence of a plane wave does not destroy the homogeneity of space in the yz plane, the polarization plane of the wave ($\mathbf{p}_\perp = \text{const}$). Clearly, in the free-particle solution of Eq. (2.1), Λ is the constant of the motion well known from classical electrodynamics, $E - cp_x/n = \text{const}$ ($\Lambda = (nE - cp_x)/2n$), only here, in the quantum case, $\mathbf{p} = \{p_x, \mathbf{p}_\perp\}$ and E must be interpreted as the initial momentum and total energy of a free particle.

Thus, we can look for a solution of Eq. (2.1) in the form

$$\Psi(\mathbf{r},t) = \Phi(\tau)\exp\left(-\frac{i}{\hbar}\Lambda\eta + \frac{i}{\hbar}\mathbf{p}_\perp\cdot\mathbf{r} \right), \tag{2.3}$$

$$\Lambda = \frac{1}{2}\left(E - \frac{c}{n}p_x \right) = \text{const},\quad \mathbf{p}_\perp = \text{const}, \tag{2.3a}$$

and for the function $\Phi(\tau)$ we have an ordinary second-order differential equation:

$$\hbar^2(n^2-1)\frac{d^2\Phi}{d\tau^2}+2i\hbar\Lambda(n^2+1)\frac{d\Phi}{d\tau}-\left[c^2\left(\mathbf{p}_\perp-\frac{e}{c}\mathbf{A}(\tau)\right)^2+\Lambda^2(n^2-1)+m^2c^4\right]\Phi=0. \tag{2.4}$$

This equation can be reduced to a one-dimensional Schrödinger equation with a ''potential energy'' $U(\tau)$ if we eliminate the term with the first derivative. To this end we conduct the following transformation:

$$\Phi(\tau)=F(\tau)\exp\left(-\frac{i}{\hbar}\frac{n^2+1}{n^2-1}\Lambda\tau\right). \tag{2.5}$$

Then for the unknown function $F(\tau)$ we have the following equation:

$$\frac{d^2F}{d\tau^2}+\frac{1}{\hbar^2(n^2-1)^2}\left\{4n^2\Lambda^2-(n^2-1)c^2\left(\mathbf{p}_\perp-\frac{e}{c}\mathbf{A}(\tau)\right)^2-(n^2-1)m^2c^4\right\}F=0, \tag{2.6}$$

and the solution of the Klein–Gordon equation (2.1) can be written

$$\Psi(\mathbf{r},t)=F(\tau)\exp\left(\frac{i}{\hbar}\mathbf{p}_\perp\cdot\mathbf{r}-\frac{i}{\hbar}\Lambda\eta-\frac{i}{\hbar}\frac{n^2+1}{n^2-1}\Lambda\tau\right), \tag{2.7}$$

with $F(\tau)$ the solution of Eq. (2.6).

Thus, establishing the interaction of a scalar particle and a given field of electromagnetic radiation (2.2) reduces to solving Eq. (2.6). The form of the latter depends on the polarization of the wave. In the general case of elliptic polarization this equation is the well-known Hill equation; for a circularly polarized wave Eq. (2.6) is the Mathieu equation, and for a linearly polarized wave the equation is again of the Hill type. Here we examine all cases of interaction (and the dependence on the wave intensity) in which the Klein–Gordon equation reduces to the Mathieu equation

$$\frac{d^2F}{d\alpha^2}=(B+D\sin^2\alpha)F. \tag{2.8}$$

In the case of a circularly polarized wave ($g=\pm1$ in (2.2)), there is azimuthal symmetry about the direction of wave propagation (the x axis), so that we can, without loss of generality, select the particle's initial momentum in the xz plane ($p_y=0$), i.e., $p_\perp=p_z$ in (2.6) and (2.7), and $\alpha=\omega\tau/2$. Then for the coefficients of Eq. (2.8) we have

$$B_c=\frac{4(n^2-1)m^2c^4+4(n^2-1)c^2p_z^2-16n^2\Lambda^2-8ecp_zA_0+4e^2A_0^2}{\hbar^2\omega^2(n^2-1)^2},$$

$$D_c=\frac{16ecp_zA_0}{\hbar^2\omega^2(n^2-1)}. \tag{2.9}$$

When the wave is linearly polarized, in the general case of arbitrary intensity and interaction angle Eq. (2.6) does not reduce to the Mathieu equation (2.8). But in realistic cases where the interaction angles are not too small, ξ is essentially always much smaller than $(E/mc^2)\sin\theta$, and in Eq. (2.6) we can ignore, with accuracy still remaining high, the term proportional to the wave intensity ($\propto A^2$) in comparison to the term proportional to $\mathbf{p}_\perp\cdot\mathbf{A}=pA\sin\theta$, where θ is the angle between the particle momentum $\mathbf{p}$ and the wave vector $\mathbf{k}$ of the given radiation field (2.2). Thus, if

$$\frac{E}{mc^2}\sin\theta\gg\xi \tag{2.10}$$

and the particles interact with a linearly polarized wave ($g=0$ in (2.2)), Eq. (2.6) reduces to the Mathieu equation (2.8) with

$$B_{l,\theta}=\frac{4(n^2-1)m^2c^4+4(n^2-1)c^2p_y^2-16n^2\Lambda^2-8ecp_yA_0}{\hbar^2\omega^2(n^2-1)^2},$$

$$D_{l,\theta}=\frac{16ecp_yA_0}{\hbar^2\omega^2(n^2-1)}. \tag{2.11}$$

If the conditions are opposite to those of (2.10), e.g., $\theta=0$, the interaction is due to the wave intensity, and for the coefficients of Eq. (2.8) we have

$$B_{l,0}=\frac{(n^2-1)m^2c^4-4n^2\Lambda^2}{\hbar^2\omega^2(n^2-1)^2},\quad D_{l,0}=\frac{e^2A_0^2}{\hbar^2\omega^2(n^2-1)}. \tag{2.12}$$

3. BUILDING AN APPROXIMATE SOLUTION OF THE MATHIEU EQUATION FROM THE EXACT SOLUTION OF THE LAMÉ EQUATION

As noted in Sec. 2, in the present paper we examine all cases in which the relativistic equations of particle motion are reducible to the Mathieu equation (2.8). As is known,[17] for $k\ll1$ Jacobi's elliptic function $\mathrm{Sn}(\alpha,k)$ tends to the ordinary sine function:

$$\mathrm{Sn}(\alpha,k)|_{k\ll1}\to\sin\alpha. \tag{3.1}$$

Using this property of the doubly periodic function $\mathrm{Sn}(\alpha,k)$, we can replace the Mathieu equation (2.8) by the generalized Lamé equation[17]

$$\frac{d^2U}{d\alpha^2}=[B+N(N+1)k^2\mathrm{Sn}^2(\alpha,k)]U, \tag{3.2}$$

where the coefficient $N(N+1)k^2$ of $\mathrm{Sn}^2(\alpha,k)$ is a constant quantity that depends on the parameters of the particle and the field (see Eqs. (2.9), (2.11), and (2.12)). For arbitrary values of B, Eq. (3.2) can be solved exactly for positive integers N.

Thus, passing to the limit (3.1), we can build an approximate solution of Eq. (2.8) by employing an exact solution of Eq. (3.2). Mathematically, when the passage to the limit (3.1) is completed, the divergence along the imaginary axis of the ordinary sine function at the infinitely distant point (in the analytic continuation into the complex plane) transforms into a removable divergence in the form of a simple pole in the elementary cell of the doubly periodic function Sn. In terms of the classification of algebraic equations, Eq. (2.8) has two regular points and one irregular point, while Eq. (3.2) has four regular points. When the Mathieu equation is replaced by a generalized Lamé equation, the irregular point of the Mathieu equation transforms into two regular points of the Lamé equation, with the result that the latter allows an exact solution (since Eq. (3.2) transforms into (2.8) in such a way that $N(N+1)k^2=D$ as $k\to 0$ and $N\to\infty$, in the Mathieu equation two regular points actually merge into one irregular).

Physically, the approximate solution of the Mathieu equation (2.8) found in this manner means that we allow for the poles of ''Bragg'' resonances in the field of the monochromatic wave (Cherenkov resonances in the reference frame comoving with the wave), and there are infinitely many such resonances. Hence such an essentially nonlinear solution will describe, at least in principle and to arbitrarily high accuracy, the nonlinear interaction of charged particles with a wave in the medium (no matter how weak the wave may be). This corresponds to $N=1$ in Eq. (3.2). The solution for strong fields corresponds to $N\gg 1$. This means, however, that any process becomes nonlinear and, on the other hand, that its quantum nature is suppressed and the interaction pattern approaches the classical one. Hence we study in detail the case in which $N=1$, which corresponds to extremely weak fields ($D\ll 1$), and which reveals the fundamental nonlinearity of the process and the quantum features of stimulated interaction noted in Sec. 1.

The exact solution of Eq. (3.2) in this case has the following form:[17]

$$U(\alpha)=C_1\frac{H(\alpha+\alpha_1)}{\theta(\alpha)}\exp[-Z(\alpha_1)\alpha]+C_2\frac{H(\alpha-\alpha_1)}{\theta(\alpha)}\exp[Z(\alpha_1)\alpha], \tag{3.3}$$

where α_1 is determined by the equation

$$\frac{\mathrm{Cn}^2\alpha_1\cdot\mathrm{dn}^2\alpha_1}{\mathrm{Sn}^2\alpha_1}-\frac{1}{\mathrm{Sn}^2\alpha_1}=B. \tag{3.4}$$

Here Cn α_1 and dn α_1 are also Jacobi's elliptic function (the cosinus amplitudinis and the delta amplitudinis), C_1 and C_2 are normalization constants, and the functions $H(\alpha)$, $\theta(\alpha)$, and $Z(\alpha)$ are defined as follows:

$$H(\alpha)=\vartheta_1(\alpha\vartheta_3^{-2}|\zeta),\quad \theta(\alpha)=\vartheta_4(\alpha\vartheta_3^{-2}|\zeta),$$

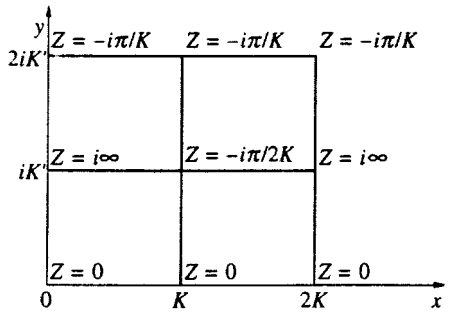


FIG. 1. Nodal values of the function $Z(\alpha)$ in the elementary cell of the complex variable α.

$$Z(\alpha)=\frac{\theta'(\alpha)}{\theta(\alpha)}, \tag{3.5}$$

where

$$\begin{aligned}\vartheta_1(\beta,q)&=2q^{1/4}\sin\beta-2q^{9/4}\sin 3\beta+2q^{25/4}\\&\quad\times\sin 5\beta-\cdots,\\ \vartheta_2(\beta,q)&=2q^{1/4}\cos\beta+2q^{9/4}\cos 3\beta+2q^{25/4}\\&\quad\times\cos 5\beta+\cdots,\\ \vartheta_3(\beta,q)&=1+2q\cos 2\beta+2q^4\cos 4\beta+2q^9\\&\quad\times\cos 6\beta+\cdots,\\ \vartheta_4(\beta,q)&=1-2q\cos 2\beta+2q^4\cos 4\beta-2q^9\\&\quad\times\cos 6\beta+\cdots.\end{aligned} \tag{3.6}$$

Obviously, $\vartheta_1(\beta,q)$ is an odd function and $\vartheta_{2,3,4}(\beta,q)$ are even functions. The arguments of the ϑ-functions incorporate $\vartheta_3\equiv\vartheta_3(0,q)$ ($\vartheta_i\equiv\vartheta_i(0,q)$, generally speaking), and $q=\exp(i\pi\zeta)$, where ζ is an arbitrary complex number with a positive imaginary part, so that $|q|<1$. On the other hand, the parameter q is related to the argument k of $\mathrm{Sn}(\alpha,k)$ in the following manner:

$$k=\frac{\vartheta_2^2(0,q)}{\vartheta_3^2(0,q)}. \tag{3.7}$$

Since the solution of Eq. (3.2) is the wave function of a particle and thus must be bounded, $Z(\alpha_1)$ in (3.3) must be either purely imaginary or zero, $Z(\alpha_1)=0$. The function $Z(\alpha)$ has the following properties:

$$Z(\alpha+2K)=Z(\alpha),\quad Z(\alpha+2iK')=Z(\alpha)-\frac{i\pi}{K}.$$

Hence, it is enough to study the function $Z(\alpha)$ in the rectangle $[2K;\ 2iK']$ of the complex plane (Fig. 1). Equations (3.5) imply that $Z(\alpha)$ is either purely imaginary or vanishes on the straight lines $x=0$ and $x=K$.

Now let us determine the range of values of B for a fixed value of K satisfying the condition of boundedness of the wave function (3.3). To this end we employ the relationship between α_1 and B given by Eq. (3.4). The latter yields

$$k^2\mathrm{Sn}^2\alpha_1=B+k^2+1, \tag{3.8}$$

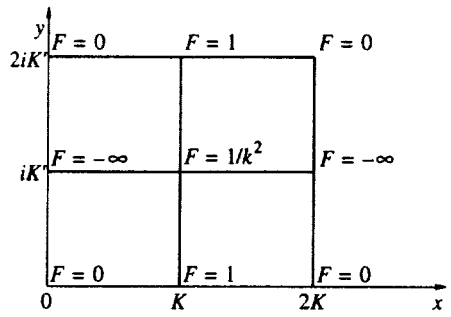


FIG. 2. Nodal values of the function $F(\alpha)=\mathrm{Sn}^2\ \alpha$ in the elementary cell of the complex variable α.

i.e., we must find all the values of B satisfying (3.8) for which α_1 varies along the straight lines $x=0$ and $x=K$. To this end we build the nodal values in the elementary cell $[2K;\ 2iK']$ of the complex variable α, since $\mathrm{Sn}^2(\alpha,k)$ is a doubly periodic function with real $(2K)$ and imaginary $(2K')$ periods (Fig. 2).

Figure 2 shows that on the straight lines $x=0$ and $x=K$ the function $\mathrm{Sn}^2(\alpha,k)$ varies within the half-closed region $(-\infty,0]$ and the interval $[1,1/k^2]$. From Eq. (3.8) in the $(-\infty,0]$ region we obtain

$$B+k^2+1\leqslant 0. \tag{3.9}$$

In the interval $[1,1/k^2]$ at $\alpha_1=K$ we obtain

$$B+1=0, \tag{3.10}$$

and at $\alpha_1=K+iK'$,

$$B+k^2=0, \tag{3.11}$$

i.e., in the interval $[1,1/k^2]$ the admissible values of the parameter B lie within the region

$$-1\leqslant B\leqslant -k^2. \tag{3.12}$$

Thus, the bounded solutions of the generalized Lamé equation (3.2) lie in the ranges (3.9) and (3.12) of variation of parameter B, i.e., there are allowed and forbidden bands. Let us show that at the edges of these bands the two linearly independent solutions in (3.3) coincide and are transformed, respectively, into the doubly periodic functions $\mathrm{Sn}(\alpha,k)$, $\mathrm{Cn}(\alpha_1)$, and $\mathrm{dn}\ \alpha_1$. The latter agrees with the general property that at the edges of the allowed bands, a particle's wave function must be periodic.

Thus, at $\alpha_1=0$ the solution (3.3) yields

$$U(\alpha)=C\frac{H(\alpha)}{\theta(\alpha)}=C\frac{\vartheta_2}{\vartheta_3}\ \mathrm{Sn}\ \alpha(C=C_1+C_2).$$

At $\alpha_1=K$ we have

$$U(\alpha)=C_1\frac{H(\alpha+K)}{\theta(\alpha)}+C_2\frac{H(\alpha-K)}{\theta(\alpha)}=C\frac{\vartheta_2}{\vartheta_4}\ \mathrm{Cn}\ \alpha.$$

At $\alpha_1=K+iK'$ we have

$$U(\alpha)=C_1\frac{H(\alpha+K+iK')}{\theta(\alpha)}\exp\left(-i\frac{\pi}{2K}\alpha\right)$$
$$+C_2\frac{H(\alpha-K-iK')}{\theta(\alpha)}\exp\left(i\frac{\pi}{2K}\alpha\right).$$

Allowing for the fact that $H(\alpha)$ is an odd function and

$$H(\alpha+K+iK')=q^{-1/4}\exp\left(i\frac{\pi}{2K}\alpha\right)\vartheta_3(\alpha\vartheta_3^{-2}),$$

we find that

$$U(\alpha)=C'q^{-1/4}\frac{\vartheta_3}{\vartheta_4}\ \mathrm{dn}\ \alpha(C'=C_1-C_2).$$

4. BAND STRUCTURE OF PARTICLE STATES IN A STIMULATED CHERENKOV PROCESS

The existence of well-defined ranges of B in Eq. (3.2), which at $k^2\ll 1$ becomes Eq. (2.8), restricts the region over which a charged particle can move in the field of a transverse electromagnetic wave in the medium. Indeed, the conditions (3.9) and (3.12) determine the admissible values of Λ (see Eq. (2.3a)) and the transverse components $p_\perp$ of the generalized momentum, i.e., the initial values of the energy and momentum of a free particle (the velocity and the Cherenkov angle) at which the particle remains in the wave. Thus, in the field of a plane-transverse wave in the medium there are allowed and forbidden regions of motion, or bands, for the charged particle. Below we will see that such a band structure of particle states exists only in a medium with refractive index $n>1$, due to the nonlinear Cherenkov interaction. Physically this means that a particle may be in a bound state with the transverse wave, i.e., is captured by the wave, which in any case is impossible in a plasma $(n<1)$, where the phase velocity of the wave, v_{ph}, is higher than c.

Let us now study the quantum dynamics of stimulated Cherenkov interaction. We wish to display the band structure of the particle state mentioned earlier as a function of the wave's intensity and polarization and of the interaction angle θ.

Suppose that we are dealing with a circularly polarized wave. If we combine the conditions (3.9), (3.12), and (2.9), we have (with such a wave) the following band structure for the conserved quantity $\Lambda=(nE-cp_x)/2n$:

$$nE-cp_x\geqslant\left[(n^2-1)(c^2p_z^2+m^2c^4+e^2A_0^2)+\frac{(n^2-1)^2\hbar^2\omega^2}{4}\right]^{1/2},$$

$$[(n^2-1)(c^2p_z^2+m^2c^4+c^2A_0^2)]^{1/2}\leqslant nE-cp_x\leqslant\left[(n^2-1)(c^2p_z^2+m^2c^4+e^2A_0^2-2eA_0cp_z)+\frac{(n^2-1)^2\hbar^2\omega^2}{4}\right]^{1/2}. \tag{4.1}$$

Such a band structure of the particle states has a simple physical explanation in the reference frame R moving together with the wave $(V=c/n)$. Since in such a reference frame there can only be a static magnetic field, the problem is steady-state, with the result that the particle energy is con-

served. Then the conditions (4.1) imply that the energy spectrum of the particle also has a band structure:

$$E' \geqslant \left[c^2 p_z^2 + m^2 c^4 + e^2 A_0^2 + \frac{n^2 \hbar^2 \omega'^2}{4} \right]^{1/2},$$

$$[c^2 p_z^2 + m^2 c^4 + e^2 A_0^2]^{1/2} \leqslant E' \leqslant \left[c^2 p_z^2 + m^2 c^4 - 2eA_0 c p_z + e^2 A_0^2 + \frac{n^2 \hbar^2 \omega'^2}{4} \right]^{1/2}. \quad (4.2)$$

Here $\omega' = (\omega/n)\sqrt{n^2-1}$ is the frequency of the wave in the reference frame R.

To establish the physical pattern of the band structure of the particle states in the laboratory reference frame, we express the conserved quantities Λ and $p_\perp$ in terms of the initial detuning of the Cherenkov resonance, $v\ \cos\theta - c/n$. Then the conditions (4.1) yield the following band structure for the initial longitudinal velocity of the particle:

$$v\ \cos\theta \leqslant \frac{c}{n} - \frac{c}{n}\left[(n^2-1)\left(\frac{mc^2}{E}\right)^2 \xi^2 + \left[\frac{\hbar\omega}{2E}(n^2-1)\right]^2 \right]^{1/2}, \quad (4.3a)$$

$$v\ \cos\theta \geqslant \frac{c}{n} + \frac{c}{n}\left[(n^2-1)\left(\frac{mc^2}{E}\right)^2 \xi^2 + \left[\frac{\hbar\omega}{2E}(n^2-1)\right]^2 \right]^{1/2};$$

$$v\ \cos\theta \geqslant \frac{c}{n} + \frac{c}{n}\left[(n^2-1)\left(\frac{mc^2}{E}\right)^2 \xi^2 \right]^{1/2}, \quad (4.3b)$$

$$v\ \cos\theta \leqslant \frac{c}{n} - \frac{c}{n}\left[(n^2-1)\left(\frac{mc^2}{E}\right)^2 \xi^2 \right]^{1/2};$$

$$v\ \cos\theta \leqslant \frac{c}{n} + \frac{c}{n}\left[(n^2-1)\left(\frac{nc^2}{E}\right)^2 \xi^2 - (n^2-1) \times \left(\frac{mc^2}{E}\right)^2 \frac{2p\ \sin\theta}{mc}\xi + \left[\frac{\hbar\omega}{2E}(n^2-1)\right]^2 \right]^{1/2}, \quad (4.3c)$$

$$v\ \cos\theta \geqslant \frac{c}{n} - \frac{c}{n}\left[(n^2-1)\left(\frac{nc^2}{E}\right)^2 \xi^2 - (n^2-1) \times \left(\frac{mc^2}{E}\right)^2 \frac{2p\ \sin\theta}{mc}\xi + \left[\frac{\hbar\omega}{2E}(n^2-1)\right]^2 \right]^{1/2}.$$

We see that due to the effect of the wave intensity in the stimulated Cherenkov process there is a forbidden band whose width depends neither on the interaction angle nor on quantum recoil (condition (4.3b)), and hence is of a classical nature and is always present, no matter how weak the radiation field. For this reason we call it the main forbidden band, with a width

$$\Delta_0(\xi^2) = 2\frac{c}{n}\left[(n^2-1)\left(\frac{mc^2}{E}\right)^2 \xi^2 \right]^{1/2}. \quad (4.4)$$

This constitutes the characteristic feature of a coherent spontaneous process, which leads to a substantial nonlinearity in the corresponding stimulated process (see footnote 3) discussed in Sec. 1. Because of the forbidden band (4.4) there is nonlinear ''reflection'' of the particle from the leading edge of the plane-transverse wave. Indeed, the condition (4.3b) shows that the edges of this band determine the critical value (1.2) of the wave intensity (at $\theta = 0$), above which the wave becomes a potential barrier for the particle and ''reflects'' it.[1,3] The other two forbidden bands are related to the stimulated Cherenkov interaction in the field, i.e., when the interaction angle θ is nonzero. In this case finite motion of the particle in the wave (capture) becomes possible, with the interaction of the charged particle and the transverse electromagnetic wave in the medium being entirely of a quantum nature. Here, because of the periodic structure of the field, the discrete levels of the bound states of the particle in the capture mode are transformed into bands, and structure similar to the band structure of electronic states in solids emerges. The forbidden bands, which appear at $\theta \neq 0$, are symmetrically disposed about the phase velocity of the wave, c/n, and have the same width $\Delta_1(\xi,\theta) = \Delta_2(\xi,\theta) \equiv \Delta(\xi,\theta)$:

$$\Delta(\xi,\theta) \approx \frac{c}{n}(n^2-1)\left(\frac{mc^2}{E}\right)^2 \frac{p}{mc}\xi\ \sin\theta \left[(n^2-1) \times \left(\frac{mc^2}{E}\right)^2 \xi^2 + \left[\frac{\hbar\omega}{2E}(n^2-1)\right]^2 \right]^{-1/2}. \quad (4.5)$$

The presence of such forbidden bands leads to ''reflection'' of the particle from the corresponding phase planes and to the possibility of formation of bound states in the wave, which are already first-order effects in the field strength; in the classical limit this agrees with our previous results.[1,5]

Note that formula (4.5) was written on the assumption that the parameter of the problem k^2 is much smaller than unity, which is a condition for the validity of the above results. This condition can also be written

$$\frac{8eA_0 c p\ \sin\theta}{\hbar^2\omega^2(n^2-1)} \ll 1. \quad (4.6)$$

If (4.6) holds, Eqs. (2.7) and (3.3) together with (2.9) determine the final form of the wave function of a particle with spin $S=0$ in the field of a plane-transverse monochromatic wave in a medium:

$$\Psi(\mathbf{r},t) = \exp\left[\frac{i}{\hbar}\mathbf{p}_\perp \cdot \mathbf{r} + \frac{i}{\hbar}\frac{bE - cp_x}{n^2-1}\left(\frac{x}{c} - nt\right) \right] \times \left\{ C_1 \frac{H((\omega/2)(t-nx/c)+\alpha_1)}{\theta((\omega/2)(t-nx/c))} \times \exp\left[-Z(\alpha_1)\ \frac{\omega}{2}\left(t - \frac{nx}{c}\right) \right] + C_2 \frac{H((\omega/2)(t-nx/c)-\alpha_1)}{\theta((\omega/2)(t-nx/c))} \times \exp\left[Z(\alpha_1)\ \frac{\omega}{2}\left(t - \frac{nx}{c}\right) \right] \right\}. \quad (4.7)$$

Let us now discuss the case in which the wave is linearly polarized. The conditions (3.9), (3.12), and (2.11) determine the following regions of particle motion in the wave as functions of the particle's initial velocity and the interaction angle (provided that condition (2.10) is met, i.e., that the effect of wave intensity is negligible):

$$v\cos\theta \leqslant \frac{c}{n}-\frac{c}{n}\frac{\hbar\omega}{2E}(n^2-1),$$
$$v\cos\theta \geqslant \frac{c}{n}+\frac{c}{n}\frac{\hbar\omega}{2E}(n^2-1); \tag{4.8a}$$

$$v\cos\theta \leqslant \frac{c}{n}+\frac{c}{n}\left\{\left[\frac{\hbar\omega}{2E}(n^2-1)\right]^2-(n^2-1)\times\left(\frac{mc^2}{E}\right)^2\frac{2p\sin\theta}{mc}\xi\right\}^{1/2},$$
$$v\cos\theta \geqslant \frac{c}{n}-\frac{c}{n}\left\{\left[\frac{\hbar\omega}{2E}(n^2-1)\right]^2-(n^2-1)\times\left(\frac{mc^2}{E}\right)^2\frac{2p\sin\theta}{mc}\xi\right\}^{1/2}. \tag{4.8b}$$

At $\theta=0$, or in conditions opposite to (2.10), when the interaction is due to the wave intensity, we have

$$v \leqslant \frac{c}{n}-\frac{c}{n}\left\{(n^2-1)\left(\frac{mc^2}{E}\right)^2\frac{\xi^2}{2}+\left[\frac{\hbar\omega}{E}(n^2-1)\right]^2\right\}^{1/2},$$
$$v \geqslant \frac{c}{n}+\frac{c}{n}\left\{(n^2-1)\left(\frac{mc^2}{E}\right)^2\frac{\xi^2}{2}+\left[\frac{\hbar\omega}{E}(n^2-1)\right]^2\right\}^{1/2}; \tag{4.9a}$$

$$v \geqslant \frac{c}{n}+\frac{c}{n}\left[(n^2-1)\left(\frac{mc^2}{E}\right)^2\frac{\xi^2}{2}\right]^{1/2},$$
$$v \leqslant \frac{c}{n}-\frac{c}{n}\left[(n^2-1)\left(\frac{mc^2}{E}\right)^2\frac{\xi^2}{2}\right]^{1/2}; \tag{4.9b}$$

$$v \leqslant \frac{c}{n}+\frac{c}{n}\frac{\hbar\omega}{E}(n^2-1), \quad v \geqslant \frac{c}{n}-\frac{c}{n}\frac{\hbar\omega}{E}(n^2-1). \tag{4.9c}$$

Here, if we examine the two opposite limits of the interaction angle or field strength, we see that the stimulated Cherenkov process directly depends both on wave intensity and on wave polarization. In contrast to the case of a circularly polarized wave, in the field of a linearly polarized wave the bound particle states (capture) also form due to the intensity effect at $\theta=0$. In this case the results are valid if

$$\frac{e^2A_0^2}{2(n^2-1)\hbar^2\omega^2} \ll 1. \tag{4.10}$$

The disappearance of the first forbidden band near the Cherenkov velocity at $\theta\neq 0$ can be related to the fact that the condition (2.10) ignores the wave intensity effect. As shown earlier, this main forbidden band for a charged particle in the field of a transverse wave in a refractive medium does not depend on the interaction angle, is always present no matter how weak the field intensity may be, and is of purely classical origin.

Up to this point we studied the main bands, which correspond to $N=1$ in the solution of the generalized Lamé equation. In the general case, for arbitrary N, the wave function of a scalar particle has the form

$$\Psi(\mathbf{r},t)=\exp\left[\frac{i}{\hbar}\mathbf{p}_\perp\cdot\mathbf{r}+\frac{i}{\hbar}\frac{nE-cp_x}{n^2-1}\left(\frac{x}{c}-nt\right)\right]$$
$$\times\left\{C_1\prod_{i=1}^{N}\frac{H((\omega/2)(t-nx/c)+\alpha_i)}{\theta((\omega/2)(t-nx/c))}\times\exp\left[-Z(\alpha_i)\frac{\omega}{2}\left(t-\frac{nx}{c}\right)\right]\right.$$
$$\left.+C_2\prod_{i=1}^{N}\frac{H((\omega/2)(t-nx/c)-\alpha_i)}{\theta((\omega/2)(t-nx/c))}\times\exp\left[Z(\alpha_i)\frac{\omega}{2}\left(t-\frac{nx}{c}\right)\right]\right\}, \tag{4.11}$$

where the α_i $(i=1,\ldots,N)$ can be determined from the following system of equations:[17]

$$\sum_{j=1,j\neq i}^{N}\frac{\mathrm{Sn}\,\alpha_j\mathrm{Cn}\,\alpha_j\mathrm{dn}\,\alpha_j+\mathrm{Sn}\,\alpha_i\mathrm{Cn}\,\alpha_i\mathrm{dn}\,\alpha_i}{\mathrm{Sn}^2\alpha_h-\mathrm{Sn}^2\alpha_i}=0,$$
$$\left[\sum_{i=1}^{N}\mathrm{Cn}\,\alpha_i\frac{\mathrm{dn}\,\alpha_i}{\mathrm{Sn}\,\alpha_i}\right]^2-\sum_{i=1}^{N}\frac{1}{\mathrm{Sn}^2\alpha_i}=B. \tag{4.12}$$

Equation (4.11) is valid if

$$\frac{16eA_0cp\sin\theta}{N(N+1)\hbar^2\omega^2(n^2-1)} \ll 1, \tag{4.13}$$

which shows that for $N\gg 1$ the wave function (4.11) is valid in the most highly refractive media possible, and in laser fields. This equation also shows that the discussed results for the main bands in the field of a circularly polarized wave are also valid in the case $N=1$ in strong fields at small interaction angles, $\theta\ll 1$.

5. DISCUSSION

Achieving a quantum description of the interaction of a charged particle and a plane-transverse electromagnetic wave in a medium, based on the relativistic equation of motion, can be reduced to finding the solution of the Mathieu equation (a one-dimensional problem in the wave coordinate). The method of building a nonlinear approximate solution of the Mathieu equation developed in the present paper is based on the exact solution of the generalized Lamé equation for small values of the modulus of Sn when the latter tends to the ordinary sine function (see (3.1)). Using this property of the doubly periodic function $\mathrm{Sn}(\alpha,k)$, we replace the Mathieu equation by the generalized Lamé equation, which has the same basic properties as the Mathieu equation, but in contrast to the latter allows for an exact solution. We have studied the class of bounded solutions of the Lamé equation and used them to build the particle wave function. According to the latter, the spectrum of eigenvalues of the quantities characterizing a state of a particle in the field of a transverse monochromatic wave in a dielectric medium has band structure. In the reference frame attached to the wave there can only be a static magnetic field, i.e., the problem is steady-state, with the result that energy is conserved (and so is the generalized momentum of the particle in the polarization plane of the plane-transverse wave), and the particle energy spectrum consists of bands. In the laboratory reference frame the conserved quantities are

$$\Lambda = \frac{1}{2}\left(E - \frac{c}{n}p_x\right) = \text{const}, \quad \mathbf{p}_\perp = \text{const}$$

(see Eqs. (2.3a)), a situation that leads to band structure in the initial longitudinal particle velocity. This velocity determines the admissible width of the Cherenkov resonance, or the proximity of the Cherenkov cone to an external wave in the stimulated Cherenkov process. This is a typically nonlinear resonance in forced oscillations, wherein the resonance width is a nonlinear function of the amplitude of the external periodic force—in our case the field strength in the wave. The admissible width of the nonlinear Cherenkov resonance is determined by the conditions (4.3) and (4.8), (4.9) for circular and linear polarization of the wave, respectively. For smaller values of this width, i.e., as we approach closer than a certain ''critical'' value from the initial Cherenkov cone (the limits of these conditions), a particle is unable to penetrate the wave further, i.e., there are forbidden regions for the motion of the particle in the field of the wave. Depending on the interaction angle, wave intensity, and polarization, the presence of these forbidden bands leads to ''reflection'' of the particle from the respective phase planes and to formation of bound states, or the capture of the particle by the transverse wave. The latter, in particular, means that there can be no band structure in plasma-like media, since bound states of a particle and a transverse wave propagating with phase velocity $v_{\rm ph} > c$ are impossible.

The nonlinear solutions of the Mathieu equation obtained in this paper on the basis of an exact solution of the Lamé equation are valid if the modulus of Sn is small, $k^2 \ll 1$, whereupon the generalized Lamé equation turns into the Mathieu equation. This is the only approximation used in building the nonlinear solution of the Mathieu equation on the basis of the method developed in this paper. When the wave is circularly polarized, the equation of particle motion has the form of the Mathieu equation, so that the results are valid if $k^2 \ll 1$, which amounts to (4.6) for the first band ($N=1$) and to (4.13) for a band with an arbitrary N. When the wave's polarization is linear, the equation of particle motion reduces to the Mathieu equation if we ignore the term proportional to wave intensity (condition (2.10)), so that our results hold in this case if the condition (4.6) is augmented by (2.10). However, in realistic cases for the Cherenkov process, the condition (2.10) (whereby the equation of motion reduces to the Mathieu equation) almost always holds, so we are again left with $k^2 \ll 1$ (condition (4.6)). At $\theta = 0$, instead of (4.6) we have (4.10).

The above equations for the first band ($N=1$) correspond to extremely weak fields. The essential nonlinearity of the stimulated Cherenkov process was established in precisely such fields. This nonlinearity is due to a buildup of quantum effects of tunneling origin or above-barrier reflection from an infinite number of shallow ''potential wells of the wave.'' In the field we consider in this paper (see (4.6)), the probabilities of such effects in each well are extremely low. But due to the coherent nature of the Cherenkov process (which has, at least in principle, an infinite coherent length) and the translational symmetry of the field of a monochromatic wave, there is a resonant increase in the probabilities of these effects, which leads to band stricture. The method developed in this paper allows for simultaneous contributions of infinite poles of the ''Bragg resonance'' type.

As for spin–spin interaction, the problem involving the Dirac equation will be discussed in a separate paper.

[1)] Actually, the spectrum of spontaneous Cherenkov radiation is determined by the dispersion of the medium, $n = n(\omega)$. But since in this paper we study the stimulated Cherenkov effect in an external field of monochromatic radiation, of which laser light is a good example, by n we mean the refractive index of the medium at the frequency of the stimulated wave.

[2)] The same coherence conditions as in Cherenkov process are applicable to stimulated Compton and undulator processes, so that these processes have all the features we have just described (resulting in an interaction with a retarded interference wave),[7,8] which can be generalized to include stimulated interaction of particles and a wave propagating with a phase velocity $v_{\rm ph} < c$.

[3)] In a plasma, where $n < 1$, an inertial reference frame moving with velocity $V = cn$ has physical meaning, being associated with the center of mass (C) of the particle–antiparticle system in pair production by the photon field in the plasma.[9,10]

[1] V. M. Arutyunyan and G. K. Avetisyan, Kvant. Elektron. (Moscow) No. 7, 54 (1972) [Sov. J. Quantum Electron. **2**, 39 (1972)].
[2] V. M. Arutyunyan and G. K. Avetisyan, Zh. Éksp. Teor. Fiz. **62**, 1639 (1972) [Sov. Phys. JETP **35**, 854 (1972)].
[3] V. M. Haroutunian and H. K. Avetissian, Phys. Lett. A **44**, 281 (1973).
[4] G. K. Avetisyan and S. G. Oganesyan, Izv. Akad. Nauk Arm. SSR, Fiz. **8**, 12 (1973).
[5] H. K. Avetissian, Phys. Lett. A **69**, 399 (1978).
[6] G. K. Avetisyan, A. K. Avetisyan, and Kh. V. Sedrakyan, Zh. Éksp. Teor. Fiz. **94**, No. 4, 21 (1988) [Sov. Phys. JETP **67**, 660 (1988)].
[7] V. M. Haroutunian and H. K. Avetissian, Phys. Lett. A **59**, 115 (1976).
[8] H. K. Avetissian, A. A. Jivanian, and R. G. Petrossian, Phys. Lett. A **66**, 161 (1978).
[9] G. K. Avetisyan, A. K. Avetisyan, and Kh. V. Sedrakyan, Zh. Éksp. Teor. Fiz. **99**, 50 (1991) [Sov. Phys. JETP **72**, 26 (1991)].
[10] K. Avetissian, A. K. Avetissian, A. Kh. Bagdasarian, and Kh. V. Sedrakian, Phys. Rev. D **54**, 5509 (1996).
[11] V. M. Arutyunyan and S. G. Oganesyan, Usp. Fiz. Nauk **164**, 1089 (1994) [Phys. Usp. **37**, 1005 (1994)].
[12] V. M. Harutunian and S. G. Oganesyan, Phys. Rep. **270**, 217 (1996).
[13] G. K. Avetisyan, Usp. Fiz. Nauk **167**, 793 (1997) [Phys. Usp. **40**, 755 (1997)].
[14] S. G. Oganesyan and G. K. Avetisyan, Izv. Akad. Nauk Arm. SSR, Fiz. **8**, 19 (1975).
[15] H. K. Avetissian, Phys. Lett. A **63**, 9 (1977).
[16] H. K. Avetissian, Phys. Lett. A **58**, 144 (1976).
[17] E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis*, Cambridge Univ. Press, Cambridge (1944).

Translated by Eugene Yankovsky

Effect of multiple scattering on the emission of ultrarelativistic electrons in a thin layer of matter

N. F. Shul'ga and S. P. Fomin

Kharkov Physicotechnical Institute National Scientific Center, 310108 Kharkov, Ukraine
(Submitted 27 May 1997)
Zh. Éksp. Teor. Fiz. **113**, 58–70 (January 1998)

We develop a quantitative theory of the effect of multiple scattering on the bremsstrahlung of ultrarelativistic electrons in a thin layer of matter. The effect is an analog of the Landau–Pomeranchuk–Migdal (LPM) effect of suppression of the radiation emitted by high-energy particles in an infinite amorphous medium, but certain differences do exist. On the basis of our approach we analyze the data recently obtained at SLAC (E-146) in experiments set up to verify the LPM effect. We show that in addition to the LPM effect, this experiment exhibited the suppression of bremsstrahlung in a thin layer of matter, theoretically predicted in our earlier papers. © *1998 American Institute of Physics.* [S1063-7761(98)00501-0]

1. INTRODUCTION

The process of emission of radiation by high-energy electrons develops in a broad spatial region in the direction of the particle's momentum. If within this region (which is known as the coherence length of the emission process[1,2]) an electron collides with a large number of atoms, the interaction with these atoms differs from the interaction with separate distant atoms.

Landau and Pomeranchuk[3] found that the multiple scattering of high-energy electrons in an amorphous medium leads to a decrease in the spectral radiation density in the low-frequency range in comparison to Bethe and Heitler's result.[4,5] Landau and Pomeranchuk[3] established the conditions needed for this effect to appear and estimated the spectral radiation density in the frequency range where the effect is significant. What is important is that the effect develops on a scale much smaller than the radiation length.

A quantitative theory of the effect of multiple scattering on electron bremsstrahlung in an amorphous medium was developed by Migdal,[6] who used the kinetic-equation method. The work done by Landau and Pomeranchuk[3] and Migdal[6] stimulated a great deal of research by other scientists, who studied the effect of the medium on bremsstrahlung at high energies. Among the areas of interest were the effect of polarization of the medium on radiation,[7] the allowance for recoil in emission,[8] the effect of absorption of radiation by the medium,[9] the spectral–angular distributions of radiation,[10,11] and the emission of radiation in finite-thickness targets[12] (see the reviews in Refs. 1, 2, 13–16 and the literature cited therein). Laskin *et al.*[17] proposed and developed a method of functional integration to describe the effect of multiple scattering on the radiation emitted by high-energy electrons in matter. At present the effect of multiple scattering on electron bremsstrahlung in an amorphous medium is known as the Landau–Pomeranchuk–Migdal (LPM) effect.[18–21]

A detailed experimental study of the LPM effect has been done only recently at SLAC for electron energies up to 25 GeV (see Ref. 19). In earlier experiments the accuracy was low due to the difficulties involved in gathering the necessary experimental statistics (see Refs. 22 and 23 and the literature cited therein). Anthony *et al.*[19] studied the spectrum of radiation emitted by electrons in the range of photon energies extending from fractions of MeV to several hundred MeVs. The measurements were done with targets ranging from carbon to uranium. The target thickness varied from fractions of one percentage point of the radiation length to several percentage points. Comparison of the experimental results with the results of Monte Carlo calculations using the Migdal formulas demonstrated good agreement of theory and experiment for relatively thick targets. With a thin target, however, significant discrepancy between the experimental results and Migdal's predictions was discovered in a number of cases, especially for gold targets with a thickness of 0.7% of the radiation length in the gamma-photon region below 30 MeV.

The experiment of Anthony *et al.*[19] aroused significant interest and stimulated the development of new approaches in studying the LPM effect based on the eikonal approximation in investigating the emission process,[20,24] on applying to the present problem the results of the theory of Moliere[25] and Bethe[26] of multiple scattering of particles in matter,[27] and on further development of the functional integration method.[28] The interest in this problem is also related to the search for analogs of the LPM effect in QCD and in other areas of physics (see, e.g., Refs. 29 and 30).

In this paper we analyze the above-mentioned experimental data that do not agree with the predictions of the Migdal theory of the LPM effect. We show that deviations from the predictions of the Migdal theory observed in the experiment of Anthony *et al.*[19] are present when the coherence length of the emission process is large compared to target thickness ($l_c \gg L$), i.e., when the condition for applicability of the Migdal theory are not met. We considered the case $l_c \gg L$ earlier, in Refs. 31 and 32, where we found that the spectrum of the radiation emitted by electrons in a thin target may differ considerably from the spectrum specified by the Migdal formula or by the Bethe–Heitler formula. In Ref. 31 we also derived asymptotic formulas for the radiation spectrum that are valid for ultrahigh electron energies, when the effect is significant. (Similar asymptotic formulas

were also obtained by Ternovskiĭ[12] on the basis of the kinetic-equation method.) However, direct application of these asymptotic formulas in the analysis of the experimental data of Ref. 19 leads to absurd results, since at SLAC energies the asymptotic behavior used in Refs. 12 and 31 is not reached. Hence analysis of the experimental data[19] requires developing a quantitative theory of the process of emission of radiation by relativistic electrons traveling through a thin layer of amorphous matter, a theory that would lead to exact results. The present paper is devoted to this problem.

In Sec. 2 we discuss the main results of the Migdal theory of the LPM effect.

Section 3 develops a general theory of the process of emission of radiation by high-energy electrons traveling through a thin layer of matter. We show that if the coherence length of the emission process exceeds the target thickness, the entire target acts as a single object with which the electrons interact and emit radiation. Here the radiation spectrum is determined solely from the distribution of the particles over the angles through which they are scattered by the target. The distribution function in turn depends on the type of target (an amorphous medium or a crystal). We derive simple asymptotic formulas for the average value of the spectral radiation density, which make it possible to analyze qualitatively the emission process in a thin layer of amorphous matter.

In Sec. 4 we develop a quantitative theory of the effect of multiple scattering on the emission of radiation by ultrarelativistic electrons in a thin layer of amorphous matter. The theory is based on the exact expression for the distribution of particles over the scattering angles in matter obtained by Moliere[25] and Bethe.[26] The distribution function takes into account both multiple scattering of particles in an amorphous medium and single scattering. Generally it differs considerably from the Gaussian distribution of particles over the scattering angles commonly used in the theory of the LPM effect. Here we discuss various limits when emission strongly depends on the effects of single and multiple scattering of particles by the atoms of the medium. We obtain several terms of the asymptotic expansion of the expression for the average value of the radiation spectrum in powers of the parameters determining the effect of multiple scattering on the emission process.

Finally, in Sec. 5 we make a quantitative comparison of the result of our theory and the experimental data of Ref. 19. We show our results and the results of the experiment agree well. We point out that the asymptotic expressions for the radiation spectrum obtained earlier in Refs. 31 and 32 are insufficient for a quantitative description of the given experiment.

All the results of our research have been obtained by describing the emission process within the classical electrodynamics approach. Such a description is valid when the energy of the gamma photons is low in comparison to the energy of the emitting particle, which means we can ignore the quantum recoil effects in emission. The conditions of the experiment conducted by Anthony *et al.*[19] obey this requirement.

2. THE LANDAU–POMERANCHUK–MIGDAL EFFECT

The spectral density of the radiation emitted by an electron moving in an external field along a path $\mathbf{r}(t)$ is determined in classical electrodynamics by the formula[1,2]

$$\frac{d\varepsilon}{d\omega}=\frac{e^2}{4\pi^2}\int do|\mathbf{k}\times\mathbf{I}|^2, \tag{2.1}$$

where $\mathbf{k}$ and ω are the wave vector and the frequency of the emitted wave, $|\mathbf{k}|=\omega$, do is the solid-angle element in the direction of the radiation's propagation, and

$$\mathbf{I}=\int_{-\infty}^{\infty}dt\ \mathbf{v}(t)\exp[i(\omega t-\mathbf{k}\cdot\mathbf{r}(t))]. \tag{2.2}$$

Here we use a system of units in which the speed of light is unity.

In an amorphous medium a particle's path is random, with the result that Eq. (2.1) must be averaged over the various paths of the electron:

$$\left\langle\frac{d\varepsilon}{d\omega}\right\rangle=\frac{e^2}{4\pi^2}\int do\langle|\mathbf{k}\times\mathbf{I}|^2\rangle. \tag{2.3}$$

The emission of radiation by a relativistic electron travelling in matter is a process that develops in a large and rapidly growing (with energy) spatial region along the direction of the particle's motion.[1,2] The length of this region is called the coherence length of the emission process.[1,2,32] In classical electrodynamics the coherence length is defined as

$$l_c=2\gamma^2/\omega, \tag{2.4}$$

where γ is the electron Lorentz factor. On the basis of qualitative estimates, Landau and Pomeranchuk found[3] that if the average value of the square of the angle of the electron deflection due to multiple scattering over the coherence length is larger than the square of the characteristic angle of emission of radiation by a relativistic electron, $\vartheta_\gamma\sim\gamma^{-1}$, the bremsstrahlung of an electron propagating in matter is suppressed in comparison to the results of Bethe and Heitler's theory.[4]

The first quantitative results concerning this effect were obtained by Migdal.[6] His reasoning was based on a procedure (which he proposed) for averaging Eq. (2.1) that used a kinetic equation for the angular distribution function of particles in the medium. Migdal succeeded in carrying out the procedure for an infinite amorphous medium in the limit of small scattering angles (in this case multiple scattering can be interpreted as a Gaussian process). He also found that

$$\left\langle\frac{d\varepsilon}{d\omega}\right\rangle=\Phi_M(s)\left(\frac{d\varepsilon}{d\omega}\right)_0, \tag{2.5}$$

where $(d\varepsilon/d\omega)_0$ is the radiation spectrum without allowance for the effect of multiple scattering on emission of radiation,

$$\left(\frac{d\varepsilon}{d\omega}\right)_0=\frac{2e^2}{3\pi}\gamma^2qL, \tag{2.6}$$

$s=(1/4\gamma^2)\sqrt{\omega/q}$ (q is the average value of the multiple scattering angle per unit length), and $\Phi_M(s)$ is the Migdal function, which takes into account the effect of multiple scattering on emission of radiation:

$$\Phi_M(s)=24s^2\left\{\int_0^\infty dx\ e^{-2sx}\coth x\ \sin(2sx)-\frac{\pi}{4}\right\}. \tag{2.7}$$

This function exhibits simple asymptotic behavior for small and large values of s:

$$\Phi_M(s)=\begin{cases}1, & \text{if } s>1,\\ 6s, & \text{if } s\ll 1.\end{cases}$$

The quantity s^2 is the ratio of the square of the characteristic angle at which radiation is emitted by a relativistic electron, $\vartheta_\gamma^2\sim\gamma^{-2}$, to the average value of the square of the electron deflection due to multiple scattering over the coherence length, $\overline{\vartheta^2}=ql_c$.

When $s>1$ holds, Eq. (2.5) coincides, within logarithmic accuracy, with the corresponding results of Bethe and Heitler:

$$\left\langle\frac{d\varepsilon}{d\omega}\right\rangle_{\rm BH}=\frac{4}{3}\frac{L}{L_R}\left[1+\frac{1}{12}(\ln 183Z^{-1/3})^{-1}\right], \tag{2.8}$$

where L_R is the radiation length,

$$L_R^{-1}=\frac{4Z^2e^6n}{m^2}\ln(183Z^{-1/3}).$$

Here n is the atom number density in the medium, $Z|e|$ is the nuclear charge, and m is the electron mass.

But if $s\ll 1$ holds, then according to (2.5) there is a decrease in the spectral density of the radiation emitted by an electron moving in matter in comparison to the results of Bethe and Heitler:

$$\left\langle\frac{d\varepsilon}{d\omega}\right\rangle\ll\left\langle\frac{d\varepsilon}{d\omega}\right\rangle_{\rm BH}. \tag{2.9}$$

This effect became known as the Landau–Pomeranchuk–Migdal effect.[18–21,27]

Formula (2.5) is valid if the medium in which the emission of radiation occurs is infinite. More precisely, the thickness L of the target must be large compared to the length within which the emission process develops:

$$L\gg l_c. \tag{2.10}$$

Formula (2.5) was derived to within logarithmic corrections. The reason for such accuracy is that q is assumed to be a constant independent of the path L on which scattering occurs. On the other hand, for small paths the value of q is a logarithmic function of t (see Ref. 33):

$$\overline{\vartheta^2}=\frac{\varepsilon_s^2}{\varepsilon^2}\frac{t}{L_R}\left(1+0.038\ \ln\frac{t}{L_R}\right). \tag{2.11}$$

(See also Refs. 1, 2, and 13, which discuss other aspects of the problem related to the necessity of refining the values of multiple scattering angles important for the emission of radiation.)

Note that formula (2.5) can also be derived via the path integral method, which was done in Ref. 17. Here the averaging procedure can be carried out analytically via integration over random paths, since the scattering is assumed to be a Gaussian process and the functional that must be averaged is an exponential expression for relativistic particles.

3. EMISSION OF RADIATION IN A THIN LAYER OF MATTER

The above theory of the LPM effect is valid if condition (2.10) is met, i.e., if we can ignore the effect of the edges of the target on the emission of radiation. We now study the opposite case, where the emission process develops over lengths much greater than the target thickness ($l_c\gg L$). The effect of the polarization of the medium on the emission and the transition radiation are ignored, which is justified if $\omega\gg\gamma\omega_p$ holds, where ω_p is the plasma frequency.

In our case of $l_c\gg L$ it is convenient to write $\mathbf{I}$ in (2.1) in the form

$$\mathbf{I}=i\int_{-\infty}^{\infty}dt\ \exp[i(\omega t-\mathbf{k}\cdot\mathbf{r}(t))]\ \frac{d}{dt}\ \frac{\mathbf{v}(t)}{\omega-\mathbf{k}\cdot\mathbf{v}(t)}. \tag{3.1}$$

This expression was obtained from (2.2) via integration by parts. The exponential factor in (3.1) can be set to unity, since for $l_c\gg L$ the difference $\omega\Delta t-\mathbf{k}\cdot\mathbf{r}(\Delta t)$ is much smaller than unity ($\Delta t=L/v$ is the time interval during which the electron interacts with the plate). Integration with respect to t in (3.1) yields

$$\mathbf{I}\approx i\left(\frac{\mathbf{v}'}{\omega-\mathbf{k}\cdot\mathbf{v}'}-\frac{\mathbf{v}}{\omega-\mathbf{k}\cdot\mathbf{v}}\right), \tag{3.2}$$

where $\mathbf{v}$ and $\mathbf{v}'$ are the electron velocities before and after the scattering by the target. Plugging (3.2) into (3.1), we arrive at the desired spectral density of the radiation emitted by an electron in a thin target. Allowing for the fact that at high energies the characteristic angles of electron scattering in a thin target are small, we can easily integrate in (2.1) over the emission angles. This yields

$$\frac{d\varepsilon}{d\omega}=\frac{2e^2}{\pi}\left[\frac{2\xi^2+1}{\xi\sqrt{\xi^2+1}}\ln(\xi+\sqrt{\xi^2+1})-1\right], \tag{3.3}$$

where $\xi=\gamma\vartheta/2$, and $\boldsymbol{\vartheta}=(\mathbf{v}'-\mathbf{v})/v$ is the angle of electron scattering by the target ($\vartheta\ll 1$).

Note that formula (3.3) was derived without using a specific law of particle–target scattering. Hence it is valid for scattering of particles in both an amorphous medium and in a crystal, as well as for the emission of radiation by a particle in a given external field. The only conditions are that the emission process must develop over a length greater than the size of the region where the external field acts on the particle and that the angle of particle–target scattering must be small compared to unity. Here, however, the ratio of the characteristic angle of emission of radiation by a relativistic electron, $\vartheta_\gamma\sim\gamma^{-1}$, to the electron scattering angle ϑ can be arbitrary.

Formula (3.3) has a simple asymptotic behavior for small and large values of the parameter ξ:

$$\frac{d\varepsilon}{d\omega}=\frac{2e^2}{3\pi}\begin{cases}\gamma^2\vartheta^2, & \text{if } \gamma\vartheta\ll 1,\\ 3[\ln\ \gamma^2\vartheta^2-1], & \text{if } \gamma\vartheta\gg 1.\end{cases} \tag{3.4}$$

The term with -1 in the asymptotic expression for $\gamma\vartheta \gg 1$ is the first-order correction to the logarithmic term (the discarded terms are of order $(\gamma\vartheta)^{-2}$).

The scattering angles for different particles are different, with the result that (3.3) must be averaged over the distribution of the particles ejected by the target in the angles ϑ:

$$\left\langle \frac{d\varepsilon}{d\omega} \right\rangle = \int d^2\vartheta f(\boldsymbol{\vartheta}) \frac{d\varepsilon\{\vartheta\}}{d\omega}, \tag{3.5}$$

where $\boldsymbol{\vartheta}=(\vartheta_x,\vartheta_y)$ is the two-dimensional angle through which the particles are scattered by the target in the plane orthogonal to the initial electron velocity **v**. The distribution functions $f(\boldsymbol{\vartheta})$ are different for different targets (amorphous media, crystals, nonuniform external fields). This leads to various effects in the emission of radiation (bremsstrahlung in an amorphous medium, coherent radiation in a crystal, etc.). But despite this difference, there are general laws that govern the emission of radiation by a particle traveling in a thin target, laws due to the asymptotic behavior (3.4). Specifically, if in (3.4) we replace ϑ^2 by the average value of the square of the particle–target scattering angle, $\overline{\vartheta^2}$, we arrive at the following estimates for the average value of the spectral density of radiation:

$$\left\langle \frac{d\varepsilon}{d\omega} \right\rangle \approx \frac{2e^2}{3\pi} \begin{cases} \gamma^2\overline{\vartheta^2}, & \text{if} \quad \gamma^2\overline{\vartheta^2} \ll 1, \\ 3[\ln\ \gamma^2\overline{\vartheta^2} - 1], & \text{if} \quad \gamma^2\overline{\vartheta^2} \gg 1. \end{cases} \tag{3.6}$$

We see from (3.6) that if $\overline{\vartheta^2}$ increases with target thickness (e.g., in an amorphous medium $\overline{\vartheta^2} \sim L$), at $\gamma^2\vartheta^2 \sim 1$ the nature of emission of radiation by a particle traveling in matter changes. Here the linear dependence of the radiation spectrum on $\overline{\vartheta^2}$ (for an amorphous medium, on L) is replaced by the weaker logarithmic dependence. Such a change in the nature of emission is characteristic of both amorphous media and crystals. (In a crystal the $\overline{\vartheta^2}$ vs. L dependence is generally more complicated than in an amorphous medium.[2] There are conditions such that in a crystal $\overline{\vartheta^2} \sim L$ also holds with $\overline{\vartheta_c^2} \gg \overline{\vartheta^2}$; see Ref. 2.)

The condition $\gamma^2\overline{\vartheta^2} \sim 1$, at which the nature of the emission of radiation by a particle in a thin layer of matter changes, coincides with the condition for the LPM effect of suppression of emission of radiation in a thick target. At the same time, the formulas that describe the effect of multiple scattering in thin and thick layers of matter differ considerably. For instance, in the case of the LPM effect the radiation spectrum depends on the frequency of the emitted photon, while (3.6) are independent of ω. Only the conditions for applicability of (3.6) depend on ω.

4. EMISSION OF RADIATION IN A THIN LAYER OF AMORPHOUS MATTER

To find the average value of the spectral density of the radiation emitted by a relativistic electron in a thin layer of matter (Eq. (3.5)), we must know the scattering-angle distribution function for the particles. The scattering-angle distribution function for fast particles traveling in an amorphous medium is specified by the following kinetic equation:[25,26]

$$\frac{df(\boldsymbol{\vartheta})}{dt} = n\int d\sigma(\boldsymbol{\chi})[f(\boldsymbol{\vartheta}-\boldsymbol{\chi},t) - f(\boldsymbol{\vartheta},t)], \tag{4.1}$$

where $d\sigma(\boldsymbol{\chi})$ is the cross section of particles scattered by an individual atom through the angle $\boldsymbol{\chi}$. For a screened Coulomb potential,

$$d\sigma(\chi) = (2\pi)^{-1}\chi d\chi d\varphi\sigma(\chi),$$

where χ and φ are the polar and azimuthal scattering angles and we have introduced $\sigma(\chi)$ used in Ref. 26:

$$\sigma(\chi) = \frac{8\pi Z^2e^4}{(pv)^2}\frac{1}{(\chi^2+\chi_a^2)^2}.$$

Here p is the electron momentum, and $\chi_a = \hbar/pR$, with R the Thomas–Fermi radius of screening of the atomic potential.

The general solution of Eq. (4.1) satisfying the condition $f(\boldsymbol{\vartheta},0) = \delta(\boldsymbol{\vartheta})$, where $\delta(\boldsymbol{\vartheta})$ is a two-dimensional delta function, has the form

$$f(\boldsymbol{\vartheta},t) = \frac{1}{(2\pi)^2}\int d^2\eta\ \exp\Bigg\{ i\boldsymbol{\vartheta}\cdot\boldsymbol{\eta} - nt \times \int d\sigma(\boldsymbol{\chi})(1-e^{i\boldsymbol{\chi}\cdot\boldsymbol{\eta}}) \Bigg\}. \tag{4.2}$$

This formula for $f(\boldsymbol{\vartheta},t)$ describes both single and multiple scattering of a particle in a medium. In the limit $t \to 0$, i.e., when we are dealing only with single scattering of a particle in a medium, we can expand (4.2) in powers of t. In the first approximation we find that for $\vartheta \neq 0$,

$$f(\vartheta) = ntd\sigma(\vartheta). \tag{4.3}$$

Inserting this distribution function in (3.5), we find (after integrating over scattering angles) a radiation spectrum that coincides (if we ignore the recoil effect) with the corresponding result of Bethe and Heitler (Eq. (2.8)). Here the screening radius R of the Coulomb potential must be set to $R = 0.81R_BZ^{-1/3}$ (see Refs. 1 and 34), where $R_B = \hbar^2/me^2$ is the Bohr radius.

The expansion in powers of t in (4.2) is valid if the target is thin compared to the mean free path of a particle in the material, $t_0 \sim 1/n\sigma_t$, where σ_t is the total electron–atom scattering cross section (for a screened Coulomb potential we have $\sigma_t = 4\pi R^2(Ze^2/\hbar v)^2$).

In the limit $t \gg t_0$ the multiple scattering of a particle by different atoms of the medium becomes important. The angular distribution function of the particles in this case can be transformed into[26]

$$f(\vartheta) = \frac{1}{2\pi B\chi_c^2}\int_0^\infty \eta d\eta J_0\left(\eta\frac{\vartheta}{\chi_c\sqrt{B}}\right)\exp\left(\frac{\eta^2}{4B}\ln\frac{\eta^2}{4} - \frac{\eta^2}{4}\right), \tag{4.4}$$

where $\chi_c^2 = 4\pi ntZ^2e^4/(pv)^2$, and B can be found from the following equation:

$$B - \ln B = \ln\frac{\chi_c^2}{\chi_a^2} + 1 - 2C. \tag{4.5}$$

Here $C = 0.577$ is Euler's constant. Since for $t \gg t_0$ we have the inequality $\chi_c \gg \chi_a$, B is large compared to unity. Keep-

ing the first two terms in the expansion of the exponential in (4.4) in powers of B^{-1}, we can write the distribution function as

$$f(\vartheta)=\frac{1}{2\pi\overline{\vartheta^2}}\left\{2e^{-\vartheta^2/\overline{\vartheta^2}}+\frac{1}{B}f^{(1)}\left(\frac{\vartheta}{\sqrt{\overline{\vartheta^2}}}\right)\right\}, \tag{4.6}$$

where $\overline{\vartheta^2}=\chi_c^2 B$ is the mean square multiple scattering angle of a particle in matter, and

$$f^{(1)}\left(\frac{\vartheta}{\sqrt{\overline{\vartheta^2}}}\right)=\int_0^\infty \eta d\,\eta J_0\left(\eta\frac{\vartheta}{\sqrt{\overline{\vartheta^2}}}\right)\frac{\eta^2}{4}\ln\left(\frac{\eta^2}{4}\right)e^{-\eta^2/4}. \tag{4.7}$$

The first term on the right-hand side of Eq. (4.6) is a Gaussian distribution of particles over the angles, related to multiple scattering. The second term is the first-order correction to the Gaussian distribution for $\vartheta^2\leqslant\overline{\vartheta^2}$. However, for $\vartheta^2\gg\overline{\vartheta^2}$ this term becomes the leading one and determines the probability of single scattering. In this range of angles,

$$f(\vartheta)\approx\frac{4ntZ^2e^4}{(pv)^2\vartheta^4},\quad \vartheta^2\gg\overline{\vartheta^2}. \tag{4.8}$$

Plugging the distribution function (4.6) into (3.5), we arrive at an expression for the radiation spectrum that allows for the effect of multiple scattering on emission of radiation:

$$\left\langle\frac{d\varepsilon}{d\omega}\right\rangle=\frac{2e^2}{\pi}\left[F_0(a)+\frac{F_1(a)}{B}-1\right], \tag{4.9}$$

where

$$a^2=\gamma^2\overline{\vartheta^2},$$

$$F_0(a)=\frac{8}{a^2}\int_0^\infty d\xi\frac{2\xi^2+1}{\sqrt{\xi^2+1}}\ln(\xi+\sqrt{\xi^2+1})e^{-4\xi^2/a^2},$$

$$F_1(a)=\frac{4}{a^2}\int_0^\infty d\xi\frac{2\xi^2+1}{\sqrt{\xi^2+1}}\ln(\xi+\sqrt{\xi^2+1})f^{(1)}(2\xi/a).$$

Let us first examine two limiting cases of formula (4.9) corresponding to $a^2\ll 1$ and $a^2\gg 1$. In the case where $a^2\ll 1$ we introduce a ξ_0 such that $a\ll\xi_0\ll 1$. Then in the region where $\xi<\xi_0$ holds we use the appropriate asymptotic formula in (3.4) for the radiation spectrum, while in the region where $\xi>\xi_0$ we use the asymptotic formula (4.8) for the distribution function. Here

$$F_0(a)\approx 1+\frac{a^2}{3}+O(\xi_0^4),$$
$$F_1(a)\approx\frac{1}{3}a^2\left[2(C-1)-\ln a^2+\frac{13}{6}+O(\xi_0^2)\right]. \tag{4.10}$$

Plugging (4.10) into (4.9) and employing (4.5), we arrive at a formula that coincides with the result (2.8) of Bethe and Heitler. Thus, both for $t\ll t_0$, where multiple scattering can be ignored, and for $t\gg t_0$ but $\gamma^2\overline{\vartheta^2}\ll 1$, the spectrum of the radiation emitted by electrons traveling in a thin layer of matter is given by the Bethe–Heitler formula.

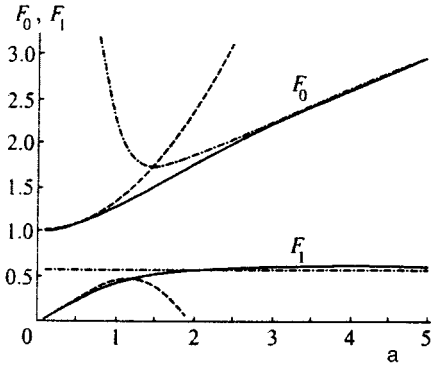


FIG. 1. The functions $F_0(a)$ and $F_1(a)$ obtained by numerically integrating the expressions (4.9) (solid curves), calculated by the asymptotic formulas (4.10) (dashed curves), and calculated by the asymptotic formulas (4.12) and (4.13) (dot–dash curves).

Now we examine the case $a^2\gg 1$. If we introduce a new variable $y=\xi+\sqrt{\xi^2+1}$ into the expression for $F_0(a)$, we get

$$F_0(a)=\frac{4}{a^2}\int_1^\infty\frac{dy}{y^3}(y^4+1)\ln y\,\exp\left[-\left(\frac{y^2-1}{ay}\right)^2\right]. \tag{4.11}$$

For $a^2\gg 1$ we can easily obtain the first terms in the series expansion of this integral in powers of a^{-2}:

$$F_0(a)\approx\left(\ln a^2-C+\frac{1}{a^2}+\cdots\right)\left(1+\frac{2}{a^2}\right)$$
$$+\frac{1}{a^2}+O(a^{-4}). \tag{4.12}$$

Reasoning along similar lines, we can obtain the first term in the series expansion of $F_1(a)$:

$$F_1(a)\approx C+O(a^{-2}). \tag{4.13}$$

We have kept only the first term of the expansion of F_1 in powers of a^{-2} since $F_1(a)$ enters into the radiation spectrum (4.9) with a small factor B^{-1}. In deriving the latter formula we used the result obtained by integrating (4.7), which can be represented in the following form:[26]

$$f^{(1)}(\vartheta)=2e^{-x}(x-1)\left[C+\int_0^x dt\,\frac{e^t-1}{t}\right]-2(1-2e^{-x}),$$

where $x=\gamma^2\vartheta^2/a^2$.

As a result we arrive at the following expression for the radiation spectrum with $a^2\gg 1$:

$$\left\langle\frac{d\varepsilon}{d\omega}\right\rangle=\frac{2e^2}{\pi}\left\{(\ln a^2-C)\left(1+\frac{2}{a^2}\right)+\frac{2}{a^2}+\frac{C}{B}-1\right\}. \tag{4.14}$$

We see that for $a^2\gg 1$, i.e., $\gamma^2\overline{\vartheta^2}\gg 1$, the spectral density of the radiation emitted by electrons in a thin layer of matter differs considerably from the Bethe–Heitler result.

In the general case where a is arbitrary, the functions $F_0(a)$ and $F_1(a)$ can be found by numerically integrating the relations determining these functions. The results of such calculations and their relationship to the asymptotic formulas (4.10) and (4.12) are depicted in Fig. 1.

5. ANALYSIS OF THE RESULTS OF THE EXPERIMENT DESCRIBED IN REF. 19

Anthony *et al.*[19] observed a decrease in the spectral density of the radiation emitted by 25-GeV electrons in a gold target of thickness $L=0.7\%L_R$ in the low-frequency range ($\omega\leqslant 20$ MeV) in comparison to the corresponding results of Bethe and Heitler. In this frequency range the coherence length l_c is large compared to L, i.e., the condition of the above theory are met. As applied to the given experiment, the values of a^2 and B in (4.9) are $a^2\approx 7.61$ and $B\approx 8.46$. Here, according to (4.14), if we allow for terms of order a^{-2} and B^{-1}, we get

$$\left\langle \frac{d\varepsilon}{d\omega}\right\rangle \approx 0.00538. \tag{5.1}$$

This asymptotic value of the radiation spectrum coincides to within a few percent with the results of exact calculations by formulas (3.5) and (4.4). Note that the asymptotic formula (4.14) and the result of exact calculations agree well if in (4.14) we keep several expansion terms, including the terms of order a^{-2} and B^{-1}. If in (4.14) we keep only the logarithmic term, the result for the radiation spectrum will be 1.4 times larger than the exact result. If the first-order correction is kept (i.e., terms of order a^{-2} and B^{-1} are ignored), the results will be lower by a factor 2 than the exact result. Hence the conditions of the experiment conducted by Anthony *et al.*[19] correspond to an intermediate case, where the Bethe–Heitler formula (2.8) ceases to be valid but the asymptotic formula (3.6), in which only the logarithmic term is taken into account, is still insufficient for describing the emission process.

Comparing the value established by formula (5.1) for the spectral density of radiation with the Bethe and Heitler's result $\langle d\varepsilon/d\omega\rangle_{\rm BH}\approx 0.0093$, we find that

$$\left\langle \frac{d\varepsilon}{d\omega}\right\rangle \approx 0.578\left(\frac{d\varepsilon}{d\omega}\right)_{\rm BH}, \tag{5.2}$$

which is in good agreement with the experimental result (see Fig. 2).

Thus, the experiment of Anthony *et al.*[19] confirms not only the Landau–Pomeranchuk–Migdal effect but also the effect of suppression of electron bremsstrahlung in a thin layer of matter predicted in Ref. 31.

As the electron energy increases, the frequency range and the range of target thicknesses in which the proposed theory is valid rapidly grow. Figure 3 depicts the dependence of the average spectral density of the radiation emitted by 250-GeV electrons on target thickness. It also presents the results of calculations of the radiation spectrum for different frequencies by the Migdal formula (2.5).

The results show that as the target gets thinner, all the LPM lines corresponding to different frequencies transform into the result of calculations of the radiation spectrum by formula (4.9), which is independent of ω. Only the range of applicability of the latter result is ω-dependent, and within a broad range of target thicknesses we have

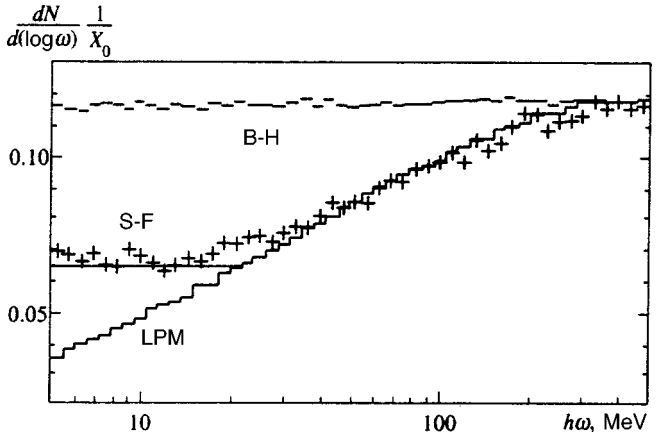


FIG. 2. Spectral density of the radiation emitted by 25-GeV electrons in a gold target whose thickness is 0.7% of the radiation length. The + designate the experimental data taken from Ref. 19; the LPM and B–H histograms represent the results of Monte Carlo calculations[19] done in the Landau–Pomeranchuk–Migdal theory and in the Bethe–Heitler theory; and the solid S–F line represents the results of our calculations by formula (3.5) with the distribution function (4.4).

$$\left\langle \frac{d\varepsilon}{d\omega}\right\rangle < \left(\frac{d\varepsilon}{d\omega}\right)_{\rm BH}.$$

When the thickness L decreases further, the mean square multiple scattering angle, $\overline{\vartheta^2}$, also decreases. Beginning with a thickness $L\sim 10^{-3}L_R$, the spectrum of radiation emitted by electrons in a thin layer of matter transforms into the corresponding Bethe–Heitler result.

Note that for $l_c\gg L$ the entire target acts as a single object with which the electrons interact and emit radiation. This means that the interference of the electromagnetic waves emitted by an electron from different sections of its path in the target is important. Here the target cannot be separated into several layers, since the radiation emitted in one layer cannot be assumed to be independent from that emitted in another layer. If we were to partition a target of thickness L

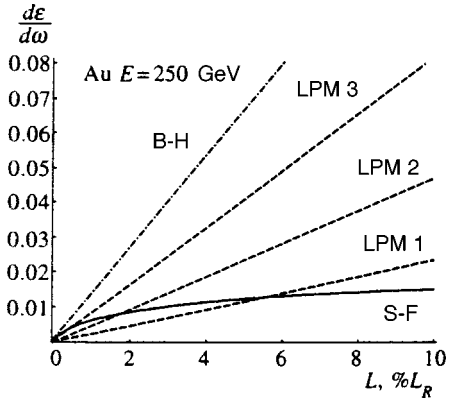


FIG. 3. Dependence of the spectral density of radiation on target thickness. The dot–dash line B–H represents the results of calculations by the Bethe–Heitler theory (formula (2.8)). The dashed lines LPM represent the results of calculations by Migdal's theory of the LPM effect (formula (2.5)) for different values of the energies of the emitted gamma quanta: $\omega=0.2$ GeV (1), 1 GeV (2), and 5 GeV (3). The solid curve S–F represents the results of our calculations by formula (3.5) with the distribution function (4.4), which describes the effect of bremsstrahlung suppression in a thin layer of matter.

into layers $\Delta L \sim 10^{-3} L_R$ and move these layers apart along the electron path to distances exceeding the coherence length l_c, the radiation emitted in each layer could be assumed to be independent, and the total spectral density in all these layers would equal the corresponding result of Bethe and Heitler (see also Ref. 24).

The present work was supported financially by the Ukrainian State Fund for Basic Research (the ''LPM-effect'' Project).

[1] M. L. Ter-Mikaelian, *High-Energy Electromagnetic Processes in Condensed Media*, Wiley-Interscience, New York (1972).
[2] A. I. Akhiezer and N. F. Shulga, *High-Energy Electrodynamics in Matter*, Gordon & Breach, London (1996).
[3] L. D. Landau and I. Ya. Pomeranchuk, Dokl. Akad. Nauk SSSR **92**, 535, 735 (1953).
[4] H. A. Bethe and W. Heitler, Proc. R. Soc. London, Ser. A **146**, 83 (1934).
[5] W. Heitler, *The Quantum Theory of Radiation*, Clarendon Press, Oxford (1954).
[6] A. B. Migdal, Dokl. Akad. Nauk SSSR **96**, 49 (1954).
[7] M. L. Ter-Mikaelian, Dokl. Akad. Nauk SSSR **94**, 1033 (1954).
[8] A. B. Migdal, Phys. Rev. **103**, 1811 (1956).
[9] V. M. Galitskiĭ and V. V. Yakimets, Zh. Éksp. Teor. Fiz. **46**, 1066 (1964) [Sov. Phys. JETP **19**, 723 (1964)].
[10] V. M. Galitsky and I. I. Gurevich, Nuovo Cimento **32**, 396 (1964).
[11] V. E. Pafomov, Zh. Éksp. Teor. Fiz. **49**, 1222 (1965) [Sov. Phys. JETP **22**, 848 (1966)].
[12] F. F. Ternovskiĭ, Zh. Éksp. Teor. Fiz. **39**, 171 (1960) [Sov. Phys. JETP **12**, 1123 (1961)].
[13] J. S. Bell, Nucl. Phys. **8**, 613 (1958).
[14] E. L. Feĭnberg, Usp. Fiz. Nauk **58**, 193 (1956).
[15] M. I. Ryazanov, Usp. Fiz. Nauk **114**, 193 (1974).
[16] V. A. Bazylev and N. K. Zhevago, *Emission by Fast Particles in Matter and External Fields*, [in Russian] Nauka, Moscow (1987).
[17] N. V. Laskin, A. S. Mazmanishvili, and N. F. Shul'ga, Dokl. Akad. Nauk SSSR **277**, 850 (1984) [Sov. Phys. Dokl. **29**, 638 (1984)]; N. V. Laskin, A. S. Mazmanishvili, N. N. Nasonov, and N. F. Shul'ga, Zh. Éksp. Teor. Fiz. **89**, 763 (1985) [Sov. Phys. JETP **62**, 438 (1985)]; A. I. Akhiezer and N. F. Shul'ga, Usp. Fiz. Nauk **151**, 385 (1987) [Sov. Phys. Usp. **30**, 197 (1987)].
[18] CERN Courier **34**, 12 (1994).
[19] P. L. Anthony, R. Becker-Szendy, P. E. Bosted, M. Cavalli-Sforza, L. P. Keller, L. A. Kelley, S. R. Klein, G. Niemi, M. L. Perl, L. S. Rochester, and J. L. White, Phys. Rev. Lett. **75**, 1949 (1995).
[20] R. Blankenbecler and S. D. Drell, Phys. Rev. D **53**, 6265 (1996).
[21] R. Baier, Yu. L. Dokshitzer, A. H. Mueller, S. Peigne, and D. Schiff, Nucl. Phys. B **483**, 291 (1997).
[22] A. Varfalomeev, V. I. Glebov, E. I. Denisov, A. M. Frolov, and A. S. Khlebnikov, Zh. Éksp. Teor. Fiz. **69**, 429 (1975) [Sov. Phys. JETP **42**, 218 (1975)].
[23] T. Stanev, Ch. Vankov, R. E. Streitmatter, R. W. Ellsworth, and Th. Bowen, Phys. Rev. D **25**, 1291 (1982).
[24] R. Blankenbecler, Phys. Rev. D **55**, 190 (1997).
[25] G. Moliere, Z. Naturforsch. **3a**, 78 (1948).
[26] H. Bethe, Phys. Rev. **89**, 1256 (1953).
[27] N. F. Shul'ga and S. P. Fomin, JETP Lett. **63**, 873 (1996).
[28] B. G. Zakharov, JETP Lett. **63**, 952 (1996); **64**, 781 (1996).
[29] X.-N. Wang, M. Gyulassy, and M. Plumer, Phys. Rev. D **51**, 3436 (1995).
[30] M. Gyulassy and X.-N. Wang, Nucl. Phys. B **420**, 583 (1994).
[31] N. F. Shul'ga and S. P. Fomin, JETP Letters **27**, 117 (1978).
[32] S. P. Fomin and N. F. Shulga, Phys. Lett. A **114**, 148 (1986).
[33] R. M. Barnett, C. D. Carone, D. E. Groom *et al.* Phys. Rev. D **54**, 1 (1996).
[34] L. Schiff, Phys. Rev. **83**, 252 (1951).

Translated by Eugene Yankovsky

Weak-pulse transparency enhancement in an optically dense three-level medium induced by a 2π pulse in a neighboring transition (*V*-scheme)

N. V. Denisova, V. S. Egorov, V. V. Kozlov, N. M. Reutova, P. Yu. Serdobintsev, and E. E. Fradkin

Physics Research Institute, 198904 St. Petersburg, Russia
(Submitted 15 April 1997)
Zh. Éksp. Teor. Fiz. **113**, 71–88 (January 1998)

The coherent *V*-configuration interaction between an optically dense resonantly-absorbing three-level medium (neon) and two ultrashort superradiance pulses with converging wave fronts is investigated experimentally and theoretically. Both separate and combined propagation of pulses with wavelengths $\lambda_1=614.3$ nm (strong field, $\theta_1\geqslant\pi$) and $\lambda_3=594.5$ nm (weak field, $\theta_3\approx\pi/20$) are studied. For propagation of a separate strong-field pulse, supertransparency of the absorbing medium was observed, which is associated with the generation of a soliton-like pulse at the difference frequency ($\Delta\nu\approx1700$ MHz) and the dispersion–diffraction stabilization effect. Under these conditions a weak-field pulse is completely absorbed. Combined propagation of the pulses leads to novel effects. A below-threshold pulse (weak field) was observed to pass through the absorber while interacting coherently with a strong-field pulse at a neighboring transition. It is shown theoretically that absorption of the weak pulse is reduced for two reasons: first, as a result of incoherent transparency of the resonance transition caused by emptying of the lower level by the field of the strong pulse, and second, as a result of coherent transfer of polarization between the upper levels via the two-photon processes. When the conditions for combined propagation are met, the latter mechanism ensures inversionless amplification of a weak pulse over a wide band of frequencies. In this case, the gain can even exceed the linear absorption coefficient in absolute value. A difference in propagation velocities of the weak and strong pulses was recorded experimentally, along with a shift in the carrier frequency of the weak field towards the red (≈600 MHz). A mechanism for transfer of phase modulation from a strong pulse to a weak pulse via the common lower level is discussed theoretically. [S1063-7761(98)00601-5]

1. INTRODUCTION

The possibility of light amplification and laser oscillation without population inversion, due to interference between atomic transitions leading to suppression of resonance absorption, was predicted by Kocharovskaya and Khanin in their 1988 paper (see Ref. 1), and independently by Harris[2] and Scully *et al.*[3] Among the systems that admit inversionless amplification, those that exhibit the most striking behavior are systems with Λ-type transition sequences, as a consequence of coherent population trapping. In contrast, we know of no papers, either theoretical or experimental, that deal with inversionless amplification of ultrashort light pulses by three-level media with a *V*-type configuration. In fact, the only discussions of pulse propagation for neighboring transitions in the *V*-scheme have been from the point of view of generating two-frequency solitons (simultons).[4,5]

In this paper we investigate theoretically and experimentally the combined propagation of two ultrashort light pulses (i.e., pulses whose durations are shorter than all relaxation times of the atomic system), produced by neighboring transitions, through an optically dense three-level medium with the *V*-configuration and an inhomogeneously broadened absorption line. Before the arrival of such pulses, all atoms in the medium are in the lower state (see Fig. 1). First, a pulse that is resonant with the lower-frequency transition and has intensity $\mathscr{E}_1$ sufficient to create an inversion between levels *2* and *1* (i.e., with area greater than π) is injected into the medium. A weak pulse of the neighboring high-frequency transition with intensity $\mathscr{E}_3$ then enters the absorbing medium after a certain time delay; this pulse is too weak to change the population of level *3*. Experimentally, an abrupt attenuation of the absorption for the weak field is observed. We have proven theoretically that when the weak pulse is shorter than the strong pulse, there exists a wide band of frequencies for which the spectral components of the weak field can undergo amplification.

2. DESCRIPTION OF EXPERIMENT

2.1. Source of superradiance

Two superradiance pulses, a long-wavelength pulse ($\lambda_1=614.3$ nm, $2p_6-1s_5$ transition, referred to as the ''red'' pulse) and a short-wavelength pulse ($\lambda_3=594.5$ nm, $2p_4-1s_5$ transition, referred to as the ''yellow'' pulse), arise essentially simultaneously when high-voltage nanosecond discharges (whose voltage growth rate is $\approx6\times10^{11}$ V/s) are excited in a narrow thick-walled glass capillary with internal diameter $d=0.4$ mm and length $L=27$ cm. The geometric parameters of the capillary were chosen in such a way as to ensure high spatial coherence of the radiation field: the Fresnel numbers $F_{1,3}=d^2/L\lambda_{1,3}$ for both wavelengths are ~1. In this system, the gas is excited by an ionizing potential gradient wave. We described and investigated the superradi-

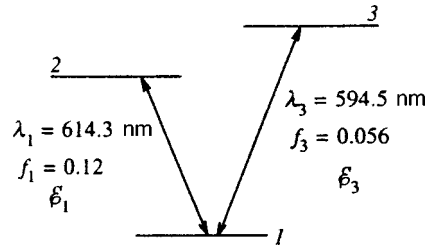


FIG. 1. Three-level V-scheme with notation used in the text.

ant operating regime of this source in Ref. 6; however, in this paper we also obtained simultaneous ''single-mode oscillation'' at the two wavelengths discussed above, a first in such a narrow capillary. The optimal neon pressure was $p \approx 0.7-0.9$ Torr. The durations of the red superradiance pulse and the yellow superradiance pulse were 1–1.5 ns and 0.5–0.8 ns respectively. The maximum power of the red pulse was 0.7 W; depending on the excitation conditions, that of the yellow pulse was approximately 10–40 times less. For this reason, in what follows we will refer to these as the strong-field (red) pulse and weak-field (yellow) pulse. The yellow pulse appears 0.5 ns later than the red pulse, but because the red pulse is wider than the yellow, the two pulses overlap completely in time. The carrier frequency of the red pulse is shifted toward the violet end of the spectrum by ≈ 200 MHz $\approx 1/7 \Delta \nu_{\mathrm{Doppl}}$, while that of the yellow pulse is shifted toward the red end by 300 MHz. The spectral widths were ≈ 1.2 GHz for the red pulse and ≈ 1.4 GHz for the yellow pulse.

2.2. Absorbing medium

Pulses from the light source were directed into the absorbing medium, which was the plasma of a positive glow-discharge neon column containing metastable neon atoms in the $1s_5(2p^5 3s)$ state with a maximum concentration of about 10^{12} cm^{-3}. The discharge was created in a glass tube of length 30 cm and diameter 10 mm at a working pressure of 1.6 Torr, a current of 2–7 mA, and a tube voltage of about 1.5 kV. The input windows of the tube were sealed into it at Brewster's angle. To measure the concentration, a mirror was fastened behind a holder oriented perpendicular to the tube axis. The reflection coefficient of the mirror was measured at a wavelength $\lambda = 585.2$ nm. The concentration of metastable neon atoms in the $1s_5$ level was measured using a mirror behind the tube[7] at wavelengths $\lambda = 614.3$ and 594.5 nm with an FEU-38 photomultiplier and a digital voltmeter. Since the distribution of metastable atoms over the tube diameter is nonuniform in a positive glow-discharge column (it is described by a zero-order Bessel function), in reality we measured the mean of the concentration over the diameter.

Before a pulse arrives at the medium, the population is already distributed in a certain way among the three levels of the V-configuration. However, the populations of levels $2p_4$ and $2p_6$ are quite low—of order 10^{10} cm^{-3}—which is much lower than the population of level $1s_5$; therefore, we neglect these in the following discussion. The lower metastable level $1s_5$ is common to the two transitions, so the arrangement of the three levels corresponds to a medium with the V-configuration.

The light beam passing through the absorbing medium was focused by a lens in such a way that it converged directly behind the output window of the absorbing cell. This beam geometry (i.e., with a nonplanar wave front) creates conditions that can lead to a supertransparency regime; see Refs. 8–10.

If the interaction between the light and the resonant absorbing medium is to be coherent, the condition $\tau \ll T_1, T_2$ must be fulfilled, where τ is the duration of either of the pulses. In dilute atomic gases and low-pressure discharges, this inequality is fulfilled for the nanosecond regime of pulse durations. Under the conditions of our experiment ($I \leqslant 6$ mA), the phase memory time of the medium for the $2p_6-1s_5$ and $2p_4-1s_5$ transitions is $T_2 = 11$ ns, according to our estimates. In calculating T_2 we included radiative and collisional relaxation processes in the absorbing medium; collisions with electrons were not taken into account due to their low concentration in the discharge under the conditions stated above, namely $n_e < 10^{12}$ cm^{-3} (see Ref. 11). The population relaxation time T_1 for the transitions we used was measured in Ref. 12 and is approximately 19 ns. These estimates show that the duration of the pulses under study here is shorter than the relaxation time by more than an order of magnitude, and hence the interaction of these pulses with the two transitions will be coherent in nature. The temporal coherence of the interaction is supplemented by the condition of spatial coherence, which is ensured by the single-mode character of the light source.

It is important to note that in our experiments we used the same two transitions in neon both for generation and for absorption. Therefore, both the yellow and red pulses satisfy the condition for resonant interaction.

2.3. Experimental setup

Because of the high time resolution of our experimental setup ($\sim$300 ps), a general sketch of which is shown in Fig. 2, we were able to make measurements of the spectral–temporal characteristics of both pulses, and also to monitor the optical density of the absorbing layer as these measurements were made. We investigated both combined and separate interactions of the pulses with the three-level medium in the V-configuration. To ensure that the pulses propagated independently, the beams were offset from one another at the input to the medium by the prism P (see Fig. 2); the separation was approximately 2 mm, which equalled their diameters. This enabled us to create individual interaction channels for the red and yellow pulses in the absorption cell.

To make the timing measurements, we directed the fundamental beams passing through the prism and the absorbing cells into an optical delay line and then recorded them using a photodiode (LFD-2) and an S7-8 sampling oscilloscope. The second light beam reflected from the input surface of the prism served to synchronize and trigger the oscilloscope; as a light detector we used an RCA photomultiplier with time resolution ~ 1 ns.

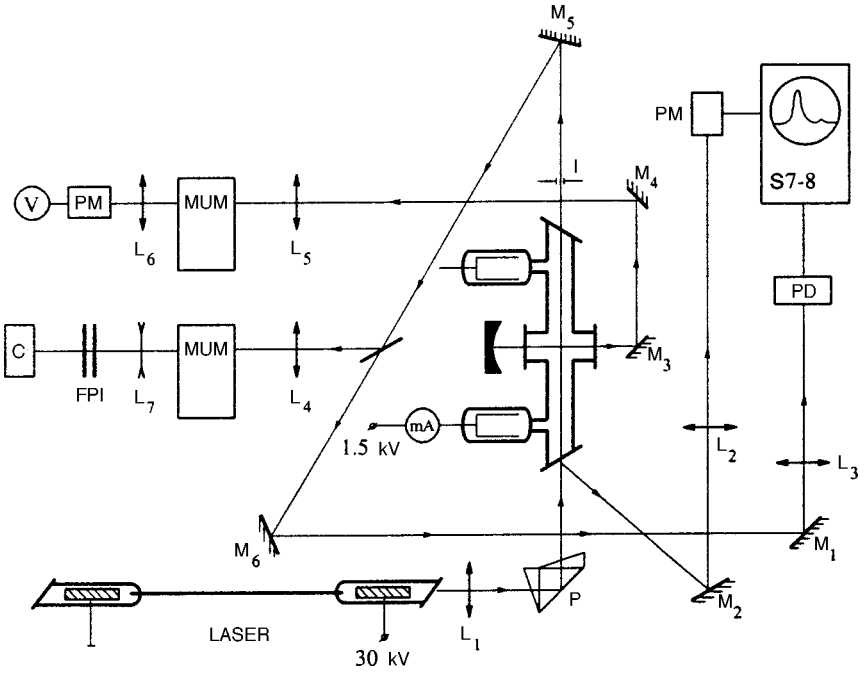


FIG. 2. Experimental setup for observing individual propagation of the pulses. M_1–M_6—mirrors, L_1–L_7—lenses, P—prism, PD—LFD–2 photodiode, S7-8—stroboscopic oscilloscope, V—digital voltmeter, mA—milliammeter, I—iris diaphragm, FPI—Fabry–Perot interferometer, C—camera.

The spectral characteristics of the pulses were investigated using an IT28-30 Fabry–Perot interferometer with a 30 mm ring brace and planar wideband mirrors with dielectric coatings (the free spectral range was 5000 MHz, the half-width of the system feedback loop was 260 MHz). For preliminary dispersion, we used an MUM monochromator. The interferometer operated with diverging beams. In order to eliminate the effect of spontaneous emission we used an iris diaphragm at the output of the absorbing tube, and the tube itself was blackened. The detecting system was entirely screened from external sources of light.

To bring about the combined propagation of red and yellow pulses in the absorbing medium, the beams emitted by the source were directed by a mirror into one propagation channel. The prism P that separated the yellow and red beams was positioned behind the absorbing cell, which enabled us to tune the recording system by rotating this prism and thereby measure the parameters of the yellow and red pulses.

3. EXPERIMENTAL RESULTS

3.1. Independent pulse propagation

In the geometry we used to observe independent propagation of the pulses, the beams did not overlap at the input to the medium. By investigating experimentally the temporal and spectral characteristics of the red and yellow pulses before and after interacting with the absorbing medium, we were able to compare the results of the interaction for various input parameters of the pulses.

Our studies of the propagation of the red pulse through the absorbing medium showed that this pulse entered a regime of supertransparency when its pulse energy exceeded a certain threshold ($\theta_1 \geqslant \pi$),[8,10] i.e., the pulse was observed to propagate over an anomalously long distance compared to the classical limits associated with self-induced transparency. We noted a red shift in the carrier frequency of the strong-field pulse away from resonance, with a value of order $\sim \Delta\nu_{\mathrm{Doppl}}$, which is typical of the supertransparency effect.

The oscilloscope traces shown in Fig. 3 allow us to compare the temporal shapes of the input pulse and the red pulse passing through the absorbing medium. The maximum delay experienced by the pulse as it passed through the absorbing medium was 1.5 ns, at which point the combined oscilloscope trace shows that the red pulse has completely emerged from the profile of the incident pulse. Pulses with high input amplitude were delayed less than pulses with low input amplitude, and passed through with less absorption.

For concentrations of absorbing atoms of approximately 10^{12} cm^{-3}, the red pulse was attenuated by about a factor of 8–20 (depending on the input amplitude). At lower concentrations ($\approx 3\times 10^{11}$ cm^{-3}) the pulse passed through with almost no attenuation.

When the yellow pulse propagated through the optically

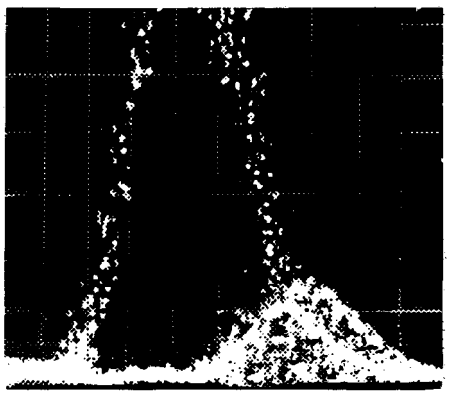


FIG. 3. Oscilloscope traces of the input red pulse (upper trace) and transmitted pulse through the absorbing medium (lower trace), obtained for independent propagation of the pulses. The time scale is 0.5 ns/div, t_d=1.3 ns.

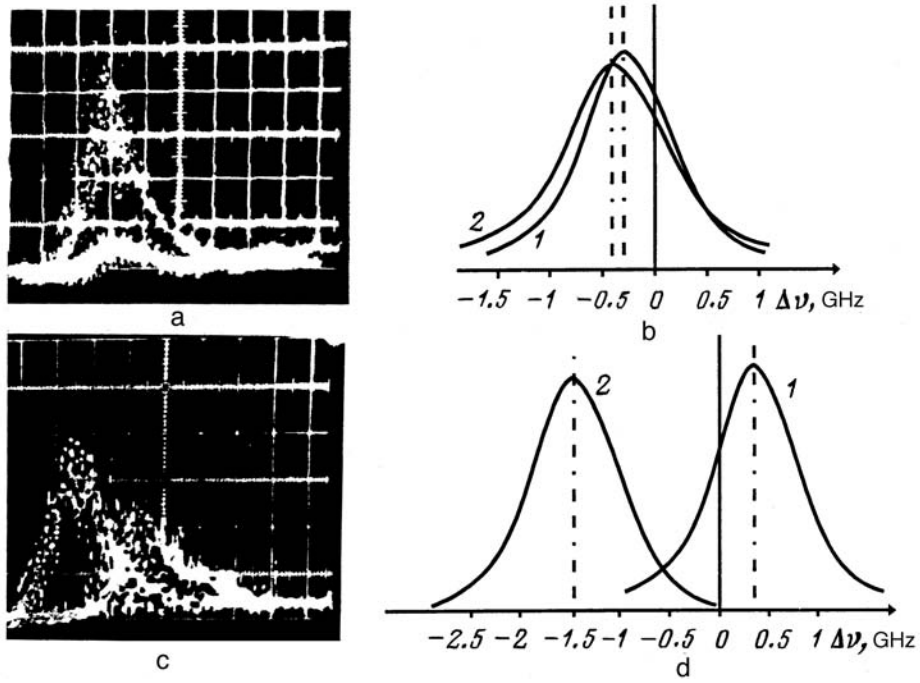


FIG. 4. Spectral and temporal characteristics of the yellow pulse (a,b) and red pulse (c,d) at the input to the absorbing medium (*1*) and at the output from it (*2*), obtained for combined propagation of the two pulses. The time scale is 0.5 ns/div. In Fig. 4c, the input pulse (left-hand trace) is attenuated by a filter.

dense absorbing layer, we observed an entirely different picture. In a medium with concentration $\geqslant 5\times 10^{11}$ cm^{-3}, the pulse was attenuated so strongly that we were unable to detect the pulse at all at the output of the absorbing cell, even with the sensitivity of the detection system set to maximum. It is obvious that in this case the energy of the yellow pulse at the input to the medium is insufficient to overcome the threshold for self-induced transparency ($\theta_3 \sim \pi/20$), in which case it is completely absorbed.

3.2. Combined propagation of two resonance pulses

During the combined propagation of red and yellow pulses in the absorbing medium we noticed that they interacted with one another. This interaction was revealed by changes in how the pulses interacted with the medium compared to the independent propagation case described in the previous section.

The most striking proof that the pulses interacted was the fact that in this regime of propagation the yellow superradiance pulse passing through an absorbing medium with optical density k_3L of 15 (where k_3 is the linear absorption coefficient for transition *1-3*) was attenuated by only a factor of 8, which enabled us to identify changes in its temporal and spectral profile since they lay within the resolution limits of our detection system.

In Fig. 4a we show an oscilloscope trace of the yellow pulse passing through the absorbing medium against the background of the input signal. There is clearly no temporal delay in the pulse compared to the incident pulse as the former passed through the medium, which corresponds to a velocity of the pulse in the medium close to the velocity of light.

Figure 4b shows changes in the spectrum of the yellow pulse relative to the spectrum of the input pulse caused by its passage through the absorbing medium. The center of the spontaneous emission line of neon with wavelength $\lambda = 594.5$ nm served as a reference for the yellow line. Spectral analysis of the yellow pulse showed that after passing through the absorbing medium, its carrier frequency was redshifted by ~ 600 MHz; in this case the offset of the carrier frequency of the yellow pulse from resonance was $\Delta\nu_{03} \approx 300$ MHz.

Analogous spectral and temporal measurements were also made for the red pulse. Oscilloscope traces of the red pulse at the input and output to the medium are shown in Fig. 4c.

The peak of the pulse passing through the medium is obviously delayed relative to the peak of the incident pulse by ~ 1 ns, which is typical of an input pulse whose amplitude is not too high. Noteworthy is the fact that in the presence of the yellow pulse, there is somewhat more energy reradiated by the red pulse at its trailing edge than is the case when the pulses propagate independently.

The spectra of a red input pulse and a red pulse transmitted through the absorbing cell are compared in Fig. 4d. In this figure, the spontaneous line center for neon at $\lambda = 614.3$ nm is shown as a reference. The red shift of the pulse carrier frequency at the output of the absorbing medium relative to the spectrum of spontaneous emission of neon is the same as for the case in which the red pulse propagates independently: $\Delta\nu \approx 1700$ MHz. The initial offset of the red pulse carrier frequency from resonance is $\Delta\nu_{01} \approx -200$ MHz.

The abrupt decrease in absorption of the weak field observed in these experiments can be explained theoretically by solving the combined Maxwell–Bloch equations for a three-level medium. This model, which we describe below, cannot include all the peculiarities of the experiment, and in particular is built around the plane-wave approximation. Nevertheless, even within the framework of this simple model we will

be able to describe the fundamental regularities of the effect, and thereby achieve the double goal of finding amplification mechanisms in a situation that is maximally close to experiment and at the same time arriving at concrete recommendations for optimizing the experimental setup.

4. THEORETICAL DISCUSSION

4.1. Fundamental equations of the model

Our theoretical discussion is based on the joint solution of the two equations for the fields in transitions *1–2* ($\mathscr{E}_1(t,z)$) and *1–3* ($\mathscr{E}_3(t,z)$) and the system of Bloch equations for the three-level system (*V*-scheme) (see Fig. 1). The experimental results clearly indicate that the strong field plays a dominant role in the dynamics of pulse propagation for the yellow transition. On the other hand, the feedback from the weak field is small and causes only an insignificant change in the temporal profile at the trailing edge of the red pulse. In our theoretical model we assume that the short-wavelength pulse $\mathscr{E}_3(t,z)$ is weak, so that its interaction with transition *1–3* can be treated in the linear approximation. We also neglect its effect on the conditions for propagation of the strong long-wavelength pulse, thereby abandoning *a priori* any attempt to describe secondary effects exerted by the weak field on the strong field. The system of Maxwell–Bloch equations then separates into two subsystems, one of which describes the interaction of pulse $\mathscr{E}_1(t,z)$ with the transition *1–2*:

$$\frac{\partial}{\partial z}\mathscr{E}_1=-\frac{n_1}{c}\frac{\partial}{\partial u}\mathscr{E}_1+\kappa_1\langle\mathscr{P}_1(t,z,\Omega)\rangle,$$

$$\frac{\partial}{\partial u}\mathscr{P}_1=-i\Omega\mathscr{P}_1+\mathscr{E}_1\mathscr{N}_1, \tag{1}$$

$$\frac{\partial}{\partial u}\mathscr{N}_1=-\frac{1}{2}(\mathscr{E}_1\mathscr{P}_1^*+\mathscr{E}_1^*\mathscr{P}_1),$$

while the other subsystem

$$\frac{\partial}{\partial z}\mathscr{E}_3=-\frac{n_3}{c}\frac{\partial}{\partial u}\mathscr{E}_3+\kappa_3\langle\mathscr{P}_3(t,z,\Omega)\rangle,$$

$$\frac{\partial}{\partial u}\mathscr{P}_3=-i\Omega\mathscr{P}_3-\gamma_3\mathscr{P}_3+\mathscr{E}_3\mathscr{N}_3+\frac{i}{2}\mathscr{E}_1\mathscr{P}_2, \tag{2}$$

$$\frac{\partial}{\partial u}\mathscr{P}_2=-i\Omega\mathscr{P}_2-\gamma_2\mathscr{P}_2+\frac{i}{2}(\mathscr{E}_1^*\mathscr{P}_3+\mathscr{E}_3\mathscr{P}_1^*),$$

describes the interaction of pulse $\mathscr{E}_3$ with transition *1–3*. In writing these equations we use the notation

$$\kappa_1=2\pi\omega_1 d_1^2 N/n_1 c\hbar,\quad \kappa_3=2\pi\omega_3 d_3^2 N/n_3 c\hbar,$$

where d_1 is the dipole moment of transition *1–2*, d_3 is the dipole moment of transition *1–3*, N is the concentration of atoms in level *1*, $\mathscr{P}_1$ is the polarization of transition *1–2*, $\mathscr{P}_2$ is the polarization of transition *2–3* (more precisely, *2–1–3*, since the transition *2-3* is forbidden in the dipole approximation), $\mathscr{P}_3$ is the polarization of transition *1–3*, $\mathscr{N}_1$ is the population difference between levels *2* and *1*, $\mathscr{N}_3$ is the population difference between levels *3* and *1*, γ_2 and γ_3 are polarization relaxation times, and n_1 and n_3 are the nonresonant refractive indices.

The total field is related to its envelopes by the expressions

$$\tilde{\mathscr{E}}_j(t,z)=\frac{\hbar}{d_j}\mathscr{E}_j(t,z)\exp[-i(\omega_j t-K_j z)],\quad j=1,3,$$

where $K_j=\omega_j/c$ are the wave numbers for the corresponding frequencies in vacuum. In these field equations, angle brackets indicate averaging over the resonant frequencies for the entire ensemble of atoms:

$$\langle\ldots\rangle=\int\ldots f(\Omega)\,d\Omega.$$

An important step in deriving Eqs. (1) and (2) is the neglect of relaxation effects in the density matrix equations for transition *1–2* while including their effect on the atomic variables for transition *1–3*. At first glance this may appear strange, since the duration of the yellow pulse in our experiments was equal to or even somewhat shorter than that of the red pulse. Actually, the situation we are dealing with here involves a manifestation of the fundamental property of self-induced transparency, in which a short pulse that satisfies the threshold conditions for this effect phases an entire ensemble of atoms with a wide inhomogeneously broadened absorption line. For an ensemble of dipoles with differing carrier frequencies distributed over a wide inhomogeneously broadened absorption line, attenuation of the macroscopic polarization takes place within a time of order T^*. In this case the polarization of each individual dipole is preserved at the micro-level for a much longer time, up to γ_1^{-1} (where γ_1 is the rate of attenuation of the polarization $\mathscr{P}_1$).

McCall and Hahn were the first to note[13] that a pulse whose duration satisfies the condition $T^*\ll\tau\ll\gamma_1^{-1}$, and whose area at the input to the absorbing medium exceeds π, propagates with low absorption due to this same ability to induce phasing of the dipole oscillations of the entire ensemble of atoms; its energy is significantly attenuated only when the propagation time through the thickness of the absorbing medium is greater than γ_1^{-1}. On the other hand, for weak pulses in the same range of duration, the pulse energy is incoherently scattered due to dephasing of the dipole oscillations over times of order T^*. Under our experimental conditions, we estimate the difference in attenuation rates for the strong and weak field to be $H(T^*)^{-1}/\gamma_1\approx 15$. This estimate proves that we are correct in including relaxation effects for the weak field and neglecting these effects for the strong field.

It is known (see, e.g., Ref. 14), that when a pulse entering the medium has a sufficiently smooth envelope profile whose area exceeds π, in traversing a distance equal to approximately 2–3 absorption lengths it is transformed into a self-induced transparency soliton with area 2π. Under our experimental conditions, the optical density of absorbing atoms for transition *1–2* was 30, and effective interaction of both pulses took place over essentially the entire length. From a physical point of view, the most transparent results

are obtained when we choose the shape of the field envelope at transition *1–2* to be a self-induced transparency soliton,

$$\mathscr{E}_1(t,z)=A_0\ \mathrm{sech}[(t-t_0-z/v)/\tau]\exp[i(\delta t-kz)], \quad (3)$$

in which case the polarization and inversion will have the form

$$\mathscr{P}_1(t,z,\Omega)=2\frac{[i(\Omega\tau+\delta\tau)-\tau\partial/\partial u]\mathrm{sech}[(t-t_0-z/v)/\tau]}{1+(\Omega\tau+\delta\tau)^2}\times\exp[i(\delta t-kz)], \quad (4)$$

$$\mathscr{N}_1(t,z,\Omega)=-1+2\frac{\mathrm{sech}^2[(t-t_0-z/v)/\tau]}{1+(\Omega\tau+\delta\tau)^2}. \quad (5)$$

We will return to the question of how correct it is to choose the field in the form of a self-induced transparency soliton in discussing the results. Substituting the expression for the field in the form (3) into Eq. (1) determines the relation between the pulse amplitude and its duration (where $A_0\tau=2$ and the pulse velocity is v),

$$\frac{1}{v}=\frac{n_1}{c}+\kappa_1\tau^2\left\langle\frac{1}{1+(\Omega\tau+\delta\tau)^2}\right\rangle, \quad (6)$$

and the dispersion relation

$$k(\delta)=\frac{\delta n_1}{c}-\kappa_1\tau\left\langle\frac{\Omega\tau+\delta\tau}{1+(\Omega\tau+\delta\tau)^2}\right\rangle. \quad (7)$$

Here δ is the offset of the carrier frequency of the self-induced transparency soliton from the center of transition *1–2*, and k is a correction to the wave number K_1. The quantity t_0 determines the position of the self-induced transparency soliton at the input to the medium relative to the position of the ''weak'' pulse. It is convenient to choose for t_0 the initial delay of pulse $\mathscr{E}_3$ compared to the soliton $\mathscr{E}_1$.

4.2. Evolution of the weak field

We now consider the system of equations for a weak field. The evolution of the inversion $\mathscr{N}_3$ entering into Eq. (2) is entirely determined by the dynamics of the self-induced transparency soliton, because we assume that the field $\mathscr{E}_3$ is so weak that it cannot cause any change in the population difference. Thus, $\mathscr{N}_3$ can be written in the form

$$\mathscr{N}_3(t,z,\Omega)=-\frac{[1-\mathscr{N}_1(t,z,\Omega)]}{2}=-1+\frac{\mathrm{sech}^2[(t-t_0-z/v)/\tau]}{1+(\Omega\tau+\delta\tau)^2}. \quad (8)$$

The simplest and physically most transparent results are obtained if we assume that the duration of the pulse $\mathscr{E}_3(t,z)$ is much shorter than that of the self-induced transparency soliton.[1)] We can then neglect the time derivatives of $\mathscr{E}_1$, $\mathscr{P}_1$, $\mathscr{N}_3$, and $\mathscr{N}_1$ compared to derivatives of the more rapid variables $\mathscr{E}_3$, $\mathscr{P}_3$ and $\mathscr{P}_2$. In this approximation the equation for the evolution of the weak field reduces to the form

$$\frac{\partial\mathscr{E}_3}{\partial z}=\kappa_3\left\langle\exp[-(i\Omega+\gamma)u]\left\{\mathscr{N}_3\int^u\mathscr{E}_3\times\exp[(i\Omega+\gamma)u]du-\frac{1}{4}\mathscr{E}_1\mathscr{P}_1^*\int^u\int^u\mathscr{E}_3\times\exp[(i\Omega+\gamma)u]du\ du\right\}\right\rangle, \quad (9)$$

where in place of time t we introduce the wave variable $u=t-z/(c/n_3)$. In deriving Eq. (9) we have used the fact that the duration of pulse $\mathscr{E}_3(t,z)$ is small compared with A_0^{-1}, and have set $\gamma_2=\gamma_3=\gamma$. We note immediately that the polarization $\mathscr{P}_3(t,z)$, which controls the evolution of the field, has two components. The first describes the linear interaction of the field with a medium whose concentration of atoms $\mathscr{N}_3$ is modulated in time by the field of the self-induced transparency soliton. We can tentatively call this interaction between the pulses incoherent. The presence of the second component of the polarization $\mathscr{P}_3$ is due to the coherent transfer of the interaction between the two transitions via the polarization $\mathscr{P}_2$. It is this transfer of coherence that essentially distinguishes this model from the two-level system.

The wave equation (9) describes the propagation of nonmonochromatic waves through a dispersive medium in the z direction. Let us assume that a wave is excited at the boundary of the medium by an incident pulse

$$\mathscr{E}_3(z=0,t)=e_0(u) \quad (10)$$

with frequency spectrum

$$F(\tilde{\omega})=\frac{1}{2\pi}\int_{-\infty}^{+\infty}e_0(u)e^{i\tilde{\omega}u}du. \quad (11)$$

Because the spectral components propagate independent of one another in a linear medium, the behavior of the wave is given by a superposition of harmonic waves

$$\mathscr{E}_3(z,u)=\int_{-\infty}^{+\infty}F(\tilde{\omega})\exp[-i\tilde{\omega}u+ik(\tilde{\omega})z]d\tilde{\omega}. \quad (12)$$

The relation $k(\tilde{\omega})$ for this dispersive medium can be found from Eq. (9):

$$k(\tilde{\omega})=\kappa_3\left\langle-\frac{(\Omega-\tilde{\omega})+i\gamma}{(\Omega-\tilde{\omega})^2+\gamma^2}\mathscr{N}_3-\frac{i}{4}\mathscr{E}_1\mathscr{P}_1^*\frac{(\Omega-\tilde{\omega})^2-\gamma^2+2i\gamma(\Omega-\tilde{\omega})}{[(\Omega-\tilde{\omega})^2+\gamma^2]^2}\right\rangle. \quad (13)$$

The imaginary part of the dispersion relation reveals that the field–medium system is not conservative, so that either amplification ($\mathrm{Im}[k(\tilde{\omega})]<0$) or attenuation ($\mathrm{Im}[k(\tilde{\omega})]>0$) of the field are possible within certain frequency ranges.

Within the broad line for the red pulse ($\tau>T^*$), and when conditions ensure that this pulse propagates at exact resonance with transition *1–2*, we can calculate the absorption coefficient of the field for the yellow transition:

$$\mathrm{Im}[k(\tilde{\omega})]=\kappa_3T^*\left\{\frac{1}{1+(\tilde{\omega}T^*)^2}\right.$$

$$-\left[1+\frac{1-(\tilde{\omega}\tau)^2}{1+(\tilde{\omega}\tau)^2}\,(1-\tanh[(t-t_0-z/v)/\tau])\right]$$
$$\times\frac{\cosh^{-2}[(t-t_0-z/v)/\tau]}{1+(\tilde{\omega}\tau)^2}\Bigg\}. \tag{14}$$

In deriving Eq. (14) we have taken advantage of the smallness of $\gamma\tau$ and have chosen the averaging function $f(\Omega)$ in the form

$$f(\Omega)=\frac{T^*}{\pi}\,\frac{1}{1+(\Omega T^*)^2}. \tag{15}$$

The larger value of the second term in Eq. (14) is responsible for the smaller absorption coefficient of the weak field. It is important to note that the strong field always leads to a decrease in the absorption coefficient of the yellow pulse. For those spectral components of the weak field that lie within an interval whose width equals that of the strong-field spectrum ($\omega\tau<1$), the polarization $\mathscr{P}_1$ at the leading edge of the red pulse increases the gain, while at the trailing edge it reduces it. Conversely, for the remaining spectral components ($\omega\tau>1$), the gain is less pronounced at the leading edge and more so at the trailing edge. We also note that when the line is homogeneously broadened (for which case we substitute $f(\Omega)=\delta(\Omega)$ into Eq. (13)), depending on the relative positions of the pulses, both amplification of the weak field and attenuation of the latter are possible. The decrease in absorption due to emptying of level *1* occurs for all spectral components of the weak field at any instant in time. On the other hand, the contribution of the polarization to amplification of a weak pulse is approximately $(\gamma\tau)^{-1}$ times smaller than that due to the dynamics of the population difference, and in contrast to inhomogeneous broadening lines, it is always negative at the leading edge of the red pulse (i.e., the strong field draws energy from the weak field) and positive at its trailing edge (the strong field delivers energy to the weak field).

We now discuss the phase characteristics of the pulses. In Ref. 10 we described supertransparency, which occurs when a coherent pulse with a converging wave front propagates through a dense resonant absorbing medium. The pulse is propagated over an anomalously long distance because the pulse carrier frequency is driven toward the low-frequency end of the spectrum beyond the limits of the inhomogeneously broadened line. In Ref. 10 we were able to explain this phenomenon by invoking the mechanism of dispersive–diffractive stabilization. In this paper our interest centers on the dynamics of the yellow-transition pulse, so we regard the phase characteristics of the strong field as given, and concentrate our attention on how they affect the spectrum of the weak pulse.

In these experiments we detected a shift in the spectrum of the field for the yellow transition. The most faithful measure of this effect is the pulse frequency averaged over the pulse spectrum, or equivalently the modulation frequency averaged over time and normalized by the field amplitude spectrum:

$$\langle\delta_3(z)\rangle_u=\int_{-\infty}^{+\infty}\frac{\partial}{\partial u}\,\varphi(u,z)\mathscr{E}_3(u,z)\,du\Bigg/\int_{-\infty}^{+\infty}\mathscr{E}_3(u,z)\,du, \tag{16}$$

where $\varphi(u,z)$ is the phase of the pulse. In the previous expression for the field we separated out a factor that oscillates at the optical frequency, so Eq. (16) is written not for the average frequency, but rather for δ_3, the average offset from the frequency ω_3. Using the spectral representation (12) and integrating overtime, we obtain

$$\langle\delta_3(z)\rangle_u=-\frac{\int_{-\infty}^{+\infty}\tilde{\omega}F(\tilde{\omega})\exp\{\mathrm{Im}[k(\tilde{\omega})]z\}\cos\{\mathrm{Re}[k(\tilde{\omega})z]\}\,d\tilde{\omega}}{\int_{-\infty}^{+\infty}F(\tilde{\omega})\exp\{\mathrm{Im}[k(\tilde{\omega})]z\}\cos\{\mathrm{Re}[k(\tilde{\omega})z]\}\,d\tilde{\omega}}, \tag{17}$$

where the expression for $k(\tilde{\omega})$ is obtained from Eq. (13) by making the substitution $t=z/(c/n_3)$. In deriving Eq. (17), we assumed that the weak pulse is not phase modulated at the input to the medium. We also assume the spectrum of the input pulse is symmetric. These two conditions taken together reflect the fact that the carrier frequency of the field coincides with the center frequency ω_3 of the transition.

In the approximation we used above ($\gamma^{-1}\gg\tau>T^*$) and for moderate offsets δ ($\delta\tau<1$), we obtain the following expression for $\mathrm{Re}[k(\tilde{\omega})]$:

$$\mathrm{Re}[k(\tilde{\omega})]=\kappa_3T^*\Bigg\{\frac{\tilde{\omega}T^*}{1+(\tilde{\omega}T^*)^2}+\cosh^{-2}\left[t_0+\left(\frac{1}{v}-\frac{1}{c}\right)z\right]\times\Bigg(\delta\tau\left[\frac{1}{1+(\tilde{\omega}\tau)^2}-2\,\frac{3(\tilde{\omega}\tau)^2-1}{[1+(\tilde{\omega}\tau)^2]^3}\right]-\tilde{\omega}\tau\left[\frac{1}{1+(\tilde{\omega}T^*)^2}+\frac{2}{[1+(\tilde{\omega}\tau)^2]^2}\right]\Bigg)\Bigg\}. \tag{18}$$

Here $n_1=n_3=1$. The antisymmetry of Eq. (18) with respect to $\tilde{\omega}$ at $\delta=0$ and the symmetry of expression (14) under the same conditions makes the right-hand side of (17) vanish. That is, we conclude that the carrier frequency of the ''weak'' field does not shift if the red pulse propagates in exact resonance with the transition *1–2* ($\delta=0$). Conversely, an offset in the red pulse carrier necessarily implies a shift in the frequency of the yellow pulse.

5. DISCUSSION OF EXPERIMENTAL RESULTS

Rather than proceed directly to a discussion of the interaction of a two-frequency pulse with the three-level medium, we first touch briefly on the distinctive features of the regime of individual pulse propagation. Up to now there have been fairly detailed studies of the numerous coherent phenomena that occur when short light pulses interact with a two-level medium.[14] These phenomena can be classified as either linear and nonlinear. In our experiments we have found both types of phenomena: the interaction of the yellow pulse with transition *1–3* is linear, while the interaction of the red pulse with transition *1–2* is nonlinear.

The yellow pulse power was approximately 0.02 W. Knowing the beam diameter at the input of the absorption cell $D\sim 1$ mm and the value of the $2p_4-1s_5$ transition dipole moment, $d_3\approx 3\times 10^{-18}$ cgs electrostatic units, we can estimate the value of the input area of the yellow pulse: $\theta_3=\Omega_3\tau_3\approx 5\times 10^{-2}\pi$. In this case, the inverse absorption length can be estimated by the magnitude of the inverse linear absorption coefficient

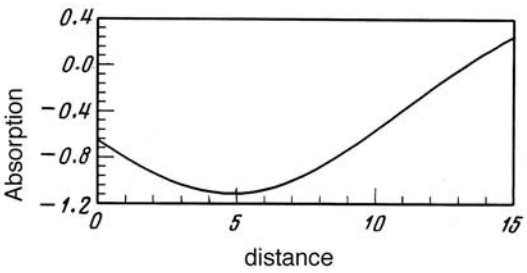
$$k_3=\frac{2}{\Delta\nu_{\mathrm{Doppl}}}\sqrt{\frac{\ln 2}{\pi}}\,\frac{\pi e^2}{m_0 c}\,fN\approx 1.152\times 10^{-12}\,N\ \ \mathrm{cm}^{-1},$$

where f is the oscillator strength of the transition and N is the concentration of absorbing atoms in the $1s_5$ state. When the concentration of absorbing atoms is $N\approx 5\times 10^{11}$ cm^{-3}, $k_3L\approx 15$. Calculations like those of Ref. 15, for example, that take into account classical anomalous absorption resulting from the large width of the pulse spectrum, show that the energy of a pulse traversing a layer with optical thickness $k_3L\approx 15$ should be attenuated by more than a factor of 1000, which surely explains the absence of the yellow pulse at the output of the absorbing medium.

For the red pulse, the propagation is qualitatively different in character. At a power of about 0.7 W, its input area $\theta\geqslant\pi$. Because the duration of the pulse is much less than T_1 and T_2, we might expect generation of a 2π–pulse in the red transition. However, it is not possible to completely identify the experimentally observed pulse dynamics with self-induced transparency, because the use of a converging beam geometry gives rise to new effects when light interacts with two-level systems. These novel effects are manifestations of a common phenomenon—supertransparency.[10] This phenomenon, which is fundamental in nature, can be said to generalize the effect of self-induced transparency to the case of three spatial measurements. In addition to the well-known dynamics of self-induced transparency solitons, in which atoms of the medium are excited to the upper state and subsequently return their excess energy to the field pulse by stimulated emission, supertransparency also exhibits a new mechanism—dispersive–diffractive stabilization—which further stabilizes the self-induced transparency pulse.

Since our analysis of the red-pulse propagation dynamics is similar in general outline to what we published in our previous paper Ref. 10, it will not be discussed in detail here. We merely note that although the carrier frequency of the red pulse is shifted from the resonance frequency toward longer wavelengths, our value of this shift turns out to be somewhat smaller than that found in the experiments of Ref. 8 since the parameters of the source and focusing of the beams were different in our case.

Now let us return to our discussion of the distinctive features of the interaction of a weak pulse with transition *1–3* in the presence of a strong pulse in the adjacent transition *1–2*, i.e., the three-level *V*-scheme. We mentioned before that the large difference in pulse energies allows us to neglect in first approximation the effect of the weak field on the strong field and regard the interaction between fields as a unilateral effect. Within this approach it is impossible to describe the larger reradiation of energy of the red pulse at the trailing edge compared to the case of independent propagation of the pulses, which is in fact detected experimentally. However, the most important effects, namely the sharp decrease in absorption of the yellow pulse and the shift of its carrier frequency, find their explanation within the framework of the approximations used. The theory set forth above makes it possible to make both qualitative and quantitative estimates of these phenomena.



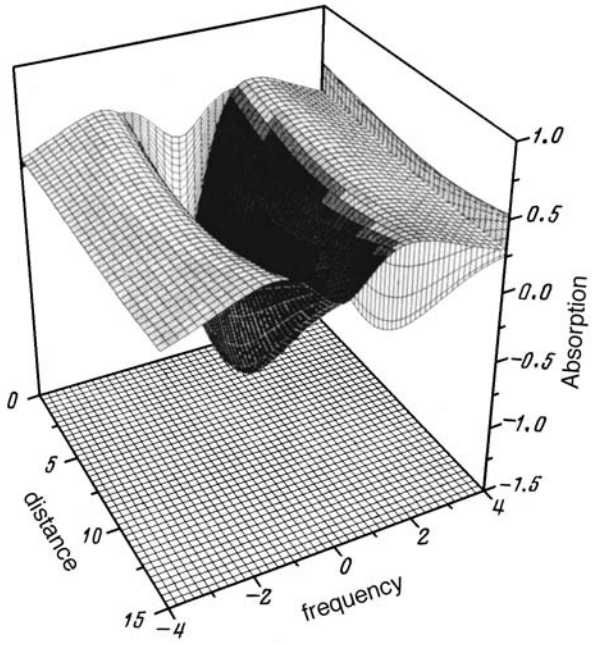


FIG. 5. Absorption coefficient of the weak field at the short-wavelength transition as a function of the normalized frequency $\omega\tau$ and the distance traversed (in units of the Beer absorption length for transition *1–3*) for the following values of the parameters: $T^*=0.35\,ns$, $\tau_3=0.4$ ns, $\tau=0.7$ ns, $t_0=0.5$ ns. Projections of the surface are shown in three mutually perpendicular planes. Individually (above) we present plots of the gain at the central frequency of transition *1–3* as a function of distance traversed by the yellow pulse in the medium.

To explain the most striking result of the experiment, i.e., the fact that a weak pulse can propagate through a dense, strongly absorbing resonant medium, we turn to the expression for the absorption coefficient of the yellow pulse in the presence of a red pulse in a neighboring transition. The latter is calculated using Eq. (14) with the substitution of Eq. (15) for values of the field and medium parameters used in the experiment. The results of this calculation, which are shown in Fig. 5, demonstrate that the absorption coefficient at transition *1–3* depends on the frequency and distance travelled by the weak-field pulse in the absorber. It should be noted

that under optimal conditions this absorption is replaced by amplification in the case of the central spectral component of the yellow pulse, since the amplification coefficient exceeds the absolute value of the linear unperturbed absorption coefficient $k_3=2\kappa_3 T^*$ by a factor of 1.11. A coefficient of field amplification this large cannot be achieved even for total inversion of the *1–3* transition. In other words, the efficiency of the interaction between the fields turns out to be higher than that of the medium with each of the fields taken individually. As we should expect, the red pulse influences the yellow pulse most strongly in the region where the strong field is a maximum, and the affected frequency components of the yellow-pulse spectrum lie within a band whose width is the same as that of the red-pulse spectrum.

The amplification effect under discussion here is purely coherent, and is caused by a two-quantum process in which the interaction is transferred via the polarization $\mathscr{P}_2$, which couples levels *1* and *3* by way of the common lower level. This is easy to verify if we neglect the second term in Eq. (13), saving only the first term arising from changes in the population of the lower level, which is $\propto \dot{\mathscr{N}}_3$. In this case, the maximum achievable effect is complete transparency of the medium at the central spectral component of the yellow pulse. If we add to this effect, which decreases the absorption, the contribution from the polarization $\mathscr{P}_2$, we obtain a net field gain in the *1–3* transition, which occurs over a wide range of frequencies against a background of no inversion between levels *1* and *3* at all times. This leads us to identify the effect as inversionless amplification. However, this does not lead to amplification of the weak pulse as a whole. The amplification effect, which occurs because the interaction involves a common lower level, is not disrupted for times of order T^*, due to the phasing of all the atomic oscillators of the ensemble induced by the strong pulse. The simultaneous phasing of the two polarizations $\langle\mathscr{P}_1\rangle$ and $\langle\mathscr{P}_2\rangle$ for all the three-level atoms of the ensemble follows directly from the theory developed here, and as far as we know has not been discussed previously.

Our fundamental theoretical conclusions are based on the assumption that the shape of the field envelope for transition *1–2* is that of a self-induced transparency soliton. This choice was determined by two factors. The first of these is the tendency for a short high-power pulse to acquire the form of a self-induced transparency soliton as it propagates in an absorber. The second is connected with the fact that a self-induced transparency soliton generates a deep modulation of the population difference between levels *1* and *2*, up to complete emptying of the lower level, which thereby ensures optimal conditions for amplification of the weak pulse. Thus, the self-induced transparency soliton is essentially ideal as a pump pulse.

The conditions for pulse propagation were by no means optimal in our experiment. Because the strong pulse was focused by a lens, which led to expulsion of its carrier frequency from the inhomogeneously broadened absorption line, the delay was decreased, the efficiency of the resonant interaction lowered, and the modulation depth for the population of level *1* less marked. Under these conditions the decrease in absorption, and hence the amplification, become less pronounced than for the ''ideal'' case illustrated in Fig. 5. Nevertheless, the experiments recorded an anomalously low loss of energy by the weak pulse (by a factor of 8 in all), which indirectly confirms the significance of these effects.

In stressing the special role of the self-induced transparency soliton in the dynamics of coherent amplification processes, we must emphasize that this same formalism can be applied to calculate the characteristics of the yellow pulse as it interacts with a red pulse of arbitrary shape,[2)] once this shape is given at any instant in time at any point of the medium. It is only important that the red pulse be strong enough to pass through the optically dense medium without appreciable absorption. The specific shape of the strong-field envelope determines only the degree of coherent amplification and incoherent brightening, not the qualitative dynamic features of the combined propagation of the coherent pulses. Our approach allows us to estimate the optimal optical thickness of absorbing medium that ensures maximum gain for the weak field at its center frequency, the optimum relative position of the pulses at the input to the medium at the center frequency, optimal relations between their durations, etc. Moreover, the theory does not distinguish transitions *1–2* and *1–3*; therefore, propagation of the strong pulse on transition *1–3* and the weak pulse on transition *1–2* in no way affects the dynamics of amplification of the weak pulse by the strong pulse.

6. CONCLUSION

We obtained these results, which are novel both from a theoretical and an experimental point of view, in the course of our investigations of the coherent interaction of light pulses with a three-level medium in the *V*-configuration. We experimentally observed combined propagation of a yellow and red pulse for neighboring transitions in an absorbing medium, and recorded how they affected one another. The most remarkable feature to emerge from this work is the ability of the yellow pulse to pass through the absorbing cell with anomalously small absorption, especially since this pulse is completely absorbed when it propagates ''independently.'' In order to explain this abrupt decrease in the absorption of a pulse generated by the $2p_4$—$1s_5$ transition we invoke a mechanism involving two factors. The first factor is the efficient lowering of the population of the lower $1s_5$ level, which is emptied by the passage of the $2p_6$—$1s_5$ red pulse. The nature of this process leads us to describe it as incoherent brightening of the adjacent transition. The second factor, which is purely coherent in nature, is connected with the transfer of the interaction via polarization $\mathscr{P}_2$, which interferes constructively with the field of the yellow pulse $\mathscr{E}_3$ and leads to enhancement of the spectral components of the weak field over a wide band of frequencies.

We experimentally detected a shift in the carrier frequency of the yellow pulse away from resonance toward longer wavelengths, caused by modulation of the refractive index of the medium in the presence of the strong field of the neighboring transition of the *V*-scheme, whose carrier frequency in turn is redshifted. We have shown theoretically that if the carrier frequency of the red pulse does not deviate from exact resonance with the transition $2p_6$—$1s_5$ as it

propagates, it will not give rise to modulation of the refractive index and there will be no frequency shift of the yellow pulse.

[1)]Under our experimental conditions they differ by approximately a factor of 2.

[2)]Within the approximations used, although by necessity the formalism can be generalized to an arbitrary relation between durations.

[1] O. A. Kocharovskaya and Ya. I. Khanin, JETP Lett. **48**, 630 (1988).

[2] S. E. Harris, Phys. Rev. Lett. **62**, 1033 (1989).

[3] M. O. Scully, S.-Y. Zhu, and A. Gavridiles, Phys. Rev. Lett. **62**, 2813 (1989).

[4] M. J. Konopnicki, P. D. Drummond, and J. H. Eberly, Bull. Am. Phys. Soc. **20**, 1124 (1980); Opt. Commun. **36**, 313 (1981).

[5] L. A. Bol'shov and V. D. Likhanskiĭ, Kvant. Elektron. (Moscow) **15**, 1339 (1985) [Sov. J. Quantum Electron. **12**, 889 (1985)].

[6] V. S. Egorov, V. D. Laptev, N. M. Reutova, and I. V. Sokolov, Kvant. Elektron. (Moscow) **16**, 729 (1986) [Sov. J. Quantum Electron. **13**, 474 (1986)].

[7] S. É. Frish and O. P. Bochkova, Vestn. Leningr. Univ., Fiz., Khim. **16**(13), 40 (1961).

[8] V. S. Egorov, É. E. Fradkin, V. V. Kozlov, and N. M. Reutova, Laser Phys. **2**, 973 (1992).

[9] V. V. Kozlov and É. E. Fradkin, Zh. Éksp. Teor. Fiz. **103**, 1902 (1993) [JETP **76**, 940 (1993)].

[10] V. V. Kozlov, É. E. Fradkin, V. S. Egorov, and N. M. Reutova, Zh. Éksp. Teor. Fiz. **110**, 1688 (1996) [JETP **83**, 927 (1996)].

[11] N. P. Penkin, Opt. Spektrosk. **2**, 545 (1957).

[12] W. R. Bennett Jr. and P. J. Kindlmann, Phys. Rev. **149**, 38 (1966).

[13] S. L. McCall and E. L. Hahn, Phys. Rev. Lett. **18**, 908 (1967).

[14] A. Allen and J. H. Eberly, *Optical Resonance and Two-level Atoms*, Wiley, New York (1975); S. L. McCall and E. L. Hahn, Phys. Rev. **183**, 457 (1969); G. L. Lamb Jr., Rev. Mod. Phys. **43**, 99 (1971); I. A. Poluktov, Yu. M. Popov, and V. S. Roĭtberg, Usp. Fiz. Nauk **114**, 97 (1974) [Sov. Phys. Usp. **17**, 673 (1974)]; A. I. Maimistov, A. M. Basharov, S. O. Elyutin, and Yu. M. Sklyarov, Phys. Rept. C **191**, 1 (1990).

[15] M. D. Crisp, Phys. Rev. A **1**, 1604 (1970).

Translated by Frank J. Crowne

Quantum theory of resonant scattering of atoms by a pulsed field under coherent population trapping conditions

A. V. Taĭchenachev, A. M. Tumaĭkin,*) and V. I. Yudin

Novosibirsk State University, 630090 Novosibirsk, Russia
(Submitted 30 April 1997)
Zh. Éksp. Teor. Fiz. **113**, 89–110 (January 1998)

We develop an entirely quantum mechanical analytical description of scattering of atoms with angular momenta $j_g \to j \to j_e = j$ (j is an integer) by a pulsed $\sigma_+ - \sigma_-$ field. In the stationary-atom approximation with exact accounting for recoil effects, we solve the problem of the change in the distribution of atoms among the internal and translational degrees of freedom initiated by a single pulse for $j_g = 1, 2$. We find in analytical form recurrence formulas that make it possible to calculate the distribution of the atoms after an arbitrary sequence of pulses has acted on the system. We show that for discrete (resonant) values of the time interval between the pulses, the action of N pulses leads to effective formation and narrowing of peaks at discrete points in momentum space and to a broadening of the envelope of these peaks. In the case of a broad initial momentum distribution we derive explicit formulas for the peaks and the envelope and study their asymptotic behavior for $N \gg 1$. Finally, in the weak-field limit we study numerically the dependence of the contrast of the scattering diagram on pulse length. © *1998 American Institute of Physics.* [S1063-7761(98)00701-X]

1. INTRODUCTION

The pioneering works of Aspect *et al.*[1,2] initiated intensive studies of the kinetic manifestations of coherent population trapping (CPT). Today we know of various methods of laser cooling of atoms below the one-photon recoil energy by velocity-selective coherent population trapping in fields with spatial polarization and intensity gradients. The methods have been developed theoretically[2–9] and to a certain extent experimentally.[1,10–12,9] The idea of these cooling methods is as follows. Atoms in a stationary inhomogeneously polarized field are trapped into a CPT state (dark state) $|\Psi_{NC}\rangle$ that does not interact with the field. The state is a coherent superposition of the wave functions of the magnetic sublevels of the ground state with different momentum values (e.g., in a $\sigma_+ - \sigma_-$ field the momentum values are $p, p \pm \hbar k, p \pm 2\hbar k, \cdots$, known collectively as the p-family[2]). When the atoms are in free motion, the components of this superposition acquire different phases. As a result, if the atom was in the CPT state at a certain moment, at subsequent times the CPT state usually disintegrates due to the translational motion of the atoms, and the atoms begin to interact with the field. However, the degree of this disintegration and the intensity of the atom–field interaction are selectively dependent on the momentum p. For instance, the population of the excited state, considered as a function of momentum, has a dip near $p = 0$. Due to random kicks in absorbing and emitting photons, the atoms gather in this region of momentum space, where the interaction with the field is at its minimum. Thus, cooling in this case is a consequence of a special diffusion process.[13–16]

Recently two groups of researchers proposed a new method of pulsed cooling[17] (or Ramsey cooling[18]), which as shown by the results of experiments[18] and quantum simulations[17] makes it possible to obtain narrower structures in the velocity distribution of atoms and in shorter time intervals than in the case of continuous irradiation. Generalizing the discussions in Refs. 17 and 18, we discuss the physics of the laser pulse cooling method. During the time τ (the pulse length of the $\sigma_+ - \sigma_-$ field) the atoms are pumped into the dark state

$$|\Psi_{NC}\rangle = \sum_{\mu_g = -j_g, -j_g+2, \cdots, j_g} \psi_{\mu_g} |p - \hbar k \mu_g, j_g, \mu_g\rangle,$$

where j_g and μ_g are the quantum numbers representing the ground-state angular momentum and its projection. Then, in the course of the time interval T the field is switched off and the atoms move freely. Here, in view of the dephasing of the different components of the superposition, there is a transition from the CPT state to states that can interact with the field:

$$|\Psi_{NC}, T\rangle = \sum_{\mu_g = -j_g, -j_g+2, \cdots, j_g} \psi_{\mu_g} \times \exp\left(-\frac{iT}{\hbar} \frac{(p - \hbar k \mu_g)^2}{2M} \right) \times |p - \hbar k \mu_g, j_g, \mu_g\rangle.$$

This formula shows that if the time interval between pulses is chosen to be $T_n = \pi n / 4\omega_r$ (here $\omega_r = \hbar k^2 / 2M$ is the frequency corresponding to the recoil energy, and n is a nonzero integer), for certain discrete values of momentum ($p_m = 2\hbar k m / n$ for j_g odd and $p_m = (2m/n + 1)\hbar k$ for j_g even) the states $|\Psi_{NC}, T\rangle$ and $|\Psi_{NC}\rangle$ differ only in a general phase factor. Thus, for discrete values of time, T_n, and momentum, p_m, the CPT state is restored. (An exception is $j_g = 1$, when there are no restrictions on the time T, with $p_m = 2\pi\hbar k m / 4\omega_r T$.) If at the end of the time interval T_n a second light pulse is switched on, the atoms with the selected momenta p_m do not interact with the field, while the other atoms scatter with a change in momentum due to spontane-

ous and stimulated recoil effects. As a result the momentum representation acquires a well-defined comblike structure.[17] By using a sequence of light pulses one can increase the contrast of the atomic scattering diagram.[18] Note that the sharp selectivity in atomic momentum in this case is ensured by the large value of T, by analogy with Ramsey resonances.[19] We also note that the transverse motion of the atoms (in relation to the wave vector) leads to the emergence of a common phase factor in the function $|\Psi_{NC},T\rangle$ (the factor contributes nothing to dephasing) and therefore in no way affects our reasoning.

In this paper we develop a quantum theory for atoms with angular momenta $j_g=j\rightarrow j_c=j$ (j is an integer) scattered by a $\sigma_+-\sigma_-$ field. Our goal is to give a complete analytical description of Ramsey cooling. The main approximation that we employ is that the lifetime of the atoms in the dark state, limited by the translational motion effect, is much longer that the pulse length τ. In perturbation-theory terms this condition can be written as $\gamma\tau(kv/\Omega)^2\ll 1$ (here γ is the radiation width of the excited level, kv is the Doppler shift, and Ω is the Rabi frequency),[2,16] which means that either we must use precooled atoms or the laser field must be strong. Moreover, we assume (except in Sec. 5) that the interaction in stationary, i.e., $\gamma\tau\gg 1$ and $\gamma S\tau\gg 1$, with $S=\Omega^2/(\gamma^2/4+\delta^2)$ the saturation parameter and δ the detuning from resonance. In these conditions the density matrix of the atoms after a light pulse has acted on the system has the following form:

$$\hat{\rho}=|\Psi_{NC}\rangle W\langle\Psi_{NC}|.$$

The function W depends on the initial (before the first light pulse has acted on the system) density matrix and can be found exactly (outside the scope of the expansion in recoil momentum) by a method described in Ref. 16. The evolution of the density matrix of the atoms that are in the ground state and propagate freely is determined by the kinetic-energy operator $\hat{H}_K$. The solution of the problem of calculating the corresponding unitary operator $\exp(-i\hat{H}_KT/\hbar)$ is well-known. Applying the above transformations in the appropriate order, we can calculate the atomic distribution after an arbitrary sequence of field pulses.

In the present paper we solve this problem for two values of angular momentum, $j_g=1,2$. We find in analytical form recurrence formulas linking the distribution $W^{(N+1)}$ after $N+1$ pulses have acted on the system with $W^{(N)}$, the distribution after N pulses. We show that a sequence of pulses applied to the system leads to the formation and narrowing of peaks at discrete points in the momentum space and to a broadening of the envelope of these peaks. In the case of a broad (in comparison to the photon momentum) initial momentum distribution, which is important for practical reasons, we derive explicit formulas for the peaks and envelope. We also examine the asymptotic behavior of the solution for $N\gg 1$. Moreover, for weak saturation, $S\ll 1$, we study the dependence of the contrast of the scattering diagram on the parameter $\gamma S\tau$. Finally, we compare our results with those of Refs. 17 and 18 and discover that qualitatively the results coincide.

2. STATEMENT OF THE PROBLEM

Let us examine one-dimensional (along the z axis) motion of atoms whose ground and excited states form an optical transition $j_g=j\rightarrow j_e=j$ (j is an integer) in the interaction between the system of atoms and a pulsed $\sigma_+-\sigma_-$ field. Within the pulse of length τ the field is assumed monochromatic:

$$\mathbf{E}(z,t)=\mathbf{e}(z)E_0\ \exp(-i\omega t)+\text{c.c.},$$
$$\mathbf{e}(z)=\frac{1}{\sqrt{2}}(\mathbf{e}_{-1}\ \exp(ikz)-\mathbf{e}_{+1}\ \exp(-ikz)), \tag{1}$$

where $\mathbf{e}_{\pm 1}=\mp(\mathbf{e}_x\pm i\mathbf{e}_y)/\sqrt{2}$ are unit cyclic vectors. At each point in space the field (1) is linearly polarized. The direction of the polarization vector $\mathbf{e}(z)$ at point $z=0$ coincides with the x axis, and for an arbitrary z the vector is rotated through an angle kz, i.e., the field is a linearly polarized helix. In view of this it is convenient (as shown in Ref. 16) to shift from the laboratory reference frame to the local reference frame in which the x' axis rotates together with $\mathbf{e}(z)$. The corresponding transformation formulas are

$$\hat{O}_{\text{lab}}=\hat{U}(z)\hat{O}_{\text{loc}}\hat{U}^{\dagger}(z),$$
$$\hat{O}_{\text{loc}}=\hat{U}^{\dagger}(z)\hat{O}_{\text{lab}}\hat{U}(z), \tag{2}$$
$$\hat{U}(z)=\exp(-ikz\hat{J}_z),$$

where $\hat{J}_z$ is the operator of the projection of angular momentum, and $\hat{O}_{\text{lab}}$ and $\hat{O}_{\text{loc}}$ are matrices representing an arbitrary operator in the laboratory and local reference frames, respectively. In particular, the Hamiltonian of a free atom in the rotating reference frame is

$$\hat{H}_0=\hat{H}_K+\hbar\omega_0\hat{\Pi}_e, \tag{3}$$

where

$$\hat{H}_K=\frac{(\hat{p}-\hbar k\hat{J}_z)^2}{2M} \tag{4}$$

is the kinetic-energy operator (which now depends on the values of the projection of angular momentum), ω_0 is the transition frequency, and

$$\hat{\Pi}_e=\sum_{\mu_e=-j_e}^{j_e}|j_e,\mu_e\rangle\langle j_e,\mu_e| \tag{5}$$

is the projector operator on the excited state, with $|j_e,\mu_e\rangle$ the wave vectors of the magnetic sublevels. In the local reference frame the Hamiltonian (1) of the resonant atom–field interaction is spatially homogeneous:

$$\hat{H}_{A-F}=\hbar\Omega\hat{V}\ \exp(-i\omega t)+\text{H.c.}, \tag{6}$$

where Ω is the Rabi frequency (which without loss of generality can be assumed positive), and the dimensionless operator $\hat{V}$ is defined in terms of Clebsch–Gordan coefficients as follows (the quantization axis is directed along the z axis):

$$\hat{V}=\frac{\hat{V}_{-1}-\hat{V}_{+1}}{\sqrt{2}}, \tag{7}$$

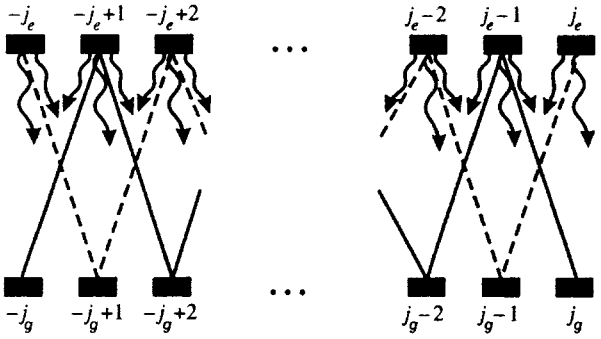


FIG. 1. Diagrams representing the radiative transitions between the Zeeman sublevels of the ground and excited states. The solid and dashed lines represent stimulated transitions between the sublevels of the Λ-system, and between the sublevels of the V-system, respectively. The wavy arrows depict spontaneous transitions.

$$\hat{V}_q=\sum_{\mu_e,\mu_g}|j_e,\mu_e\rangle\langle j_e(j_g,1),\mu_e|j_g,\mu_g;1,q\rangle\langle j_g,\mu_g|. \quad (8)$$

The transitions induced by the field (1) are depicted in Fig. 1. We see that there are two independent systems of interacting sublevels.[20] One consists of Λ-segments and begins with the $|j_g,-j_g\rangle$ sublevel. The other consists of V-segments and begins with the $|j_e,-j_e\rangle$ sublevel. We call these systems the Λ- and V-system, respectively.

Separating in the usual way the fast time dependence at the frequency of the field, we obtain a quantum kinetic equation that describes the evolution of the slow components of the density matrix in the rotating reference frame:

$$\begin{aligned}\frac{\partial}{\partial t}\hat{\rho}(z_1,z_2)=&-\frac{i}{\hbar}[\hat{H}_K,\hat{\rho}(z_1,z_2)]\\&-i\Omega[(\hat{V}+\hat{V}^\dagger),\hat{\rho}(z_1,z_2)]-\left(\left(\frac{\gamma}{2}-i\delta\right)\right.\\&\left.\times\hat{\Pi}_e\hat{\rho}(z_1,z_2)+\left(\frac{\gamma}{2}+i\delta\right)\hat{\rho}(z_1,z_2)\hat{\Pi}_e\right)\\&+\gamma\sum_{q=\pm1,0}Q_q(k(z_1-z_2))\hat{V}_q^\dagger\hat{\rho}(z_1,z_2)\hat{V}_q,\end{aligned} \quad (9)$$

where $\delta=\omega-\omega_0$ is the detuning from resonance, and the functions $Q_q(k(z_1-z_2))$ describe the stimulated and spontaneous recoil effects:

$$\begin{aligned}Q_{\pm1}(kz)&=\frac{3}{2}\left(\frac{\sin kz}{kz}+\frac{\cos kz}{(kz)^2}-\frac{\sin kz}{(kz)^3}\right)\exp(\mp ikz),\\Q_0(kz)&=3\left(-\frac{\cos kz}{(kz)^2}+\frac{\sin kz}{(kz)^3}\right).\end{aligned} \quad (10)$$

In (9) we used the standard notation for commutators, $[\hat{A},\hat{B}]$. Equation (9) accounts exactly for quantum effects due to momentum transfer from field to atoms in radiative processes and to the translational motion of atoms. Note that the convenience of using the coordinate representation for translational degrees of freedom is due in this case to the fact that the semiclassical expansion in powers of $k(z_1-z_2)$ is invalid.

A remark is in order concerning the necessity to justify the model of one-dimensional motion (often used in problems dealing with the mechanical action of a plane electromagnetic wave on atoms; see, e.g., Ref. 22). Obviously, due to the interaction with the vacuum modes of the field, the Cartesian coordinates in the equations of motion generally cannot be separated. However, these correlations, which are due to the spontaneous recoil effect, are small if the variation of the transverse momentum in an elementary emission act is small. This case, when the transverse kinetic energy is much higher than the one-photon recoil energy ($k_BT_\perp\gg\hbar\omega_r$), is studied in the present paper.

3. SOLUTION FOR IMMOBILE ATOMS

If while a light pulse (of length τ) acts the atoms are, on the average, displaced by a distance much shorter than the wavelength of the light,

$$v\tau\ll\lambda, \quad (11)$$

they can be assumed to be immobile, so that the first term on the right-hand side of Eq. (9) can be dropped. As a result we arrive at a system of first-order ordinary differential equations $\partial\rho/\partial t=\mathscr{L}\rho$, whose solution can be written as

$$\rho_{\mu_1,\mu_2}(z_1,z_2|t+\tau)=\sum_{\nu_1,\nu_2}\mathscr{R}^{\nu_1,\nu_2}_{\mu_1,\mu_2}(z_1,z_2|\tau)\rho_{\nu_1,\nu_2}(z_1,z_2|t), \quad (12)$$

where $\mathscr{R}$ is the matrix exponential function of the corresponding Liouville operator, $\mathscr{R}(\tau)=\exp(\tau\mathscr{L})$. For large values of τ, i.e.,

$$\gamma\tau\gg1,\quad \gamma S\tau\gg1, \quad (13)$$

where $S=\Omega^2/(\gamma^2/4+\delta^2)$ is the saturation parameter, the atoms are pumped completely into the dark state, with the corresponding steady-state solution (the limit $\tau\to\infty$ in (12))

$$\hat{\rho}(z_1,z_2)=|\Psi_{NC}\rangle W(z_1,z_2)\langle\Psi_{NC}|. \quad (14)$$

The atom–field interaction operator annihilates the CPT state $|\Psi_{NC}\rangle$

$$\hat{H}_{A-F}|\Psi_{NC}\rangle=0, \quad (15)$$

which is a superposition of ground-state Zeeman wave functions:

$$|\Psi_{NC}\rangle=\sum_{\mu_g}\psi_{\mu_g}|j_g,\mu_g\rangle. \quad (16)$$

The simplest way to find the coefficients ψ_{μ_g} is to direct the quantization axis along the field polarization vector. Then the dark state coincides with the Zeeman sublevel $|j_g,0\rangle$. If we then carry out a rotation through $\pi/2$ about the y axis (thus returning to the original choice of the quantization axis z), we find that ψ_{μ_g} can be expressed in terms of Wigner d-functions (elements of the rotation matrix):[21]

$$\psi_{\mu_g}=d^{j_g}_{\mu_g,0}(\pi/2). \quad (17)$$

The function $W(z_1,z_2)$ has the meaning of a two-point distribution function in the local reference frame. In our case the relationship between W and the ordinary distribution function F in the laboratory reference frame is given by the following expressions:

$$F(z_1,z_2)=\mathrm{Tr}\{\hat{U}(z_1)\hat{\rho}(z_1,z_2)\hat{U}^{\dagger}(z_2)\}=\langle\Psi_{NC}| \times\exp(-ik(z_1-z_2)\hat{J}_z)|\Psi_{NC}\rangle W(z_1,z_2), \tag{18}$$

where the trace is over the internal atomic degrees of freedom. The value of W is determined by momentum transfer in spontaneous and stimulated photon scattering in the course of a field pulse and by the initial (prior to the light pulse) distribution over the internal and translational degrees of freedom. According to Ref. 16, after the pulse has acted, we have

$$W(z_1,z_2|t+\tau)=\mathrm{Tr}\{\hat{C}(z_1-z_2)\hat{\rho}(z_1,z_2|t)\}, \tag{19}$$

where the matrix $\hat{C}(z_1-z_2)$ is the left eigenvector of $\mathscr{L}$ corresponding to a zero eigenvalue:

$$0=i\Omega[(\hat{V}+\hat{V}^{\dagger}),\hat{C}(z)]-((\gamma/2)+i\delta)\hat{\Pi}_e\hat{C}(z)+(\gamma/2-i\delta)\hat{C}(z)\hat{\Pi}_e)+\gamma\sum_{q=\pm1,0}Q_q(kz)\hat{V}_q\hat{C}(z)\hat{V}_q^{\dagger}. \tag{20}$$

We select the normalization condition for $\hat{C}$ in the form

$$\langle\Psi_{NC}|\hat{C}(z)|\Psi_{NC}\rangle=1. \tag{21}$$

If we ignore the translational motion of the atoms, $\mathrm{Tr}\{\hat{C}(z_1-z_2)\hat{\rho}(z_1,z_2|t)\}$ is a constant of motion for (9). Calculating its value before and after the pulse with allowance for (14) and (21), we arrive at (19). Equations (20) and (21) lead to four fundamental properties of the matrix $\hat{C}(z)$, which are valid for all integral values of angular momentum j:

(1) $C_{\mu_1,\mu_2}(0)=\delta_{\mu_1,\mu_2}$, which corresponds to conservation of the total population of the atomic sublevels in optical pumping.

(2) Asymptotically (for $|kz|\gg1$) the function $Q_q(kz)$ specified in (10) vanishes and the solution of (20) assumes the form $\hat{C}(\infty)=|\Psi_{NC}\rangle\langle\Psi_{NC}|$. This property allows, at least in principle, for correlations between arbitrarily distant points z_1 and z_2.

(3) The property $\hat{C}^{\dagger}(z)=\hat{C}(-z)$ ensures that the momentum distribution of the atoms is positive definite.

(4) In contrast to the above three properties, the fourth property $C_{\mu_1,\mu_2}(z)=C_{-\mu_1,-\mu_2}(-z)$ is specific to the configuration of the $\sigma_+-\sigma_-$ field considered here and manifests itself in the symmetry of the scattering diagram with respect to zero.

In our problem, prior to the field pulse the atoms are in the ground state, so that to find the function W we only need to know the matrix elements of $\hat{C}$ between the wave functions of the magnetic sublevels of the ground level, $|j_g,\mu_g\rangle$. Since the Λ- and V-systems of magnetic sublevels are independent (there is no coherence between them), this matrix $\hat{C}^{gg}$ can be represented in block-diagonal form:

$$\hat{C}^{gg}=|\Psi_{NC}\rangle\langle\Psi_{NC}|+\sum_{i,i'=1}^{j_g}|\Psi_i^{\Lambda}\rangle C_{i,i'}^{\Lambda}\langle\Psi_{i'}^{\Lambda}| +\sum_{i,i'=1}^{j_g}|\Psi_i^{\mathrm{V}}\rangle C_{i,i'}^{\mathrm{V}}\langle\Psi_{i'}^{\mathrm{V}}|, \tag{22}$$

where

$$|\Psi_i^{\Lambda}\rangle=\sum_{\mu_g}\frac{1}{\sqrt{2}}(d^{j_g}_{\mu_g,-i}(\pi/2)+d^{j_g}_{\mu_g,i}(\pi/2))|j_g,\mu_g\rangle \tag{23}$$

and

$$|\Psi_i^{\mathrm{V}}\rangle=\sum_{\mu_g}\frac{1}{\sqrt{2}}(d^{j_g}_{\mu_g,-i}(\pi/2)-d^{j_g}_{\mu_g,i}(\pi/2))|j_g,\mu_g\rangle \tag{24}$$

are the eigenvectors of the operator $\hat{V}^{\dagger}\hat{V}$ with eigenvalues

$$\alpha_i=\frac{i^2}{j_g(j_g+1)},\quad i=1,\cdots,j_g. \tag{25}$$

The method used to determine the coefficients in (23) and (24) is similar to that used in deriving (17). The superscripts Λ and V indicate that the vector belongs either to the Λ-system or to the V-system of the Zeeman sublevels (see Fig. 1). For instance, in (23) μ_g runs through the values $-j_g,-j_g+2,\cdots,j_g$ and in (24), through the values $-j_g+1,-j_g+3,\cdots,j_g-1$.

We found the explicit form of the matrices $\hat{C}^{\Lambda}$ and $\hat{C}^{\mathrm{V}}$ for two transitions, $j_g=1,2$.

3.1. The $j_g=1\rightarrow j_e=1$ transition

In this case

$$C^{\Lambda}=\frac{Q_1+Q_{-1}}{4-Q_1-Q_{-1}},\quad C^{\mathrm{V}}=\frac{2Q_0}{4-Q_1-Q_{-1}} \tag{26}$$

are independent of detuning and saturation, which is a specific feature of this transition.

3.2. The $j_g=2\rightarrow j_e=2$ transition

Here $\hat{C}^{\Lambda}$ and $\hat{C}^{\mathrm{V}}$ can be represented as the symmetric and antisymmetric (with respect to Q_0) parts,

$$\hat{C}^{\Lambda}=\hat{M}(Q_0)+\hat{M}(-Q_0),\quad \hat{C}^{\mathrm{V}}=\hat{M}(Q_0)-\hat{M}(-Q_0), \tag{27}$$

of the matrix

$$\hat{M}(Q_0)=\frac{2Q_0+Q_{-1}+Q_1}{2D}\begin{pmatrix}m_{1,1} & m_{1,2}\\ m_{2,1} & m_{2,2}\end{pmatrix}. \tag{28}$$

The explicit expressions for the coefficients D and m_{ij} are given in the Appendix.

As Eqs. (A1)–(A5) show, the dependence on detuning and field amplitude is due to the finite off-diagonal elements $m_{1,2}$ and $m_{2,1}$. This coherence is induced by the recoil effect and is proportional to the difference $Q_{-1}-Q_1$, which van-

ishes at $kz=n\pi$ (see Eqs. (10)). Assuming that $Q_{-1}=Q_1=Q$, we arrive at a reduced matrix, which is independent of δ and Ω:

$$\hat{M}_{\rm red}=\frac{3(Q_0+Q)}{36-Q_0^2-2Q_0Q-30Q+3Q^2}\times\begin{pmatrix}3-2Q & 0\\ 0 & (Q_0+Q)/2\end{pmatrix}. \quad (29)$$

On the other hand, as Eqs. (20) and (28) show, the coherence between the states (23) and (24) can be ignored in two limiting cases: a strong laser field $\Omega\gg\gamma$, and a large detuning $|\delta|\gg\gamma$. In both cases (28) becomes (29) with $Q=(Q_{-1}+Q_1)/2$.

Let us now discuss the range of applicability of solution (14). In the local reference frame the translational motion of atoms in described by the operator $\hat{H}_K$ given by (4). The diagonal element

$$\langle\Psi_{NC}|\hat{H}_K|\Psi_{NC}\rangle=\frac{\hat{p}^2}{2M}+\frac{\hbar\omega_r}{2}j_g(j_g+1) \quad (30)$$

has the meaning of the effective energy of an atom in the dark state. The second term on the right-hand side, the correction to the kinetic energy due to the inhomogeneity in the field's polarization, is on the order of the recoil effect and is independent of coordinates. Thus, in the homogeneous case considered here the diagonal element (30) contributes nothing to the dynamics of the atomic ensemble. The finite off-diagonal elements,

$$\langle\Psi_1^\Lambda|\hat{H}_K|\Psi_{NC}\rangle=-\frac{\hbar k\hat{p}}{2M}\sqrt{j_g(j_g+1)},$$

$$\langle\Psi_2^\Lambda|\hat{H}_K|\Psi_{NC}\rangle=\frac{\hbar\omega_r}{4}\sqrt{j_g(j_g+1)(j_g-1)(j_g+2)}, \quad (31)$$

describe nonadiabatic states, which lead to a finite lifetime of the atoms in the CPT state. If we employ the perturbation theory in the atomic velocity $kv\ll\min(\gamma,\gamma S)$, this lifetime can be estimated at $\gamma^{-1}(\Omega/kv)^2$ (see Refs. 2 and 16). The corrections due to translational motion are negligible if this lifetime is much longer than the pulse length τ, i.e., if $(\Omega/kv)^2\gg\gamma\tau$. Summing up all the restrictions, we can write

$$\min(\gamma,\gamma S)\tau\gg1,\quad kv\ll\min(\gamma,\gamma S),\quad \gamma\tau\left(\frac{kv}{\Omega}\right)^2\ll1. \quad (32)$$

These conditions are weaker than (11) and can be met with precooled atoms or a strong laser field.

4. SOLUTION FOR FREE PROPAGATION OF ATOMS; THE ACTION OF A SEQUENCE OF LIGHT PULSES

After the field is switched off, the atoms are in the ground state, with the result that their evolution in free propagation is determined by the kinetic-energy operator (4):

$$\hat{\rho}(z_1,z_2|t+T)=\exp\left(-\frac{i}{\hbar}\hat{H}_KT\right)|\Psi_{NC}\rangle W(z_1,z_2|t)\langle\Psi_{NC}|\times\exp\left(\frac{i}{\hbar}\hat{H}_KT\right). \quad (33)$$

By combining (33) and (19) we can derive a recurrence formula that relates the distribution after $(N+1)$ pulses, $W^{(N+1)}$, to $W^{(N)}$, the distribution after N pulses:

$$W^{(N+1)}(r,z)=\sum_{\mu_g,\nu_g}C_{\mu_g,\nu_g}(z)\psi_{\nu_g}\psi_{\mu_g}\times\exp[-i\omega_rT(\nu_g^2-\mu_g^2)]\times\exp\left[\omega_rT\left(\frac{2i}{k}\frac{\partial}{\partial z}+\nu_g+\mu_g\right)\frac{\partial}{k\partial r}\right]\times\exp\left[\frac{2\omega_rT(\nu_g-\mu_g)}{k}\frac{\partial}{\partial z}\right]W^{(N)}(r,z), \quad (34)$$

where we have introduced the variables $r=(z_1+z_2)/2$ and $z=z_1-z_2$. The reader will recall that the Fourier transform in the difference z is a Wigner distribution function in the phase space:

$$\widetilde{W}(r,p)=\frac{1}{2\pi\hbar}\int_{-\infty}^{\infty}\exp\left(-\frac{ipz}{\hbar}\right)W(r,z)dz.$$

In this paper we do not account for spatial localization effects. We assume that the distribution is homogeneous, $W(r,z)=W(z)$, with the result that (34) becomes

$$W^{(N+1)}(z)=\sum_{\mu_g,\nu_g}C_{\mu_g,\nu_g}(z)\psi_{\nu_g}\psi_{\mu_g}\times\exp[-i\omega_rT(\nu_g^2-\mu_g^2)]\times W^{(N)}(z+2\omega_rT(\nu_g-\mu_g)/k). \quad (35)$$

We define the initial conditions for the recurrence formula (35) as follows. Suppose that before the first pulse has acted on the atomic system the atoms are in the ground state and have an isotropic distribution over the magnetic sublevels:

$$\rho^{(0)}_{\mu_g,\nu_g}(z)=\frac{\delta_{\mu_g,\nu_g}}{2j_g+1}F^{(0)}(z),$$

where $F^{(0)}(z)$ is the initial distribution in the laboratory reference frame. After the first pulse has acted, we get

$$W^{(1)}(z)=\mathrm{Tr}\{\hat{C}^{gg}(z)\exp(ikz\hat{J}_z)\}\frac{F^{(0)}(z)}{2j_g+1}. \quad (36)$$

4.1. The case of a broad initial momentum distribution

The natural scale of length in (35) and (36) is the wavelength $\lambda=2\pi/k$. If the spread in momentum in the initial distribution is considerably greater than the photon momentum, the function $F^{(0)}(z)$ is finite in a small (in comparison to λ) neighborhood of $z=0$. In the zeroth approximation we can approximate it by a ''unit delta function'':

$$F^{(0)}(z)=\begin{cases}1 & \text{if } z=0,\\ 0 & \text{if } z\neq0.\end{cases} \quad (37)$$

In the same approximation, after N pulses have acted on the system the distribution function constitutes a regular system of peaks at the point $0, \pm 4\omega_r T/k, \pm 8\omega_r T/k, \cdots$:

$$W^{(N)}(z)=\sum_l \phi_l^{(N)} F^{(0)}(z-4\omega_r T l/k). \tag{38}$$

The amplitudes of the peaks, $\phi_l^{(N)}$, satisfy the recurrence formula

$$\phi_l^{(N+1)}=\sum_{\mu_g,\nu_g} C_{\mu_g,\nu_g}(4\omega_r T l/k)\psi_{\nu_g}\psi_{\mu_g} \times \exp[-i\omega_r T(\nu_g^2-\mu_g^2)]\phi^{(N)}_{l+(\nu_g-\mu_g)/2} \tag{39}$$

with the initial condition

$$\phi_l^{(1)}=\delta_{l,0}.$$

In view of the symmetry of $\hat{C}$ (see Sec. 3), the coefficients $\phi_l^{(N)}$ are real and symmetric: $\phi_l^{(N)}=\phi_{-l}^{(N)}$. The momentum representation corresponding to (38) has the form of the product,

$$\widetilde{W}^{(N)}(p)=\Phi^{(N)}(p)\widetilde{F}^{(0)}(p), \tag{40}$$

of the periodic (with a period $2\pi\hbar k/4\omega_r T$) and symmetric (with respect to $p=0$) function

$$\Phi^{(N)}(p)=\sum_{l=-j_g(N-1)}^{j_g(N-1)} \exp\left(-i4\omega_r T l\frac{p}{\hbar k}\right)\phi_l^{(N)} \tag{41}$$

and a smooth envelope, which in the present approximation coincides with the initial momentum representation.

Let us examine the process of formation of a comblike structure in the momentum space qualitatively using an approximate expression for the matrix $\hat{C}$:

$$C_{\mu_g,\nu_g}(4\omega_r T l/k)=\begin{cases}\delta_{\mu_g,\nu_g}, & \text{if } l=0,\\ C_{\mu_g,\nu_g}(\infty)=\psi_{\mu_g}\psi_{\nu_g}, & \text{if } l\neq 0,\end{cases}$$

which is valid for large values of the time interval between pulses, $\omega_r T\gg 1$. In this case (39) becomes

$$\begin{aligned}&\phi_0^{(N+1)}=1,\\ &\phi_{l\neq 0}^{(N+1)}=\frac{1}{2}\,\phi_l^{(N)}+\frac{1}{4}\,(\phi_{l+1}^{(N)}+\phi_{l-1}^{(N)})\end{aligned} \tag{42}$$

for $j_g=1$ and

$$\begin{aligned}&\phi_0^{(N+1)}=1,\\ &\phi_{l\neq 0}^{(N+1)}=\frac{11}{32}\,\phi_l^{(N)}+\frac{3}{16}\cos(4\omega_r T)(\phi_{l+1}^{(N)}+\phi_{l-1}^{(N)})\\ &\qquad\qquad +\frac{9}{64}\,(\phi_{l+2}^{(N)}+\phi_{l-2}^{(N)})\end{aligned} \tag{43}$$

for $j_g=2$. Within the interval $[-\pi\hbar k/4\omega_r T, \pi\hbar k/4\omega_r T]$ the function (41) describes the formation of the principal maximum at the point where all the harmonics interfere constructively, amplifying each other. As Eqs. (42) imply, at $j_g=1$ the amplitudes of all the harmonics are positive. Hence the principal maximum is at the point $p=0$ (see Fig. 2a). When $j_g=2$, the amplitudes of the even harmonics are positive, and the sign of the odd harmonics coincides with that of $\cos 4\omega_r T$. For $\cos 4\omega_r T$ positive the principal maximum is at point $p=0$ (see Fig. 2b), while for $\cos 4\omega_r T$ negative the principal maximum is at point $p=\pi\hbar k/4\omega_r T$ (see Fig. 2c). Moreover, in contrast to the case $j_g=1$, at moderate values of N an additional maximum manifests itself at point $p=\pi\hbar k/4\omega_r T$ if $\cos 4\omega_r T$ is positive or at point $p=0$ if $\cos 4\omega_r T$ is negative (see Figs. 2b and c). In the additional maximum the even and odd harmonics interfere destructively, so that the maximum disappears as the number of pulses increases. The case $\cos 4\omega_r T=0$ requires special treatment. Here the period of $\Phi^{(N)}(p)$ is $\pi\hbar k/4\omega_r T$ and the principal maximum (within the periodicity interval) is at point $p=0$ (see Fig. 2d). The number of harmonics in (41) and their amplitudes increase with the number of pulses. This, obviously, leads to an increase in the height of the principal maximum of the function $\Phi^{(N)}(p)$ and a decrease in the width (see Fig. 3).

Generally speaking, the comblike structure in the momentum distribution is formed for arbitrary values of the time lag T. However, for $j_g=2$ the effectiveness of this process reaches its maximum at $|\cos 4\omega_r T|=1$, i.e., at the resonant values

$$T_n=\frac{\pi n}{4\omega_r}. \tag{44}$$

Equations (42) and (43) yield values of $\phi_l^{(N)}$ that differ somewhat from (39). Nevertheless, the main features of peak formation are reflected correctly.

Another important fact should be mentioned. At resonant values (44) of the time interval between pulses in a sequence we have $Q_{-1}(4\omega_r T_n)=Q_1(4\omega_r T_n)$. In this case (see (29)) the matrix $\hat{C}^{gg}$ and hence the amplitudes $\phi_l^{(N)}$ are independent of detuning and saturation. In other words, when the initial momentum distribution is broad, at resonant values of the time lag the formation of the comblike structure is independent on the field parameters δ and Ω (provided, of course, that the conditions (32) are met).

When we used the approximation (37), we entirely ignored the variation of the envelope of the peaks in the momentum representation. Now we take this fact into account by writing, instead of (38),

$$W^{(N)}(z)=\sum_l \phi_l^{(N)}\mathscr{E}_l^{(N)}(z-4\omega_r T l/k), \tag{45}$$

where the functions $\mathscr{E}_l^{(N)}$ are equal to unity at zero, are finite in a small neighborhood of $z=0$, and describe the variation of the envelope at each step. The amplitudes $\phi_l^{(N)}$ still satisfy (39). Let us examine the evolution of $\mathscr{E}_0^{(N)}(0)=1$, which is equivalent to the envelope of the momentum distribution as a whole,

$$\mathscr{E}_0^{(N+1)}(z)=\sum_{\mu_g,\nu_g} C_{\mu_g,\nu_g}(z)\psi_{\nu_g}\psi_{\mu_g}\exp[i\omega_r T(\nu_g^2-\mu_g^2)] \times \phi^{(N)}_{(\nu_g-\mu_g)/2}\mathscr{E}^{(N)}_{(\nu_g-\mu_g)/2}(z). \tag{46}$$

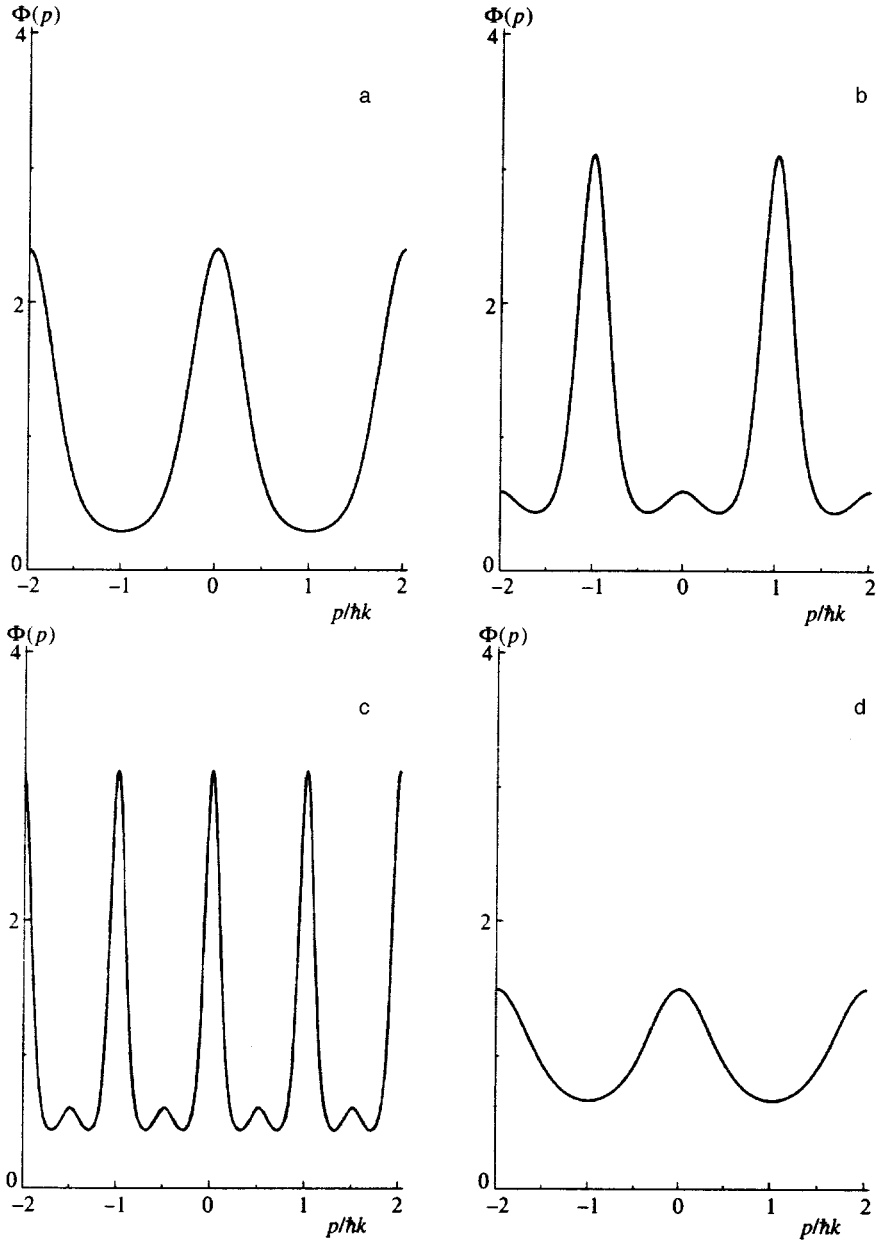


FIG. 2. Formation of a comblike structure in the velocity distribution of atoms after five light pulses have acted on the system. The periodic function $\Phi(p)$ of (41) is depicted for the following four cases: (a) $j_g=1$ and $4\omega_r T=\pi$; (b) $j_g=2$ and $4\omega_r T=\pi$; (c) $j_g=2$ and $4\omega_r T=2\pi$; (d) $j_g=2$ and $4\omega_r T=\pi/2$.

The right-hand side of this equation contains the functions $\mathscr{E}_l^{(N)}(z)$ with $l\neq 0$. In the limit $4\omega_r T\gg 1$, in (46) we can approximate $\mathscr{E}_{l\neq 0}^{(N)}(z)$ by $\mathscr{E}_0^{(N-1)}(z)$. Then for $\mathscr{E}^{(N)}(z)\equiv\mathscr{E}_0^{(N)}(z)$ to within $(kz)^2$ we have

$$\mathscr{E}^{(N+1)}(z)-\mathscr{E}^{(N)}(z)=(-kz)^2 D^{(N)}\mathscr{E}^{(N)}(z) \tag{47}$$

with the initial condition (36):

$$\mathscr{E}^{(1)}(z)=W^{(1)}(z).$$

The first-order terms in kz vanish in view of the symmetry $\phi_l^{(N)}=\phi_{-l}^{(N)}$. The ''diffusion'' coefficient

$$D^{(N)}=-\frac{1}{2}\sum_{\mu_g,\nu_g} C''_{\mu_g,\nu_g}(0)\psi_{\nu_g}\psi_{\mu_g} \times\exp[-i\omega_r T(\nu_g^2-\mu_g^2)]\phi^{(N)}_{(\nu_g-\mu_2)/2} \tag{48}$$

depends on the behavior of $\hat{C}(z)$ near zero and the values of $\phi_l^{(N)}$. Calculating the second derivative of $\hat{C}(z)$ at zero, we establish explicitly that

$$D^{(N)}=\frac{7(1-\phi_1^{(N)})}{10} \tag{49}$$

for $j_g=1$ and

$$D^{(N)}=\frac{9(347-260\cos(4\omega_r T)\phi_1^{(N)}-87\phi_2^{(N)})}{1600} -\frac{2(11+9(\Omega/\gamma)^2)(1-\cos(4\omega_r T)\phi_1^{(N)})}{|11+9(\Omega/\gamma)^2+18i\delta/\gamma|^2} \tag{50}$$

for $j_g=2$.

The solution of Eq. (47) can be represented in the form of a finite product:

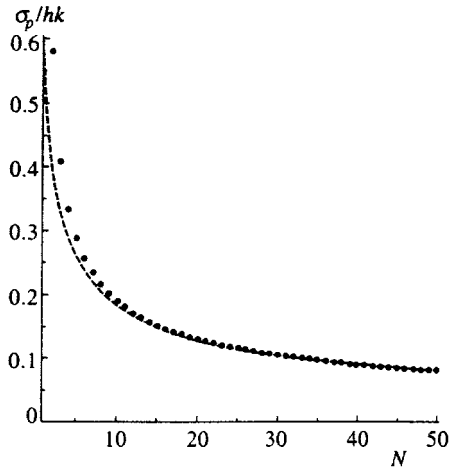


FIG. 3. Dependence of the halfwidth of the peak in the momentum representation on the number of light pulses at $j_g=1$ and $4\omega_r T=\pi$. The dots correspond to the results of calculations by Eqs. (39) and (41), and the dashed curve represents the asymptotic curve $0.58/\sqrt{N}$.

$$\mathscr{E}^{(N)}(z)=\prod_{i=0}^{N-1}(1-(kz)^2D^{(i)})F^{(0)}(z), \tag{51}$$

where $D^{(i)}$ for i positive is determined by Eqs. (49) and (50), and expanding (36) in powers of kz, we arrive at the following expressions for $D^{(0)}$:

$$D^{(0)}=\frac{11}{30} \quad \text{at } j_g=1,$$

$$D^{(0)}=\frac{17}{5}-\frac{16(11+9(\Omega/\gamma)^2)}{15|11+9(\Omega/\gamma)^2+18i\delta/\gamma|^2} \quad \text{at} \quad j_g=2. \tag{52}$$

Obviously, Eq. (51) describes the broadening of the envelope of the peaks in the momentum distribution, with the number of peaks increasing and the fraction of atoms within each peak decreasing in the process. When $j_g=2$ holds, the coefficient D depends on the field parameters δ and Ω and reaches its minimum at $\delta=0$ and $\Omega\ll\gamma$ (here, in view of (32), Ω must be much larger than kv). Thus, the envelope broadening effect can be diminished by a quantity of order of a few percent by using a weak resonant field.

The final expression for the momentum distribution function in the local reference frame in the case of a broad initial momentum distribution is

$$\widetilde{W}^{(N)}(p)=\sum_{l=-j_g(N-1)}^{j_g(N-1)}\exp\left(-i4\omega_r Tl\frac{p}{\hbar k}\right)\phi_l^{(N)}\widetilde{\mathscr{E}}_l^{(N)}(p), \tag{53}$$

where $\widetilde{\mathscr{E}}_l^{(N)}(p)$ is the Fourier transform of $\mathscr{E}_l^{(N)}(z)$. The observed distribution function (in the laboratory reference frame) is expressed in terms of $\widetilde{W}^{(N)}(p)$ by the following formulas:

$$\widetilde{F}^{(N)}(p)=\frac{\widetilde{W}^{(N)}(p+\hbar k)+\widetilde{W}^{(N)}(p-\hbar k)}{2} \quad \text{for } j_g=1, \tag{54}$$

$$\widetilde{F}^{(N)}(p)=\frac{3\widetilde{W}^{(N)}(p+2\hbar k)+2\widetilde{W}^{(N)}(p)+3\widetilde{W}^{(N)}(p-2\hbar k)}{8}$$

$$\text{for } j_g=2, \tag{55}$$

which follow from (18) and describe the splitting of each peak in the local reference frame into j_g+1 peaks in the laboratory reference frame.

4.2. Asymptotic behavior for $N\gg1$

The approximate Eqs. (42) and (43) make it possible to analyze the asymptotic behavior of the solution when the number of pulses is large. For $N\gg1$ the dependence of $|\phi_l^{(N)}|$ on N and l can be approximated by a smooth function $\phi(N,l)$ and Eqs. (42) and (43), by the second-order differential equation

$$\frac{\partial}{\partial N}\phi(N,l)=a\frac{\partial^2}{\partial l^2}\phi(N,l) \tag{56}$$

with the boundary and initial conditions

$$\phi(N,0)=1, \quad \phi(0,l)=0.$$

Thus, the problem reduces to a heat equation for a semibounded rod whose end is kept at a constant temperature. The solution of this problem has the form

$$\phi(N,l)=1-\mathrm{Erf}\left(\frac{l}{2\sqrt{aN}}\right), \tag{57}$$

where the ''thermal conductivity'' coefficient a is equal to $\frac{1}{4}$ at $j_g=1$. For $j_g=2$ the situation is more complicated. Formally at $|\cos 4\omega_r T|=1$ the ''thermal conductivity'' a is equal to $\frac{15}{32}$, but the first coefficients $|\phi_1^{(N)}|$ differs considerably from (57) because of the transition over point $l=0$, a process that cannot be described by Eq. (56). For this coefficient the asymptotic behavior is

$$|\phi_1^{(N)}|\approx 1-\frac{1.3}{\sqrt{N}}.$$

For $l>1$ formula (57) becomes valid if we interpret a as a fitting parameter close to $\frac{15}{32}$ and weakly dependent on N.

Equation (57) shows that the width of the peaks in the momentum representation decreases like $1/\sqrt{N}$ (see Fig. 3). The asymptotic behavior of the ''diffusion'' coefficient $D^{(N)}$ in Eq. (47) for the envelope of the peaks also obeys the $1/\sqrt{N}$ law. Hence the width of the envelope increases like $N^{1/4}$, while the relative fraction of atoms within each peak (the area of the peak) decreases like $N^{-1/4}$.

Interestingly, a similar asymptotic behavior for the width and area of the peaks is observed in the problem of cooling by velocity-selective coherent population trapping in a stationary $\sigma_+-\sigma_-$ field[13] if N is interpreted as the time of interaction with the field.

4.3. Scattering in the general case

Let us examine the general case where the de Broglie wavelength λ_{DB} of the atom (the characteristic length over

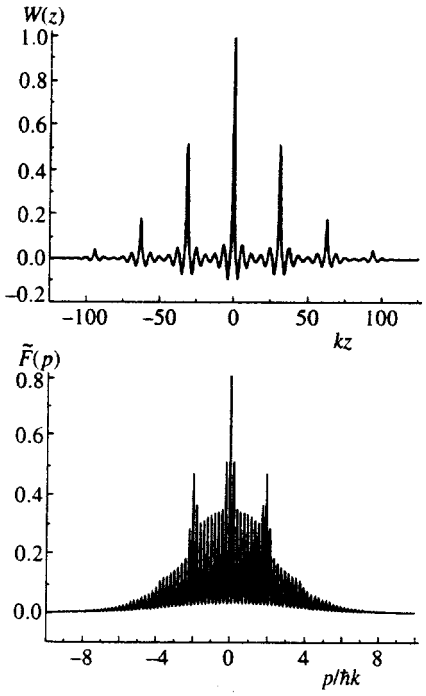


FIG. 4. Diagram of atomic scattering after five light pulses have acted on the system at $j_g=1$ and $4\omega_r T=10\pi$ in the case of a narrow initial momentum distribution, $\tilde{F}^{(0)}(p)\propto\exp(-p^2/p_0^2)$, with $p_0=\hbar k/5$.

which the function $F^{(0)}(z)$ is finite) is not necessarily small compared to the wavelength λ of the light. If

$$4\omega_r T\gg\frac{\lambda_{DB}}{\lambda}, \tag{58}$$

the formulas (35) and (36) still describe the formation of well-resolved peaks at $0,\pm 4\omega_r T/k,\pm 8\omega_r T/k,\cdots$, whose amplitudes are described by (39) and, as a consequence, of a comblike structure in momentum space. However, the simple representation (51) for the envelope becomes invalid. For $\lambda_{DB}>\lambda$ the shape of the envelope differs qualitatively from the case $\lambda_{DB}\ll\lambda$ discussed above. As the scattering diagram in Fig. 4 shows, each peak in the coordinate representation exhibits spatial oscillations corresponding to the diffraction of atoms on a standing wave (the Kapitza–Dirac resonance effect).[22] In the momentum representation this feature manifests itself in the peak envelope, which demonstrates quantum diffraction features discussed in Ref. 16. In the opposite limit where

$$4\omega_r T\leqslant\frac{\lambda_{DB}}{\lambda}, \tag{59}$$

the very property (53) of factorization of the distribution function into a periodic function $\Phi(p)$ and an envelope $\tilde{\mathscr{E}}(p)$ becomes invalid. When $\lambda_{DB}>\lambda$ holds, the two effects, Ramsey cooling proper and the Kapitza–Dirac resonance effect, may influence each other (may interfere with each other). For $4\omega_r T=(2n+1)\pi$, their interference is destructive. The action of two light pulses reduces to broadening the momentum distribution, while the quantum features induced by diffraction and the CPT effect prove to be suppressed. But for $4\omega_r T=2n\pi$, the effects interfere constructively, amplifying each other.

4.4. Ramsey cooling of atoms precooled by velocity-selective coherent population trapping

It is appropriate at this point to compare the results of the theory developed in the present paper with quantum simulations of Ramsey cooling of helium atoms (the $j_g=1\rightarrow j_e=1$ transition).[17] The following situation was examined in Ref. 17. The atoms were precooled by velocity-selective coherent population trapping in a resonant $\sigma_+-\sigma_-$ field below the quantum limit $\hbar\omega_r$ (the full width at half-maximum of the peaks in the momentum representation was $0.34\hbar k$). Then, in the course of $T=500\gamma^{-1}$ the atoms were left to themselves, after which a second pulse of light of length $\tau=100\gamma^{-1}$ was sent through the system. Here we will not consider the precooling process—we model it by a mixture of Lorentzian (for atoms in the $|\Psi_{NC}\rangle$ state) and Gaussian (for atoms in the $|\Psi_1^{\Lambda}\rangle$ state) distributions (the $|\Psi_1^{V}\rangle$ state is assumed empty):

$$\hat{\rho}(z)=A\ \exp(-ak|z|)|\Psi_{NC}\rangle\langle\Psi_{NC}|+B\times\exp(-b(kz)^2)|\Psi_1^{\Lambda}\rangle\langle\Psi_1^{\Lambda}|.$$

The following values of the parameters correspond to the data given in Ref. 17: $a=0.17$, $A=1.25$, $b=7.78$, and $B=4.75$. Since the atoms in the $|\Psi_1^{\Lambda}\rangle$ state do not participate in Ramsey cooling, the initial condition for (35) should be taken in the form of a Lorentzian peak:

$$W^{(1)}(z)=A\ \exp(-ak|z|).$$

According to (35),

$$W^{(2)}(z)=W^{(1)}(z)+\frac{2-Q_{-1}(kz)-Q_1(kz)}{8-2Q_{-1}(kz)-2Q_1(kz)}\times(W^{(1)}(z+4\omega_r T/k)+W^{(1)}(z-4\omega_r T/k)-2W^{(1)}(z)).$$

We found the Fourier transform $\tilde{W}(p)$, added to it a broad background corresponding to the Gaussian distribution, and used the peak splitting formula (54). The result was the distribution function $\tilde{F}(p)$ in Fig. 5, which agrees qualitatively with the result of Ref. 17. For instance, according to our calculations the width of the principal peaks at the points $p=\pm\hbar k$ was $0.06\hbar k$ (the total width at halfheight without allowing for background), while the data of Ref. 17 yield $0.04\hbar k$. The discrepancy can be explained by the fact that condition (32) was not met in Ref. 17, $kv\sim\gamma S$, and during the second field pulse there was additional cooling due to selective coherent population trapping.

5. DEPENDENCE OF THE CONTRAST OF THE SCATTERING DIAGRAM ON THE PARAMETER $\gamma S\tau$

In view of the experiment of Sander *et al.*[18] in pulsed cooling on the $j_g=2\rightarrow j_e\rightarrow 2$ in the D_1-line of ^{87}Rb, we consider a situation in which the pulse length is not sufficiently large for the atoms to switch to the dark state com-

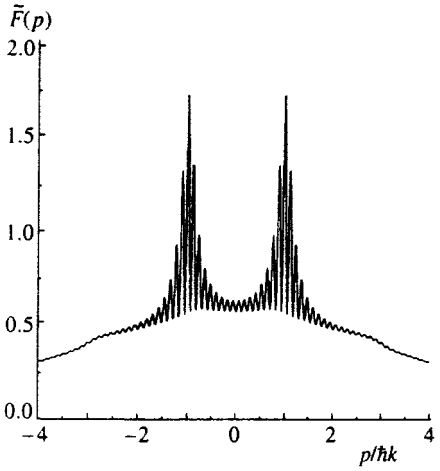


FIG. 5. Diagram of scattering of atoms precooled by velocity-selective coherent population trapping.

pletely. More precisely, we assume that in the event of weak saturation, $S \ll 1$, only one of the two conditions in (13) is met, i.e.,

$$\gamma\tau \gg 1, \quad \gamma S\tau \leqslant 1.$$

In this case instead of (35) we have the following recurrence formula for the density matrix of the atoms:

$$\rho^{(N+1)}_{\mu_1,\mu_2}(z) = \sum_{\nu_1,\nu_2=-j_g}^{j_g} \mathscr{R}^{\nu_1,\nu_2}_{\mu_1,\mu_2}(z|\tau)\exp[-i\omega_r T(\nu_1^2-\nu_2^2)] \times \rho^{(N)}_{\nu_1,\nu_2}(z+2\omega_r T(\nu_1-\nu_2)/k), \tag{60}$$

where $\mu_{1,2}$ and $\nu_{1,2}$ label the magnetic sublevels of the ground state (the density matrix of the excited state and the off-diagonal elements can be expressed in terms of the density matrix of the ground state in the $S \ll 1$ limit). The initial conditions also change appropriately:

$$\rho^{(1)}_{\mu_1,\mu_2}(z) = \sum_{\nu=-j_g}^{j_g} \mathscr{R}^{\nu,\nu}_{\mu_1,\mu_2}(z|\tau)\exp(ikz\nu)\frac{F^{(0)}(z)}{2j_g+1}. \tag{61}$$

Figures 6a, b, and c depict the results based on the numerical calculation of the matrix exponential function $\mathscr{R}$ for the $j_g=2 \to j_e=2$ transition for $\delta=7\gamma$, $N=28$, $T=\pi/4\omega_r$ (in accordance with the data of Ref. 18) and for different values of the $\gamma S\tau$ parameter: 0.4, 0.8, 1.6, and 16. The initial distribution is assumed to be Gaussian, $F^{(0)} \times(z)=\exp(-b(kz)^2)$ with $b=1.45$ (the corresponding Gaussian halfwidth σ at $e^{-1/2}$ of the maximum of the momentum distribution is $1.7\hbar k$). Table I shows that the envelope halfwidth σ_e and the contrast of the scattering diagram (the peak-to-peak ratio $\tilde{F}(\hbar k)/\tilde{F}(0)$) increase with $\gamma S\tau$, but the peak halfwidth σ_p decreases. The value of $\gamma S\tau$ can be increased by increasing the pulse length and the field inten-

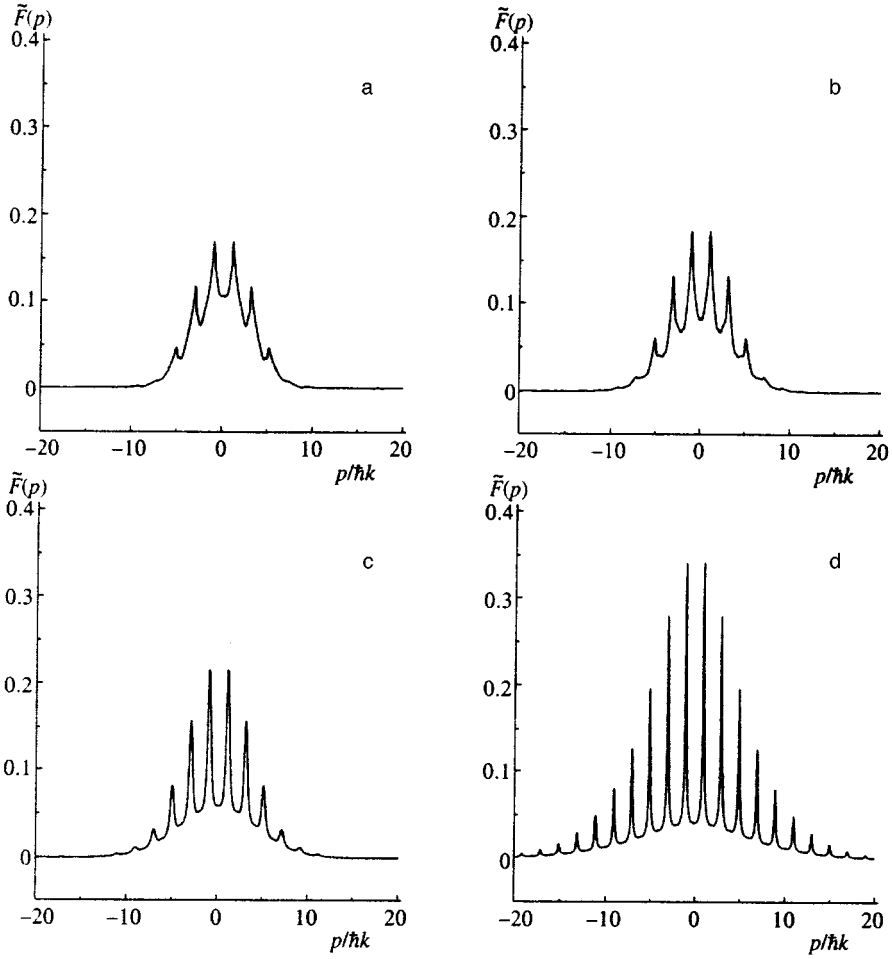


FIG. 6. Dependence of the scattering diagram for $j_g=2$ and $4\omega_r T=\pi$ on the parameter $\gamma S\tau$. Shown is the momentum distribution function for the atoms, $\tilde{F}(p)$, in the laboratory reference frame after 28 light pulses have acted on the system at (a) $\gamma S\tau=0.4$, (b) $\gamma S\tau=0.8$, (c) $\gamma S\tau=1.6$, and (d) $\gamma S\tau=16$.

TABLE I. Dependence of the characteristics of the scattering diagram on the parameter $\gamma S\tau$.

$\gamma S\tau$	$\sigma_e/\hbar k$	$\sigma_p/\hbar k$	$\tilde{F}(\hbar k)/\tilde{F}(0)$
0.4	3.2	0.15	1.5
0.8	3.5	0.15	2.2
1.6	4.0	0.15	3.7
16	6.6	0.05	31
∞	6.6	0.05	55

sity or by decreasing the detuning from resonance. The experimental data correspond to $\gamma S\tau=1.6$, which is far from the optimum values determined by the following qualitative considerations. Only the atoms trapped into a CPT state contribute to formation of the comblike structure, so that the maximum contrast is observed when all the atoms are transferred into $|\Psi_{NC}\rangle$. The condition for almost complete clearing of the medium can be written as $\gamma S\tau>1/\alpha$, where α is the minimum eigenvalue (25) (in our case $\alpha=1/6$). For instance, decreasing the detuning by a factor of $\sqrt{10}$ we find that $\gamma S\tau=16$, which yields a scattering amplitude close to the analytical results achieved in the previous sections of this paper (see Fig. 6d and Table I). Note that here one can speak only of qualitative agreement between our results and those of Ref. 18. In particular, at $\gamma S\tau=1.6$ our data yield $\sigma_e\approx 4.0\hbar k$, while the data of Ref. 18 yield $\sigma_e\approx 4.4\hbar k$; for the peak halfwidth we have $\sigma_p\approx 0.2\hbar k$, while Ref. 18 yields $\sigma_p=0.3\hbar k$. Here are some factors that lead to these discrepancies. First, the interaction in the experiment of Ref. 18 did not consist only of the closed transition $j_g=2\rightarrow j_e=2$, since there was an additional field that pumped the atoms from the hyperfine component with $j_g=1$. This ''depleting'' field acted continuously, and in the interval between light pulses in the active transition atoms returned to the sublevel with the angular momentum $j_g=2$, with the momentum distribution being further broadened because of the recoil effect. Second, the finite momentum resolution of the detecting system made it impossible to observe structures narrower than $0.3\hbar k$. Finally, our main assumption that the atoms remain immobile when a light pulse acts on the system works poorly in conditions corresponding to those of the experiment, where $kv\tau\sim\gamma S\tau\sim 1$.

6. CONCLUSION

By assuming that the translational motion of atoms can be completely ignored while a light pulse acts on the system we were able to develop a fairly simple analytical description of Ramsey cooling of atoms. The method accounts exactly for the quantum effects that are due to the recoil in the absorption (emission) of photons and the free motion of atoms in the absence of a field. We found that the interaction of atoms and light pulses in CPT conditions may generate correlations between arbitrarily distant points z_1 and z_2, a fact of fundamental importance for atomic optics and atomic interferometry. We also discovered features of the scattering diagram not detected in earlier studies,[17,18] namely,

(1) the formation (after a moderate number of light pulses have acted on the system) of additional maxima (i.e., in addition to the principal maximum);

(2) the lack of dependence in the case $j_g=1$ (provided that conditions (32) are met) of the scattering diagram on the field parameters; and

(3) the dependence in the case $j_g=2$ of the scattering diagram in the field intensity and detuning. However, in the case of a broad (on the photon momentum scale) initial distribution at resonant values of the time interval between two consecutive light pulses (Eq. (44)) this dependence manifests itself only in the peak envelope in the momentum distribution.

We examined the scattering of atoms with angular momenta $j_g=1\rightarrow j_e=1$ and $j_g=2\rightarrow j_e=2$ by a resonant $\sigma_+-\sigma_-$ field. The method can be applied (after proper modification) to transitions with larger values of angular momenta and to more complicated field configurations (including two- and three-dimensional).

The authors are grateful to Dr. Frank Sander from Garching, who read the manuscript and made useful remarks and clarifications.

APPENDIX

The coefficients D and $m_{i,j}$ have the following form:

$$
\begin{aligned}
D=81&\left(4\left(\frac{\delta}{\gamma}\right)^2+\left(\frac{\Omega}{\gamma}\right)^4\right)(-144+60Q_{-1}-3Q_{-1}^2\\
&+60Q_1-6Q_{-1}Q_1-3Q_1^2+4Q_{-1}Q_0+4Q_1Q_0\\
&+4Q_0^2)-18\left(\frac{\Omega}{\gamma}\right)^2(2160-1188Q_{-1}+225Q_{-1}^2\\
&-18Q_{-1}^3-1188Q_1+210Q_{-1}Q_1-6Q_{-1}^2Q_1\\
&+225Q_1^2-6Q_{-1}Q_1^2-18Q_1^3-60Q_{-1}Q_0-60Q_1Q_0\\
&+32Q_{-1}Q_1Q_0-60Q_0^2+8Q_{-1}Q_0^2+8Q_1Q_0^2)-(15\\
&-3Q_{-1}-3Q_1+2Q_0)(2160-1044Q_{-1}+225Q_{-1}^2\\
&-27Q_{-1}^3-1044Q_1-30Q_{-1}Q_1+15Q_{-1}^2Q_1\\
&+225Q_1^2+15Q_{-1}Q_1^2-27Q_1^3-288Q_0+60Q_{-1}Q_0\\
&-18Q_{-1}^2Q_0+60Q_1Q_0+28Q_{-1}Q_1Q_0-18Q_1^2Q_0\\
&-60Q_0^2+12Q_{-1}Q_0^2+12Q_1Q_0^2+8Q_0^3),
\end{aligned}
\tag{A1}
$$

$$
\begin{aligned}
m_{1,1}=972&\left(4\left(\frac{\delta}{\gamma}\right)^2+\left(\frac{\Omega}{\gamma}\right)^4\right)(-3+Q_{-1}+Q_1)\\
&-216\left(\frac{\Omega}{\gamma}\right)^2(45-21Q_{-1}+3Q_{-1}^2-21Q_1\\
&+2Q_{-1}Q_1+3Q_1^2)-12(15-3Q_{-1}+2Q_0-3Q_1)\\
&\times(45-18Q_{-1}+3Q_{-1}^2-6Q_0+2Q_{-1}Q_0-18Q_1\\
&-2Q_{-1}Q_1+2Q_0Q_1+3Q_1^2),
\end{aligned}
\tag{A2}
$$

$$
m_{2,2}=-243\left(4\left(\frac{\delta}{\gamma}\right)^2+\left(\frac{\Omega}{\gamma}\right)^4\right)(Q_{-1}+2Q_0+Q_1)
$$

$$+54\left(\frac{\Omega}{\gamma}\right)^2(-15Q_{-1}-30Q_0+4Q_{-1}Q_0-15Q_1$$
$$+8Q_{-1}Q_1+4Q_0Q_1)+3(15-3Q_{-1}+2Q_0$$
$$-3Q_1)(-15Q_{-1}-3Q_{-1}^2-30Q_0+4Q_{-1}Q_0$$
$$+4Q_0^2-15Q_1+10Q_{-1}Q_1+4Q_0Q_1-3Q_1^2), \quad \text{(A3)}$$

$$m_{1,2}=6(Q_{-1}-Q_1)(2Q_0+3Q_{-1}+3Q_1-6)\left(15-\frac{18i\delta}{\gamma}\right.$$
$$\left.+9\left(\frac{\Omega}{\gamma}\right)^2-3Q_{-1}+2Q_0-3Q_1\right), \quad \text{(A4)}$$

$$m_{2,1}=6(Q_{-1}-Q_1)(2Q_0+3Q_{-1}+3Q_1-6)\left(15+\frac{18i\delta}{\gamma}\right.$$
$$\left.+9\left(\frac{\Omega}{\gamma}\right)^2-3Q_{-1}+2Q_0-3Q_1\right). \quad \text{(A5)}$$

*)E-mail: Tumaikin@Univ.nsk.su

[1] A. Aspect, E. Arimondo, R. Kaiser, N. Vansteenkiste, and C. Cohen-Tannoudji, Phys. Rev. Lett. **61**, 826 (1988).
[2] A. Aspect, E. Arimondo, R. Kaiser, N. Vansteenkiste, and C. Cohen-Tannoudji, J. Opt. Soc. Am. B **6**, 2112 (1989).
[3] F. Mauri and E. Arimondo, Europhys. Lett. **16**, 717 (1994).
[4] M. A. Ol'shaniĭ, J. Phys. B **24**, L583 (1991).
[5] M. A. Ol'shanyĭ, Opt. Spektrosk. **76**, 196 (1994) [Opt. Spectrosc. **76**, 178 (1994)].
[6] C. Foot, H. Wu, E. Arimondo, and G. Morigi, J. Phys. II **4**, 1913 (1994).
[7] A. V. Taĭchenachev, A. M. Tumaĭkin, M. A. Ol'shanyĭ, and V. I. Yudin, JETP Lett. **53**, 351 (1991).
[8] P. Marte, R. Dum, R. Taïeb, P. Zoller, M. S. Shahriar, and M. Prerntiss, Phys. Rev. A **49**, 4826 (1994).
[9] H. Stecher, H. Ritsch, P. Zoller, F. Sander, T. Esslinger, and T. W. Hänsch, Phys. Rev. A **55**, 545 (1997).
[10] J. Lawall, F. Bardou, B. Saubamea, K. Shimizu, M. Leduc, A. Aspect, and C. Cohen-Tannoudji, Phys. Rev. Lett. **73**, 1915 (1994).
[11] J. Lawall, S. Kulin, B. Saubamea, N. Bigelow, M. Leduc, and C. Cohen-Tannoudji, Laser Phys. **6**, 153 (1996).
[12] T. Esslinger, F. Sander, M. Weidemüller, A. Hemmerich, and T. W. Hänsch, Phys. Rev. Lett. **76**, 2432 (1996).
[13] F. Bardou, J. P. Bouchard, O. Emile, A. Apect, and C. Cohen-Tannoudji, Phys. Rev. Lett. **72**, 203 (1994).
[14] E. Korsunsky, A. Snegiriov, V. Gordienko, B. Matisov, and L. Windholz, Z. Phys. D **30**, 23 (1994).
[15] E. Korsunsky, D. Kosachiov, B. Matisov, Yu. Rozhdestvensky, L. Windholz, and C. Neureiter, Phys. Rev. A **48**, 1419 (1993).
[16] A. V. Taichenachev, A. M. Tumaikin, V. I. Yudin, Laser Phys. **2**, 575 (1992).
[17] H. Wu, E. Arimondo, and C. Foot, Quantum Semiclassic. Opt. **8**, 983 (1996).
[18] F. Sander, T. Devolder, T. Esslinger, and T. Hänsch, Phys. Rev. Lett. **78**, 4023 (1997).
[19] N. F. Ramsey, Phys. Rev. **78**, 695 (1950).
[20] V. S. Smirnov, A. M. Tumaĭkin, and V. I. Yudin, Zh. Éksp. Teor. Fiz. **96**, 1613 (1989) [Sov. Phys. JETP **69**, 913 (1989)].
[21] D. A. Varshalovich, A. N. Moskalev, and V. K. Khersonskiĭ, *Quantum Theory of Angular Momentum*, World Scientific, Singapore (1987).
[22] A. P. Kazantsev, G. I. Surdutovich, and V. P. Yakovlev, *Mechanical Effect of Light on Atoms*, World Scientific, Singapore (1990).

Translated by Eugene Yankovsky

Ultrahigh-energy cosmic rays: possible origin and spectrum

A. V. Uryson

P. N. Lebedev Physical Institute, Russian Academy of Sciences, 117924 Moscow, Russia
(Submitted 30 June 1997)
Zh. Éksp. Teor. Fiz. **113**, 12–20 (January 1998)

The complicated shape of the cosmic ray spectrum recorded by giant arrays in the energy range $10^{17}-10^{20}$ eV is analyzed. It is shown that in the energy region $\sim 10^{18}-10^{19}$ eV the spectrum probably coincides with the injection spectrum whose exponent is equal approximately to $3.2-3.3$. The flatter component in the energy region $(3.2-5.0)\times 10^{19}$ eV is due to braking of extragalactic protons on primordial photons (the cosmic background radiation). At energies exceeding 3.2×10^{19} eV the spectrum does not have a blackbody cutoff. The possibility of determining the distances at which cosmic rays originate and investigating the evolution of their sources on the basis of ultrahigh-energy cosmic ray data is discussed. [S1063-7761(98)00201-7]

1. INTRODUCTION

The origin of cosmic rays of ultrahigh-energy, $E>10^{17}$ eV, is still unclear. The experimental data indicate that cosmic rays with energies $E>4\times 10^{19}$ eV are probably extragalactic in origin.[1–3] If this is so, then their spectrum may have a blackbody cutoff:[4,5] the recorded particle flux with energy 6×10^{19} eV will be twice as small as expected from the power-law extrapolation of the spectrum as a consequence of the interaction of the cosmic rays with primordial photons (the cosmic background radiation) in intergalactic space. However, if the proton sources are not farther away from us than 40–50 Mpc the blackbody cutoff will be absent since protons of energies up to $E\approx 10^{22}$ eV traverse such distances almost freely.[6] In Ref. 7 it was shown that the main proton sources with energies $E>E_{bb}\approx 3.2\times 10^{19}$ eV are probably the nuclei of active galaxies no farther from us than 40 Mpc if the Hubble constant is equal to 75 km/(s Mpc). In this case the proton spectrum does not have a blackbody cutoff. At present the experimental data obtained at different detectors—Yakutsk,[8] Akeno and AGASA,[9] ''Fly's Eye,''[10] Haverah Park,[11], Sydney,[12] and Volcano Ranch[13]—neither confirm nor refute its presence.

The origin of cosmic rays in the energy region $10^{17}<E\leqslant 10^{19}$ eV is determined not only on the basis of their spectrum, but also their anisotropy and chemical composition.[1–3] However, the available experimental data are not sufficiently unequivocal to determine whether cosmic rays of such energies are galactic or extragalactic.

Different models have been considered in attempts to explain the shape of the spectrum in the energy region $E>10^{17}$ eV. According to the results of Refs. 14–17, the spectrum can have a complicated shape if it is formed by extragalactic protons whose sources are hundreds of megaparsecs from us. On the other hand, modeling of charged-particle trajectories in galactic magnetic fields has shown that cosmic rays in the energy region $10^{17}-10^{18}$ eV are galactic or are of mixed origin—they are accelerated in the Galaxy and in the Local Supercluster.[18,19]

The present paper proposes two models to explain the proton spectra. The first model assumes that cosmic rays are galactic in the energy region $E<10^{19}$ eV and extragalactic for $E>10^{19}$ eV. In the second model they are assumed to be extragalactic for energies $E>10^{17}$ eV. Using these two models the paper discusses the possibility of investigating the evolution of sources of ultrahigh-energy cosmic rays.

2. EXPERIMENTAL DATA

The cosmic ray spectrum in the energy region $E>10^{17}$ eV has a complicated shape:[3,8,10] for $E\approx 5\times 10^{17}$ eV the slope of the spectrum γ grows from $\gamma\approx 3.0-3.1$ to $\gamma\approx 3.2-3.3$ (the error in the determination of γ is $0.02-0.06$), while in the energy region $E\approx 10^{19}$ eV it decreases to $\gamma\approx 2.6-2.7$, i.e., a flatter component appears in the spectrum. The error in the determination of the slope of the flatter component is 0.1. (Spectral slopes are not provided in Refs. 11–13.) Cosmic ray spectra measured at different detectors[8–12] and energy-normalized in the same way as in Ref. 3 are plotted in Fig. 1.

3. COSMIC RAY SPECTRUM FOR $E<10^{19}$ eV IN THE GALACTIC MODEL

The propagation of cosmic particles in the Galaxy can be described in the diffusion approximation if their energy does not exceed $10^{17}-10^{18}$ eV (Ref. 18). In addition, it was shown in Refs. 20–22 that particles with charge Z cease to propagate diffusively if their energy exceeds some value E_0Z, such that in the energy region $E>E_0Z$ the particle spectrum coincides with the injection spectrum and the slopes of the spectra are equal to $\gamma=\gamma_0$. (Note that this result was obtained in Refs. 20–22 by different methods: in Refs. 20 and 21 it is due to drift of ultrahigh-energy cosmic rays in large-scale magnetic fields, while in Ref. 22 it is due to a transition to collisionless propagation of particles in a medium where they excite MHD waves.) An estimate of the energy E_0 was obtained by numerical simulation of the particle trajectories in the Galactic magnetic field:[1] $E_0\approx 2\times 10^{18}$ eV.

The chemical composition of cosmic rays in the energy region $10^{18}-10^{19}$ eV is still unclear. According to the results of measurements reported in Ref. 23, the proton fraction in-

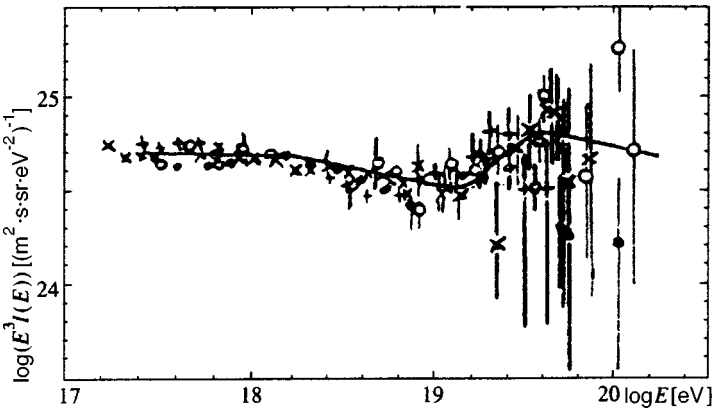


FIG. 1. Cosmic ray spectrum for $E>10^{17}$ eV, energy-normalized in the same way as in Ref. 3; measurement data at: ● — Yakutsk,[8] × — Akeno and AGASA,[9] + — ''Fly's Eye,''[10] ○ — Haverah Park.[11] Solid line — theoretical spectrum calculated in the galactic model for $E<E_{bb}$ and in the extragalactic model for $E>E_{bb}$.

creases systematically, starting from $\sim 10^{16}$ eV, so that for energies $E>10^{18}$ eV protons predominate. According to the data of Ref. 10, the composition varies in the energy range $4\times10^{17}-4\times10^{19}$ eV in the following way: to start with, iron nuclei predominate, and at the other end there are only protons.

We will assume that at energies $E\geqslant10^{18}$ eV protons predominate. Then the spectrum of the protons coincides with their injection spectrum at energies $E\geqslant2\times10^{18}$ eV.

The region $E\geqslant2\times10^{18}$ eV is the region in which the slope of the measured spectrum grows. This means apparently that at energies $E>10^{18}$ eV the slope of the proton injection spectrum γ_0 is roughly equal to $3.2-3.3$: $\gamma_0\approx3.2-3.3$.

Particles with energy $E>E_{bb}$ probably accelerate mainly from sources no farther away from us than 40–50 Mpc,[7,24,25] and as a consequence their spectrum does not have a blackbody cutoff. If this is so, then the exponent of the spectrum in this region coincides with the exponent γ_0 of the injection spectrum. We assume that in the region $E>E_{bb}$ the injection spectrum is the same as for $E\geqslant2\times10^{18}$ eV. Then the slope of the spectrum in the region $E>E_{bb}$ is equal to $\gamma=3.2-3.3$.

Particles with energies $E>E_{bb}$, propagating from sources closer than 40 Mpc, will interact with the cosmic background radiation until their energy falls to $E\approx(3.2-5.0)\times10^{19}$ eV. Particles with such energies cannot undergo any interactions in intergalactic space since their mean free paths in the cosmic background radiation field is quite large: $\lambda>1000$ Mpc (Ref. 6). This leads to the result that protons with energies $E>3.2\times10^{19}$ eV ''pump'' into the region $E\approx(3.2-5.0)\times10^{19}$ eV, and as a result the slope of the spectrum in this region changes from $\gamma>3.1$ to a value γ_1 such that

$$\int_{E_{bb}} E^{-\gamma}dE=\int_{E_{bb}}^{E_3} E^{-\gamma_1}dE,$$

where E_3 is the upper limit of the energy range of the flatter component. We find γ_1 from the experimental data.[3,8–13] The measured value of E_3 is approximately 4×10^{19} eV, and the energy of the particles is determined with an error of approximately 20–30% (Refs. 8 and 10). Therefore we estimated the exponent γ_1 for several values of E_3 in the interval $4\times10^{19}<E_3<5\times10^{19}$ eV and several values of γ in the interval $3.1<\gamma\leqslant3.3$. The calculated values of γ_1 are listed in Table I. The energy range of the flatter component with allowance for the 30% error in the determination of the energy is $E\approx(2-5)\times10^{19}$ eV. It agrees with the measurement results of Refs. 8, 9, 11–13 and does not contradict the single-measurement data of Ref. 10.

The possible existence in the spectrum of a flattened component of such a nature was predicted in Refs. 14–16.

The theoretical spectrum based on the proposed model is plotted in Fig. 1. It is normalized to the measured intensity at $E\approx10^{19}$ eV. In addition, the calculations assumed that the energy region of the flatter component is $E\approx(3.2-5.0)\times10^{19}$ eV. The theoretical spectrum agrees with the measurements within the limits of error.

Let us consider the slopes of the spectra in the region $E\geqslant10^{19}$ eV. It is clear from the table that the proposed model yields values of γ_1 in agreement with the slope of the flatter component.

To estimate the slope of the measured spectrum in the region $E>E_{bb}$ we make use of the summary of the experimental data in Ref. 3: in 1993 only 881 events with energy $E\geqslant10^{19}$ eV were recorded, and only 7 with $E\geqslant10^{20}$ eV and 2 with $E\geqslant2\times10^{20}$ eV. For a power-law spectrum where N is the number of particles with energy greater than E, $N(\geqslant E)$, the relation $N_1(\geqslant E_1)/N_2(\geqslant E_2)=(E_1/E_2)^{\gamma+1}$ holds, and from it we obtain $\gamma=3.1^{+0.2}_{-0.1}$ for $E_1=10^{19}$ eV, $E_2=10^{20}$ eV.

Some of the 881 events have energy in the interval $\approx(1.0-3.2)\times10^{19}$ eV and make up the flattened component. Therefore, in the region $E>3.2\times10^{19}$ eV the exponent of the spectrum will be greater than the estimate: $\gamma>3.1$ and, consequently, $\gamma_0>3.1$.

TABLE I. Calculated exponent γ_1 of the flatter component for different values (within the limits of experimental error) of its upper limit E_3 and spectral slope γ.

γ	E_3, eV	γ_1
3.0	4.9×10^{19}	2.6
3.05	4.8×10^{19}	2.65
3.05	4.9×10^{19}	2.7
3.1	4.5×10^{19}	2.6
3.2	4.5×10^{19}	2.7
3.3	4.2×10^{19}	2.7

If the blackbody cutoff is absent, then the proton spectrum coincides with the injection spectrum in two regions: $2\times10^{18}\leqslant E<10^{19}$ eV and $E>3.2\times10^{19}$ eV. Estimates of the slope of the spectrum in these intervals agree with each other: $\gamma=3.2-3.3$ and $\gamma>3.1$; consequently, the measured spectrum apparently does not have a blackbody cutoff. (The authors of Ref. 3, on the basis of these same experimental data, concluded that its existence was possible. They assumed that if there is no cutoff, then the slope for $E>E_{bb}$ must coincide with the slope of the flatter component.)

Cosmic ray injection spectrum

Let us consider how the cosmic ray injection spectrum varies between the different energy intervals, making use of the results presented above.

In the region $E<10^{17}$ eV the proton spectrum is related to the injection spectrum by the relation[1] $N(>E)\propto E^{-\gamma_0-\mu}$, where the parameter μ describes the dependence of the diffusion coefficient on the energy, $D\propto E^{\mu}$. The measurements of Ref. 26 yield a value of μ the range $0.3-0.7$ for energies of a few GeV/nucleon, the measurements of Ref. 27 yield $\mu=0.6$ for energies ≈1 TeV/nucleon, and analysis of the diffusion model[18] yields $\mu=0.15-0.20$ in the energy range $E=10^9-10^{17}$ eV. The slope γ of the cosmic ray spectrum for $E<3\times10^{15}$ eV is equal to approximately 2.75 and hence the exponent of the injection spectrum in this region is $\gamma_0\approx2.2$ for $\mu=0.6$ and $\gamma_0\approx2.6$ for $\mu=0.15-0.2$.

The spectral index γ_0 in the region $3\times10^{15}-10^{18}$ eV is hard to determine since it is still not clear for what reasons the slope of the cosmic ray spectrum varies for $E>3\times10^{15}$ eV. Particles with charge Z are accelerated to energies $E\leqslant10^{15}Z^{-1}$ eV, apparently, in supernova bursts.[28] According to Refs. 18, 29, and 30, the slope of the spectrum varies as a consequence of the propagation and subsequent acceleration of the particles in the Galaxy. In addition, it is possible[3] that high-energy protons accelerate in other (not yet established) processes, and their injection spectrum changes.

Thus, if protons predominate in the composition of cosmic rays in the energy region $E>10^{18}$ eV (Ref. 23), then it is possible that the slope of the injection spectrum varies in the following way: it increases to a value $\gamma_0\approx3.2-3.3$ for $E>10^{18}$ eV in comparison with the region $E<3\times10^{15}$ eV, where the slope does not exceed 2.6, $2.2\leqslant\gamma_0\leqslant2.6$.

4. EXTRAGALACTIC MODEL OF THE ORIGIN OF COSMIC RAYS IN THE ENERGY REGION $E>10^{17}$ eV

In this model we assume that the particles with energy $E>10^{18}$ eV are mainly extragalactic, that their spectrum has the single exponent $\gamma\approx3.0-3.1$, found in the region $E\approx(2-4)\times10^{17}$ eV,[3,8,10] and that for energies $E>10^{18}$ eV the spectrum is distorted as a result of interaction of the particles with the fossil radiation (cosmic background radiation) in intergalactic space.

A possible change in the shape of the spectrum in the region $E\leqslant3.2\times10^{19}$ eV was noted in Ref. 14 and investigated in Refs. 15–17. According to the results of these studies, the spectrum can have a complicated shape if it is formed by extragalactic protons: by interacting with the cosmic background radiation they lose energy as a consequence of the red shift, in processes of e^+e^- pair formation if their energy satisfies $E<10^{19}$ eV, and by photo-generation of pions for $E>E_{bb}$. As a result, the differential spectrum of the protons emitted by an isolated source can exhibit a hump and a valley from e^+e^- pair formation, a photo-pion hump, and a blackbody cutoff. However, if the sources uniformly fill the Universe, the hump and valley from e^+e^- pair formation will be weakly expressed. The sources should fill a sphere of radius corresponding to the red shift $z\approx0.2$. The photo-pion hump (without a valley) will be present in the spectrum if the proton sources uniformly fill a sphere of radius corresponding to $z\leqslant0.085$.

Hence we may assume that for a nonuniform source distribution the spectrum will have not only a photo-pion hump, but also other peculiarities. Thus, the measured spectrum could be explained by varying the shape of the source distribution. Observation data indeed indicate that the nuclei of active galaxies are distributed nonuniformly on scales reaching $\sim10^2$ Mpc (Ref. 31) and that they are most likely the main sources of ultrahigh-energy protons.[1,7]

Let us analyze the spectrum in accordance with this hypothesis. If, in accordance with Refs. 7, 24, and 25, the particles with $E>E_{bb}$ are accelerated mainly in sources separated from us by distances not exceeding $40-50$ Mpc, then their spectrum does not have a blackbody cutoff. Thus, the exponent of the spectrum in this region is equal to $\gamma\approx3.0-3.1$. This is just the rough estimate of the slope in the region $E>E_{bb}$ obtained above. It coincides with the slope for $E\approx(1-5)\times10^{17}$ eV and, consequently, in this model the measured spectrum does not have a blackbody cutoff.

The shallow component at energies $E\approx(3.2-5.0)\times10^{19}$ eV in this model is also due to ''pumping'' of protons into this region having energies $E>E_{bb}$. Values of the spectral index γ_1 of the flattened component for $\gamma\approx3.0-3.1$ are shown in the table.

In the spectrum of particles accelerated in sources with $z\approx0.2$, a notch can appear in the region $E<10^{19}$ eV as a result of particles with energies $E\approx2\times10^{18}-3\times10^{19}$ eV losing energy by creating e^+e^- pairs in the background radiation field.[16]

This model can be verified by calculating the proton spectra of sources distributed nonuniformly at distances $r>40$ Mpc from us with a nonuniformity scale of up to ~100 Mpc.

5. POSSIBLE CONSTRAINTS ON COSMIC RAY SOURCES

Particles with energies $E>10^{18}$ eV probably propagate along straight-line paths in the Galaxy[21,22] and beyond it.[32]

The energy of a proton emitted at the epoch with red shift z falls as it propagates in intergalactic space due to the red shift and formation of e^+e^- pairs and pions.[1] Let $E_0(z)$ be the energy that a proton should have at its epoch of generation in order for its energy at $z=0$ to be E. We note that the luminosity and density of sources in the accompanying volume increase with growth of their red shift z. Thus, the energy density of extragalactic particles in the interval $(E,E+dE)$ is equal to

$$w_{eg}(E)dE=\int_{z_{\min}}^{z_{\max}} n_{eg}(z)L_{CR}[E_0(z)]\tau(z)dE\ dz,$$

where $n_{eg}(z)$ is the density of extragalactic sources, $L_{CR}(E_0)dE$ is their cosmic-ray luminosity in the interval (E_0, E_0+dE), and $\tau(z)$ is the propagation time of such particles. The lower integration limit $z_{\min}$ corresponds to the distance the particles can traverse essentially without losses. According to Ref. 6, this distance is ~ 10 kpc, and therefore $z_{\min}\approx 0.003$. The upper limit $z_{\max}$ is probably $\approx 3-4$.[33,34]

The energy density $w_{eg}(>E)$ can be found from the cosmic ray spectrum:

$$w_{eg}(>E)=\frac{4\pi}{c}\int (E)EdE,$$

where $I(>E)$ is the total intensity of cosmic rays with energies $>E$ and $c=3\times 10^{10}$ cm/s is the speed of light. (Energy requirements on ultrahigh-energy particle sources are discussed in Ref. 1.)

At the present time, models of the cosmological evolution of sources are not exact enough[35,36] to allow one to extract from them estimates of the density and luminosity of sources $n_{eg}(z)$ and $L_{CR}(z)$. It is also unclear whether the energetics of the sources is not connected in some way with the efficiency of the particle acceleration. It has still not been possible to identify the most powerful extragalactic sources as possible sources of cosmic protons with $E\approx 10^{20}$ eV (Ref. 32). On the contrary, in Ref. 7 I identified the nuclei of active galaxies, emitting moderate fluxes in the radio and x-ray ranges, as the sources of such protons. From the proposed models we obtain the estimate

$$w_{eg}(>E)=\int n_{eg}(z)L_{CR}[E_0(z)]\tau(z)dz.$$

For example, according to Fig. 1, for $E=E_{bb}$ we have $I(E_{bb})E_{bb}^3\approx 10^{24.5}$ $(\mathrm{m}^2\cdot\mathrm{s}\cdot\mathrm{sr}\cdot\mathrm{eV}^{-2})^{-1}$, and hence $w_{eg}(>E_{bb})\approx 4\times 10^{-21}$ erg/cm^3. In the second model $w_{eg}(>E)$ can be estimated at lower energies: from Fig. 1 it follows for $E\approx 2\times 10^{18}$ eV that $I\cdot(2\times 10^{18}\ \mathrm{eV})\cdot(2\times 10^{18}\ \mathrm{eV})^3\approx 10^{24.7}$ $(\mathrm{m}^2\cdot\mathrm{s}\cdot\mathrm{sr}\cdot\mathrm{eV}^{-2})^{-1}$ and hence $w_{eg}(>E)\approx 1\times 10^{-19}$ erg/cm^3.

It is possible that the slope of the flattened component reflects how distant the proton sources are that form it. The farther the source is located from us, the larger the energy that the proton loses on average traversing intergalactic space. The dimensions of the voids between galaxies amount to $(2.5-100)h^{-1}$ Mpc, and between clusters of galaxies are $(100-250)h^{-1}$ Mpc for the Hubble constant $H=100h$ km$\cdot$s$^{-1}\cdot$Mpc^{-1} (Ref. 31). Therefore, if the sources beyond the Local Supercluster are located at a distances $r\geqslant 100$ Mpc, then the slope of the flattened component will be greater than for $r\leqslant 100$ Mpc. Thus, by studying the flattened component it is possible to estimate the distances from which the protons are arriving. So far such an analysis has been difficult to carry out because of the large experimental error in the slope of the spectrum for $E>E_{bb}$. If the second model is confirmed, then it will be possible to obtain from it an estimate of the distances from which these ultrahigh-energy cosmic rays are arriving using the spectrum not only in the region of the flattened component, but also for $E\approx 2\times 10^{18}-10^{19}$ eV, where the experimental errors are significantly less.

6. CONCLUSION

I have proposed two models of the origin of ultrahigh-energy cosmic rays. In the first model the cosmic rays are assumed to be extragalactic for $E\geqslant 3\times 10^{19}$ eV and galactic at lower energies. In the second model they are assumed to be extragalactic starting at $E>10^{17}$ eV. It follows from both models that the measured spectrum has apparently no blackbody cutoff and that the flattened component in the energy region $(3-5)\times 10^{19}$ eV is due to braking of protons on primordial photons.

Moreover, it follows from the first model that there is a possible changeover from galactic to extragalactic rays in the region of the notch. The data on anisotropy and chemical composition in this energy region are still not sufficiently definite to reliably confirm this conclusion.

In the first model we also found that the injection spectrum of cosmic rays with energies $E>10^{18}$ eV is different from that at lower energies. Its exponent is larger: $\gamma_0\approx 3.2-3.3$ whereas for $10^{10}<E<3\times 10^{15}$ eV it lies in the interval $2.2\leqslant\gamma_0\leqslant 2.6$. The proton spectrum has exponent $\gamma_0\approx 3.2-3.3$ in the energy region $\sim 10^{18}-10^{18}$ eV. This result was obtained from the measurements of Ref. 23, from which it follows that for $E\geqslant 10^{18}$ eV protons predominate in the composition of cosmic rays.

To check the second model it is necessary to calculate the spectra of protons with $E>10^{17}$ eV from sources at distances $r>40$ Mpc from us and nonuniformly distributed on scales up to ~ 100 Mpc.

From the ultrahigh-energy cosmic ray data it is possible to form a picture of the evolution of the sources and estimate the distances from which cosmic rays are arriving. Calculations of the proton spectra with allowance for the evolution of the sources were performed in Refs. 1 and 16. A simple dependence of the source density and luminosity on z was adopted. It was then shown that the evolution of the sources has a more complicated form.[35,36] It follows from the proposed models that the evolution of the sources can be investigated employing the extragalactic particle spectrum

$$I(>E)\frac{4\pi}{c}E=\int n_{eg}(z)L_{CR}[E_0(z)]\tau(z)dz.$$

The distances from which the cosmic rays are arriving can be estimated by analyzing the slope of the flattened component. The dimensions of the voids between galaxies is $(2.5-100)h^{-1}$ Mpc, and between clusters of galaxies $(100-250)h^{-1}$ Mpc (Ref. 31). The slope of the flattened component will be greater if the distance to the sources beyond the Local Supercluster exceeds 100 Mpc than in the case $r<100$ Mpc. However, so far it has been difficult to carry out such an analysis due to the large error in the slope of the measured spectra for $E>E_{bb}$. If the second model is confirmed, then it can be used to obtain an estimate of the distances from which cosmic rays are arriving using the

spectrum not only in the region of the flattened component, but also for $E \approx 2 \times 10^{18} - 10^{19}$ eV, where the experimental errors are significantly less.

The above conclusions can be verified in further measurements of cosmic ray spectra at energies $E > 10^{17}$ eV at the detectors described in Refs. 8–11 and also in Refs. 37 and 38, and at the ShAL-1000 detector,[39] all of which have significantly better resolution. Projected new detectors intended for recording cosmic rays with energies $E > 10^{19}$ eV are described in Ref. 40.

ACKNOWLEDGMENTS

I am grateful to S. I. Nikol'skiĭ and G. B. Christiansen for discussion of the experimental data, V. A. Dogel' and V. S. Ptuskin for discussion of the models of propagation of galactic cosmic rays, and Yu. N. Vetukhnovskiĭ, B. V. Komberg, and O. K. Sil'chenko for some remarks about extragalactic sources.

[1] V. S. Berezinskiĭ, S. V. Bulanov, V. L. Ginzburg, V. A. Dogel', and V. S. Ptuskin, *Astrophysics of Cosmic Rays* North-Holland, Amsterdam (1990).
[2] M. N. D'yakonov, T. A. Egorov, N. N. Efimov *et al.*, *Ultrahigh-Energy Cosmic Radiation* [in Russian] (Nauka, Siberian Branch, Novosibirsk, 1991).
[3] M. Teshima, in *Proceedings of the 23rd ICRC*, Calgary. (Invited, Rapporteur, and Highlight Papers), edited by D. A. Leahy, R. B. Hicks, and D. Venkatesan (World Scientific, Singapore, 1993), p. 257.
[4] G. T. Zatespin and V. A. Kuz'min, JETP Lett. **4**, 78 (1966).
[5] K. Greisen, Phys. Rev. Lett. **16**, 748 (1966).
[6] F. W. Stecker, Phys. Rev. Lett. **21**, 1016 (1968).
[7] A. V. Uryson, JETP Lett. **64**, 77 (1996).
[8] B. N. Afanasiev, M. N. D'yakonov, T. A. Egorov *et al.*, in *Proceedings of the 24th ICRC*, Rome (1995), Vol. **2, p.** 756.
[9] M. Nagano *et al.*, J. Phys. G: Nucl. Phys. **18**, 423 (1992).
[10] O. J. Bird, S. C. Corbato, H. Y. Dai *et al.*, Astrophys. J. **424**, 491 (1994).
[11] M. A. Lawrence, R. J. O. Reid, and A. A. Watson, J. Phys. G: Nucl. Phys. **12**, 653 (1986).
[12] M. M. Winn, J. Ulrichs, L. S. Deak *et al.*, J. Phys. G: Nucl. Phys. **12**, 653 (1986).
[13] J. Linsley, G. Cunningham, D. M. Edge *et al.*, *Catalogue of Highest Energy Cosmic Rays*, No. 1, World Data Center C2, Japan (1980).
[14] A. M. Hillas, Can. J. Phys. **21**, 1016 (1968).
[15] C. T. Hill and D. N. Schramm, Phys. Rev. D **31**, 564 (1985).
[16] V. S. Berezinskiĭ and S. I. Grigor'eva, Zh. Éksp. Teor. Fiz. **93**, 812 (1987) [Sov. Phys. JETP **66**, 457 (1987)].
[17] V. S. Berezinskiĭ, S. I. Grigor'eva, and V. A. Dogel', Zh. Éksp. Teor. Fiz. **96**, 798 (1989) [Sov. Phys. JETP **69**, 453 (1989)].
[18] V. S. Ptuskin, S. I. Rogovaya, V. N. Zirakashvili *et al.*, Astron. Astrophys. **268**, 726 (1993).
[19] D. N. Pochepkin, V. S. Ptuskin, S. I. Rogovaya *et al.*, in *Proceedings of the 24th ICRC*, Rome (1995), Vol. **3, p.** 136.
[20] S. I. Syrovatskii, Comments Astrophys. Space Phys. **3**, 155 (1971).
[21] V. S. Berezinsky, A. A. Mikhailov, and S. I. Syrovatskii, in *Proceedings of the 16th ICRC*, Kyoto (1979), Vol. **2, p.** 86.
[22] V. A. Dogiel, A. V. Gurevich, and K. P. Zybin, Astron. Astrophys. **281**, 937 (1994).
[23] M. N. D'yakonov, V. P. Egorova, A. A. Ivanov *et al.*, JETP Lett. **50**, 442 (1989); M. N. D'yakonov, A. A. Ivanov, S. P. Knurenko *et al.*, *Proceedings of the 23rd ICRC*, Calgary (1993), Vol. **4, p.** 303.
[24] J. Rachen, T. Stanev, and P. Biermann, Astron. Astrophys. **273**, 377 (1993).
[25] R. J. Protheroe and P. A. Johnson, in *Proceedings of the 24th ICRC*, Rome (1995), Vol. **3, p.** 309.
[26] W. R. Webber, in *Composition and Origin of Cosmic Rays*, edited by M. M. Shapiro, (D. Reidel Publishing Co., Dordrecht, 1983).
[27] S. P. Swordy, D. Muller, P. Meyer *et al.*, Astrophys. J. **349**, 625 (1990).
[28] E. G. Berezhko, in *Proceedings of the 24th ICRC*, Rome (1995), Vol. **3, p.** 372.
[29] W. I. Axford, Astrophys. J., Suppl. Ser. **90**, 937 (1994).
[30] R. A. Bell, Mon. Not. R. Astron. Soc. **257**, 500 (1992).
[31] J. Einasto, M. Einasto, and M. Gramann, Mon. Not. R. Astron. Soc. **238**, 155 (1989).
[32] N. Hayashida, K. Honda, and M. Honda, Phys. Rev. Lett. **77**, 1000 (1996).
[33] H. Kuhr, A. Witzel, and I. I. K. Pauliny-Toth, Astrophys. J., Suppl. Ser. **45**, 367 (1981).
[34] A. Hewitt and G. Burbidge, Astrophys. J., Suppl. Ser. **69**, 1 (1988).
[35] B. V. Komberg, A. V. Kravtsov, and V. N. Lukash, Mon. Not. R. Astron. Soc. **282**, 713 (1996).
[36] A. K. Singal, Mon. Not. R. Astron. Soc. **263**, 139 (1993).
[37] M. Teshima *et al.*, Nucl. Phys. B, Proc. Suppl. **28**, 213 (1992).
[38] J. W. Cronin, Nucl. Phys. B, Proc. Suppl. **28**, 213 (1992).
[39] S. S. Ameev, I. Y. Chasnikov, Yu. A. Fomin *et al.*, in *Proceedings of the 24th ICRC*, Rome (1995), Vol. **1, p.** 466.
[40] *Proceedings of the 25th ICRI*, Durban (1997), Vol. **5**.

Translated by Paul F. Schippnick

Squeezed states and quantum chaos

K. N. Alekseev*) and D. S. Priĭmak

L. V. Kirenskiĭ Institute of Physics, Siberian Branch of the Russian Academy of Sciences, 660036 Krasnoyarsk, Russia

(Submitted 20 May 1997)
Zh. Éksp. Teor. Fiz. **113**, 111–127 (January 1998)

We examine the dynamics of a wave packet that initially corresponds to a coherent state in the model of a quantum rotator excited by a periodic sequence of kicks. This model is the main model of quantum chaos and allows for a transition from regular behavior to chaotic in the classical limit. By doing a numerical experiment we study the generation of squeezed states in quasiclassical conditions and in a time interval when quantum–classical correspondence is well-defined. We find that the degree of squeezing depends on the degree of local instability in the system and increases with the Chirikov classical stochasticity parameter. We also discuss the dependence of the degree of squeezing on the initial width of the packet, the problem of stability and observability of squeezed states in the transition to quantum chaos, and the dynamics of disintegration of wave packets in quantum chaos. © *1998 American Institute of Physics*. [S1063-7761(98)00801-4]

1. INTRODUCTION

At present the problem of generating squeezed quantum states draws a lot of attention, both from the standpoint of both pure knowledge and possible applications.[1–3] Most often the topic is squeezed states of the electromagnetic field. If in the simplest case we take a single-mode quantum field, which is described by the creation and annihilation operator $a^\dagger$ and a, the variances of the quadrature field operators $a_1=a+a^\dagger$ and $a_2=-i(a-a^\dagger)$ satisfy the uncertainty relation $\Delta a_1 \Delta a_2 \geqslant 1$, where the equality holds for a coherent state or vacuum. Then, in these simple terms, a squeezed state is a state for which the variance of one of the quadrature components is less than unity. Quantum fluctuations, determined by the uncertainty relation, are represented diagrammatically in the $a_1 a_2$ plane of the quadrature components by a circle for a coherent state or by an ellipse for a squeezed state. In a more systematic description of squeezing, the quantum-noise ellipse is determined in terms of the projection onto the same plane of the horizontal section of the Wigner distribution function, which gives the quasiprobability distribution for measuring the quadratic field components.[3]

A typical situation in experiments in generation of squeezed states is one in which a large number of photons participate in a nonlinear interaction and the amplitude of quantum fluctuations is small compared to the mathematical expectations of the observables.[2,3] In this case the common approach in explaining squeezing is to use the semiclassical setting, where the Wigner quantum function is actually associated with a classical distribution function and instead of examining the dynamics of the quantum-noise ellipse one considers the evolution of the classical phase volume.[3,4]

For quite a long time it has been known that squeezing of light is amplified in systems close to the bifurcation point between two different dynamical regimes.[3–6] Buildup of squeezing in such conditions was considered, e.g., for the parametric interaction of light waves[5] and for the interaction of Rydberg atoms with an electromagnetic mode in a high-Q cavity in a dynamical regime close to the separatrix.[4,6]

The following simple argument is used to explain the buildup of squeezing near a bifurcation point: quantum fluctuations build up for the physical variable that is unstable near the threshold. As a result there is nothing to stop the strong squeezing of fluctuations of the conjugate variable since in a nondissipative system phase volume is conserved.[3]

It must be noted at this point that a number of researchers (see Refs. 3–6) studied the buildup of squeezing near the instability threshold in optical systems with only regular dynamics. However, it is well known that strong (exponential) deformation of the phase volume is one of the main manifestations of dynamical chaos in classical systems.[7] The physical reason for such strong deformations of the phase volume is the local instability of motion, which usually manifests itself within a wide range of values of the control parameter of the dynamical system and not near the bifurcation point. According to the correspondence principle, in the quasiclassical limit a quantum system must manifest the properties of a classical system. Thus, it is quite natural to expect buildup of squeezing in the transition to quantum chaos, too. On the other hand, in a quantum mechanical description we speak only of the dynamics of wave packets, whose center moves almost along a classical trajectory in the course of a certain time interval. Hence in the quasiclassical limit the strong deformations of the phase volume, which accompany the transition to chaos, must manifest themselves in squeezing along a certain direction up to the point when quantum effects produce strong smearing of the wave packet.

As far as we know, the generation of squeezed states in a system with chaotic dynamics was first examined in Refs. 8–10. By employing the $1/N$-expansion method[6,11] (here N is the number of quantum states participating in the dynamics of the system) it was found in Refs. 8 and 9 that the squeezing of light increases significantly in the transition to chaos during the time interval for which quantum–classical correspondence is well-defined.[12] This result was illustrated in Refs. 8 and 9 by the example of the generalized Janes–

Cummings model, which allows a transition from regular dynamics to chaotic dynamics in the classical limit.[13] Then this result was generalized to the case of arbitrary single-mode quantum-optical systems in Ref. 14. The squeezing of wave packets in quantum chaos was also briefly discussed in Ref. 10.

However, the main results of Refs. 8, 9 and 14 were obtained by using a form of perturbation theory (the $1/N$-expansion). In this connection it should be interesting to study the generation of squeezed states in the numerical solution of the Schrödinger equation proper for a simple quantum system that allows a transition to quantum chaos.

In the present paper we study the generation of squeezed states in the time evolution of an initially Gaussian wave packet in the model of a quantum rotator excited by a periodic sequence of kicks, called the kicked quantum rotator. The model was first introduced by Casati *et al.*[15] and at present is the main model in studies of quantum chaos (see, e.g., the review in Refs. 16–18). The quantum rotator model is attractive mainly for two reasons: first, the classical limit for this model is a well-studied standard map,[19] and second, in numerical calculations it is fairly easy to study the dynamics of the model in the quasiclassical region with a large number of quantum levels.

We examine the dynamics of narrow Gaussian packets in a rotator with 2^{17} ($\approx 10^5$) levels. We define squeezing for the generalized quadrature operator

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

where θ is a real parameter. It is this type of squeezing that is observed in the homodyne detecting scheme, where θ is determined by the phase of the reference beam.[1] We will see that as long as the wave packet is localized, the degree of squeezing correlates well with the degree of local instability in the system. Here the greater the instability, the stronger the squeezing achieved in a shorter time interval. Squeezing is much stronger in quantum chaos than it is in regular motion. We will also see that the narrower the initial wave packet, the higher the degree of squeezing that can be achieved. We attribute this to the fact that a narrow wave packet is closer in its evolution to the classical trajectory than a broad one, with the result that it is more sensitive to local instabilities in the motion, which leads to strong squeezing.

We will also consider the problem of stability and observability of squeezing in the transition to chaos. More precisely, we will study the time dependence of the optimum values of the phases θ of the generalized quadrature operator X_θ for which the squeezing is at its maximum (this is known as principal squeezing[20,21]). We will show that in strong chaos and in long time intervals the optimum values of the phases change dramatically even under a small perturbation of the parameters of the initial Gaussian packet. Such a squeezing regime is unstable and difficult to observe. On the other hand, our results suggest that in weak chaos squeezing is fairly stable.

We will also briefly discuss the dynamics of disintegration of wave packets in chaos. Here we will show that a typical scenario of disintegration of an initially localized wave packet in chaos consists of two stages: the initial spread of the packet, and the catastrophic disintegration of the packet into many small subpackets. Here our results agree on the whole with the results of Casati and Chirikov.[18]

Note that earlier the dynamics of narrow Gaussian packets in the quasiclassical region was studied numerically for the model of a quantum rotator with kicks,[22,23] and also the model of a kicked quantum top[24] and for the quantum-cat–Arnold model[25] in connection with the problem of quantum–classical correspondence in quantum chaos. However, in these papers the generation of squeezed states was not considered.

The model of a quantum rotator is extremely popular in theoretical studies of quantum chaos. On the other hand, recently possibilities of implementing variants of this model in optical systems have been discussed.[26] Moreover, the quantum rotator model has been realized in experiments in the interaction of laser light and cooled atoms.[27] Hence our results on the buildup of squeezing in the transition to quantum chaos in a rotator are also related to experimentally realizable systems.

The plan of this paper is as follows. In Sec. 2 we discuss the quantum map of the rotator model and find how to calculate principal squeezing. The method used in numerical calculations is developed in Sec. 3, and the main results in the dynamics of squeezing are given in Sec. 4. Finally, in Sec. 5 we draw the main conclusions and consider the possibility of verifying our results in experiments on squeezing buildup.

2. THE QUANTUM ROTATOR MODEL AND SQUEEZED STATES

Let us examined the model of a quantum rotator with periodic delta-function kicks. Here we follow the notation of Ref. 23. The Hamiltonian for such a model is

$$H = \frac{p^2}{2mL^2} - \delta_p(t/T)\, mL^2\omega_0^2 \cos x,$$

$$\delta_p(t/T) = \sum_{j=-\infty}^{+\infty} \delta(j - t/T), \tag{1}$$

where x is the cyclic variable with a period 2π, L is the characteristic size of the rotator, m is the rotator mass, and ω_0 is the frequency of linear vibrations. The function $\delta_p(t/T)$ describes a periodic sequence of kicks with a period T, where $\delta(x)$ is the Dirac delta function. Let us introduce new variables

$$\alpha = mL^2\omega_0^2 T, \quad \beta = \frac{T}{mL^2}, \tag{2}$$

and measure time in units of T, i.e., $t \to t/T$. Then the Schrödinger equation assumes the form

$$i\hbar \frac{\partial \Psi}{\partial t} = -\frac{\hbar^2 \beta}{2} \frac{\partial^2 \Psi}{\partial x^2} - \delta_p(t)\, \alpha \cos x \cdot \Psi. \tag{3}$$

Due to the periodicity of $\Psi(x)$ in x the solution of Eq. (3) can be written as follows:

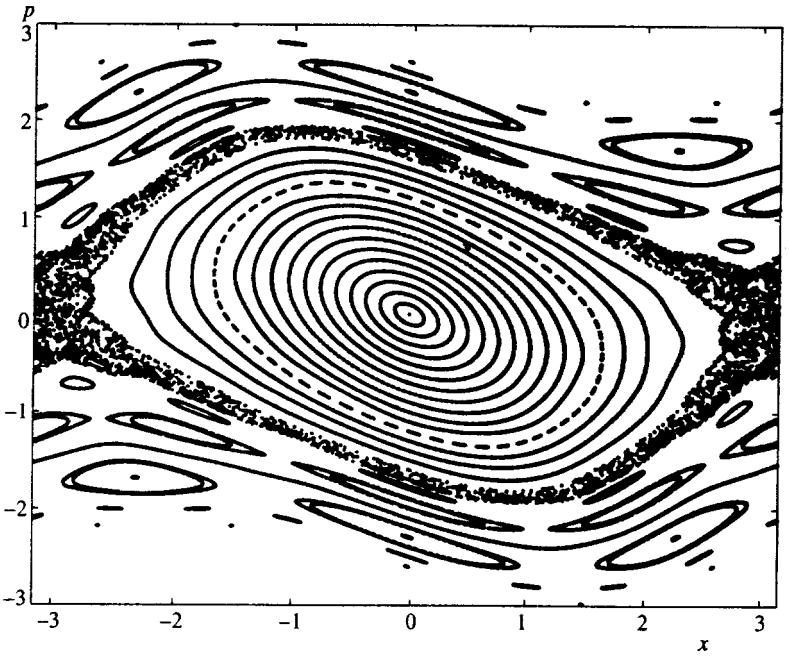


FIG. 1. Phase portrait of the classical standard map for $K=0.8$.

$$\Psi(x)=\frac{1}{\sqrt{2\pi}}\sum_{k=-\infty}^{+\infty}e^{ikx}A_k(t),$$

$$A_k(t)=\frac{1}{\sqrt{2\pi}}\int_0^{2\pi}\Psi(x)e^{-ikx}dx. \tag{4}$$

Using the standard procedure,[16,23] we obtain the quantum map in the form

$$\Psi_{n+1}=U_xU_p\Psi_n,$$

$$U_p=\exp\left(-\frac{i\beta}{2\hbar}\hat{p}^2\right),\quad U_x=\exp\left(\frac{i\alpha}{\hbar}\cos(\hat{x})\right), \tag{5}$$

where Ψ_n is the value of the wave function at the tome immediately after the nth kick. The time evolution of the wave function in the map (5) is determined solely by two parameters, $\alpha/\hbar$ and $\beta\hbar$. Since U_p is diagonalized in the p-representation, U_x is diagonalized in the x-representation, and the transition between x- and p-representations is given by the Fourier transformation (4), the map (5) actually reduces to

$$\Psi_{n+1}(x)=U_xF^{-1}U_pF\Psi_n(x), \tag{6}$$

where F and F^{-1} are the direct and inverse Fourier transforms.

Sometimes it proves useful to use the quantum map written in terms of the probability amplitudes A_k of transitions between the unperturbed levels of the rotator.[15] Combining (4) and (5), we obtain

$$A_k^{(n+1)}=\sum_{m=-\infty}^{+\infty}F_{km}A_m^{(n)},$$

$$F_{km}=(-i)^{k-m}\exp\left(-\frac{i\hbar\beta m^2}{2}\right)J_{k-m}\left(\frac{\alpha}{\hbar}\right), \tag{7}$$

where $J_l(z)$ is the Bessel function of order l and argument z, and the superscript (n) on the variable A stands for the number of the kick. Bearing in mind that the Bessel functions with $|l|\geqslant z$ rapidly decrease with increasing l, we see from (7) that, with exponential accuracy, in the course of a single kick $2\alpha\hbar^{-1}$ unperturbed rotator levels are captured. Below we consider the case where $\alpha/\hbar$ is large, which is typical of quantum chaos problems.

In the classical limit the Hamiltonian (1) reduces to the standard map

$$P_{n+1}=P_n-K\sin x_{n+1},\quad x_{n+1}=x_n+P_n\pmod{2\pi}, \tag{8}$$

where $P_n=\beta p_n$, with the subscript n denoting the values of x and P immediately after the nth kick, and $K\equiv\alpha\beta$ is the Chirikov parameter.[1)] Strong and global chaos sets in for $K>1$. For $K<1$ the larger part of the phase plane is filled with regular trajectories, although small regions with local chaos exist no matter how small K may be.[19] The phase portrait for the map (8) at $K=0.8$ is depicted in Fig. 1. The chaotic layer lies near the separatrix of the main resonance, which passes through the hyperbolic points $(\pm\pi,0)$. In our calculations we usually take a wave packet whose center of gravity lies near a hyperbolic point.

For the initial state of the quantum map (5) we take the Gaussian wave packet

$$\Psi(x)=(2\pi\sigma^2)^{-1/4}\exp\left(-\frac{(x-x_0)^2}{4\sigma^2}+ik_0(x-x_0)\right), \tag{9}$$

where

$$\langle x\rangle=x_0,\quad \langle\delta x^2\rangle\equiv\langle x^2\rangle-\langle x\rangle^2=\sigma^2,$$

$$p_0\equiv\langle p\rangle=\hbar k_0,\quad \langle\delta p^2\rangle=\hbar^2/4\sigma^2,$$

and k_0 is an integer. The packet is assumed narrow:

$$\sigma \ll x_0, \quad \langle \delta p^2 \rangle \ll \hbar k_0.$$

Note that in view of its periodicity in x the wave packet (9) is generally not a state that minimizes the uncertainty relation. But in the case of a narrow packet it is essentially indistinguishable from a minimum-uncertainty state.[22–24]

A typical initial quantum state in studies of light squeezing is a coherent state.[1–3] Such a state is an eigenfunction of the annihilation operator a, which in the present notation can be written as

$$a = \frac{1}{\sqrt{2\hbar}} \left(\sqrt{\gamma}\hat{x} + i \frac{\hat{p}}{\sqrt{\gamma}} \right), \quad \gamma = \left(\frac{\alpha}{\beta} \right)^{1/2}. \tag{10}$$

The fact that the annihilation operator has such an appearance can easily be understood if we consider the following limiting case of the harmonic oscillator that follows from (3):

$$i\hbar \frac{\partial \Psi}{\partial t} = -\frac{\hbar^2 \beta}{2} \frac{\partial^2 \Psi}{\partial x^2} + \frac{\alpha x^2}{2} \Psi. \tag{11}$$

Now we can show that the wave function (9) is a coherent state, i.e., an eigenfunction of (10), if we put

$$\sigma^2 = \frac{\hbar}{2\gamma}. \tag{12}$$

Let us now turn to the problem of squeezing.

In light squeezing experiments,[1] the observable quantity is the variance of the generalized quadrature operator

$$X_\theta = a e^{-i\theta} + a^\dagger e^{i\theta}, \tag{13}$$

where θ is the phase of the reference beam in the homodyne detecting scheme. In the particular cases where $\theta = 0$ or $\theta = \pi/2$ Eq. (13) yields the following expressions for the generalized position and momentum operators:

$$X_1 = a + a^\dagger, \quad X_2 = -i(a - a^\dagger), \quad [X_1, X_2] = 2i, \tag{14}$$

with the uncertainty relation $\langle \delta X_1^2 \rangle \langle \delta X_2^2 \rangle \geqslant 1$, where averaging is done over an arbitrary quantum state and equality is achieved for a coherent state. The standard definition of quadrature squeezing is the condition[1,3]

$$\min(\langle \delta X_1^2 \rangle, \langle \delta X_2^2 \rangle) < 1, \tag{15}$$

i.e., the variance of one of the quadrature components is smaller than for the coherent state.

In a more general case we consider the variance $\langle \delta X_\theta^2 \rangle$ of the operator (13), and the state is assumed squeezed if the value of $\langle \delta X_\theta^2 \rangle$ in this state for some value of θ is smaller than in the coherent state.[20,21] Experiments actually determine the minimum S of this variance as a function of the angle θ:

$$S = \min_{\theta \in [0, 2\pi]} \langle \delta X_\theta^2 \rangle. \tag{16}$$

Using the definition (13) of X_θ, we can show[20,21] that

$$S = 1 + 2\langle \delta a^\dagger \delta a \rangle - 2\sqrt{\langle \delta a^2 \rangle \langle \delta a^{\dagger 2} \rangle}, \tag{17}$$

and the minimum of $\langle \delta X_\theta^2 \rangle$ is reached at an optimum phase value $\theta = \theta^*$ defined as follows:[21]

$$e^{2i\theta^*} = -\sqrt{\frac{\langle \delta a^2 \rangle}{\langle \delta a^{\dagger 2} \rangle}}. \tag{18}$$

For our discussion it is convenient to express S in terms of the cumulants of the operators x and p. Using the definition (10) of operator a and Eq. (17), we obtain

$$S = \frac{1}{\hbar} \left(\frac{\langle \delta p^2 \rangle}{\gamma} + \langle \delta x^2 \rangle \gamma - \sqrt{(\langle \delta x^2 \rangle \gamma - \langle \delta p^2 \rangle / \gamma)^2 + 4c^2} \right), \tag{19}$$

where

$$c = \tfrac{1}{2}(\langle (xp + px) \rangle - 2\langle x \rangle \langle p \rangle).$$

Clearly each Gaussian packet satisfies $S = \hbar / 2\sigma^2 \gamma$, while for a coherent state we have, in view of (12), $S = 1$. Hence a state is squeezed if

$$S < 1. \tag{20}$$

The condition determines the principal squeezing attainable in homodyne detecting.[20]

The maximum of the variance $\langle \delta X_\theta^2 \rangle$ in θ can be defined in the same way the minimum was defined in (16). We denote it by $\overline{S}$. Then we can show that the dependence of $\overline{S}$ on the cumulants differs from (19) only in the sign in front of the square root, so that we have

$$S\overline{S} \geqslant 1. \tag{21}$$

Thus, squeezing in S (Eq. (20)) is accompanied by dilation in $\overline{S}$.

Note that in contrast to the quadrature squeezing (15), the definition (19) of principal squeezing contains quadrature correlators of the $\langle xp \rangle$ type. This is very important for systems with discrete time, to which the model of a quantum rotator excited by kicks belongs. The thing is that the quadrature squeezing (15) is essentially unobservable in such systems, although the principal squeezing (19) and (20) may occur.[2)] In Sec. 4 we discuss the time dependence of S.

3. THE NUMERICAL METHOD

Several features of the numerical method must be mentioned. The interval in x from 0 to 2π is partitioned into N segments $\Delta x = 2\pi/N$, and the wave function $\Psi(x)$ is represented by a discrete sequence of values (column vector $|\Psi\rangle$) of length N, so that $\Psi_l = \Psi(l\Delta x)$, $l \in [0,1,\cdots, N-1]$. Accordingly, in the sum in (4) k varies from 0 to $N-1$. In our numerical method N is an integral power of two. Here the operator F in (6) is interpreted as the fast Fourier transform, which induces the following transformations:

$$F\colon \Psi_l \to A_k, \quad F^{-1}\colon A_k \to \Psi_l. \tag{22}$$

To determine the principal squeezing, we must calculate $\langle \delta q^2 \rangle$, $\langle \delta p^2 \rangle$, and $\langle xp \rangle$ (see Eq. (19)). For instance, the calculation of $\langle xp \rangle$ proceeds along the following lines:

$$\langle xp \rangle = \langle \Psi | \mathbf{x} F^{-1} \mathbf{p} F | \Psi \rangle,$$

where $\langle\Psi|$ is obtained by transposing the vector $|\Psi\rangle$ and then finding the complex conjugate of the result, while $\mathbf{x}$ and $\mathbf{p}$ are vectors that initially have the form

$$\mathbf{x}=[0,\ \Delta x,\ 2\Delta x,\cdots,\ 2\pi-\Delta x],$$

$$\mathbf{p}=[0,\ 1,\ 2,\cdots,\ N-1].$$

The fact that x is defined modulo 2π requires following the wave packet and ensuring that it is defined correctly during the passage through the end-points of the interval $[0,\ 2\pi]$. We set up the process in the following manner. When the center of the wave packet in the x-representation approaches an edge of the half-interval $[0,\ 2\pi]$, the wave function $\Psi(x)$ is examined on a new interval, $[-\pi,\pi]$, with a new vector

$$\mathbf{x}=[0,\Delta x,\cdots,\pi,-\pi+\Delta x,-\pi+2\Delta x,\cdots,-2\Delta x,-\Delta x],$$

since $(-k\Delta x)\bmod 2\pi=(2\pi-k\Delta x)\bmod 2\pi$, where k is an integer. The transition from $[-\pi,\pi]$ to $[0,2\pi]$ is treated similarly.

Calculations in the p-representation have their own special features. For instance, although for the Hamiltonian (1) the momentum is defined in the interval from $-\infty$ to $+\infty$, in numerical calculations we deal only with a finite range of values of momentum p, a range specified by the number N of Fourier transforms in the expansion (4). To avoid the possible problem of reflection of the wave packet from an edge of the given interval in the p-representation,[3)] we select this interval in each iteration of map (6) in such a way that the maximum of the absolute value of the wave function of the packet is always at the center of the given interval (actually, we renumber the vector $\mathbf{p}$).

The process of calculating the next iteration of the quantum map (6) is terminated as soon as the packet ceases to be sufficiently localized either in the x-representation or in the p-representation, i.e., when the number of Fourier transforms actually involved in the calculation process is smaller than needed. We write the conditions for packet delocalization mentioned earlier. To this end we introduce the notation

$$\xi=\max_{[0,2\pi]}|\Psi(x)|,\quad \chi=\max\{|A_1|,|A_2|,\cdots,|A_N|\},$$

and A_{left} and A_{right} are the values of A_k belonging, respectively, to the left and right edges of the finite interval in which the wave function in momentum space, the finiteness being due to the finite number N of Fourier transforms in the expansion (4). The calculation is terminated when one of the two inequalities,

$$\max\left\{\frac{|A_{\text{left}}|}{\chi},\frac{|A_{\text{right}}|}{\chi}\right\}>\varepsilon\quad\text{or}\quad\frac{|\Psi(z)|}{\xi}>\varepsilon,$$

is valid (here $z=0$ if $x\in[0,2\pi]$ or $z=\pi$ if $x\in[-\pi,\pi]$. In this paper we used the value $\varepsilon=0.002$).

4. THE MAIN RESULTS

For the initial wave function in our calculations we took the coherent state (a Gaussian wave packet) with $\hbar=10^{-6}$ and $k_0=10\,000$, and σ was varied between 0.04 and 0.07. We fixed the initial width σ of the wave packet and the Chirikov parameter K, in terms of which the parameters α and β in the evolution operator (5) are expressed as follows:

$$\alpha=K^{1/2}\frac{\hbar}{2\sigma^2},\quad \beta=K^{1/2}\frac{2\sigma^2}{\hbar}. \tag{23}$$

These formulas are obtained by combining the definition $K=\alpha\beta$ and Eqs. (10) and (12).

In Sec. 2 we found that the number of unperturbed-rotator levels captured in one kick is roughly $2\alpha/\hbar$. From (23) it follows that in our case this number is $K^{1/2}/\sigma^2$ and amounts to several tens of thousands for the adopted widths σ of the wave packet.

In our calculations K was varied between 0.2 and 2 with a step of 0.02. We found the time dependence of the squeezing S (19) and the optimum value of the phase θ^* at which $\langle\delta X_\theta^2\rangle$ is at its minimum. To demonstrate the correlation that exists between the degree of squeezing and the chaos characteristics[9,14] we calculated

$$d=[\langle\delta x^2\rangle+\langle\delta p^2\rangle]^{1/2}. \tag{24}$$

It can be shown[9,14,28] that in the classical limit and while the wave packet is well-localized, i.e., $[\langle\delta x^2\rangle]^{1/2}\ll x_0$ and $[\langle\delta p^2\rangle]^{1/2}\ll p_0$, the d of (24) corresponds to the following separation in phase space:

$$d_{\text{cl}}(t)=[(\Delta x)^2+(\Delta p)^2]^{1/2}, \tag{25}$$

where $(\Delta x(t),\ \Delta p(t))$ is the solution of the linear small-perturbation equations near the classical trajectory $(x(t),\ p(t))$. The quantity $d_{\text{cl}}(t)$ characterizes the divergence of two initially close trajectories and enters into the definition of the largest classical Lyapunov exponent

$$\lambda=\lim_{t\to\infty}\frac{d_{\text{cl}}(t)}{t}. \tag{26}$$

For a classical standard map with strong chaos $K\gg 1$ we have the simple dependence $\lambda\approx\ln(K/2)$ (see Ref. 19). The Lyapunov exponent (26) is an asymptotic characteristic of chaos. For finite time intervals[7]

$$d_{\text{cl}}(t)\approx\exp(h(x,p)t), \tag{27}$$

where the exponent h is a function of a point in phase space and coincides, in order of magnitude, with the Lyapunov exponent λ, but in some time intervals the difference between the two may be significant. The latter fact can be explained by the strong inhomogeneity in the statistical properties of the phase space of chaotic systems and, correspondingly, by the different rates of divergence of trajectories in different regions of phase space through which the system passes in its time evolution. It must be noted at this point that the dependence of h on the parameter K is extremely complicated. What is important, however, is only the property of the strong (exponential) increase of d_{cl} specified by (27) in the presence of chaos, a property often called local instability.[7] When the motion is regular, the time dependence of d_{cl} is much weaker—it follows a power function.[7]

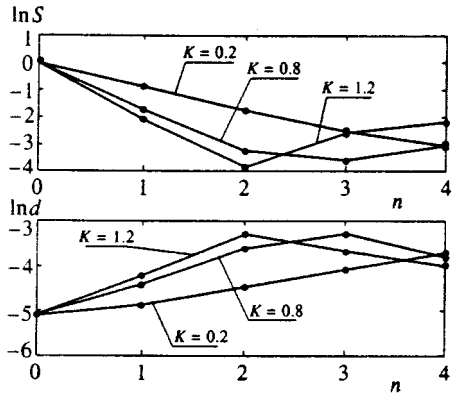


FIG. 2. Time dependence of the logarithm of squeezing S (the upper part of the figure) and $\ln d$ defined in Eq. (24) (the lower half of the figure); $x_0=\pi$ and $\sigma=0.006$.

On the other hand, it is h that determines the rate of phase-volume deformation: the stronger the local instability, the greater the phase-volume deformation in a given time interval.

Since in our case quantum–classical correspondence and the concept of chaos are well-defined only in a very short time interval, while the wave packet remains localized, it is meaningful to consider the correlations existing between the time dependence of the squeezing and that of the quantity d (see (24)), which in the classical limit becomes $d_{\rm cl}$ (see (25)).

Figure 2 depicts the time dependence of the logarithm of squeezing S and $\ln d$ for different values of K, when the center of gravity is of the wave packet is initially at the point $x_0=\pi$, $p_0=\hbar k_0=0.01$. This initial condition is close to a hyperbolic point through which the chaotic layer passes even when K is small (see Fig. 1). Figure 2 shows that the larger the squeezing (the smaller the value of S) the larger the local instability (the larger the values of $\ln d$) up to $n\approx 4$, when the packet spread becomes so large that purely quantum effects become important.

For another initial condition, $x_0=\pi/2$ and $p_0=0.01$, which is closer to an elliptic point and hence lands in the chaotic region only at large values of K, the dynamics of squeezing is depicted in Fig. 3. We see that in this case squeezing is stronger by a factor of almost two than under the conditions of Fig. 2 in the same time interval. On the other hand, both Fig. 2 and Fig. 3 exhibit an increase in squeezing as a function of the parameter K, which controls the development of chaos in the system.

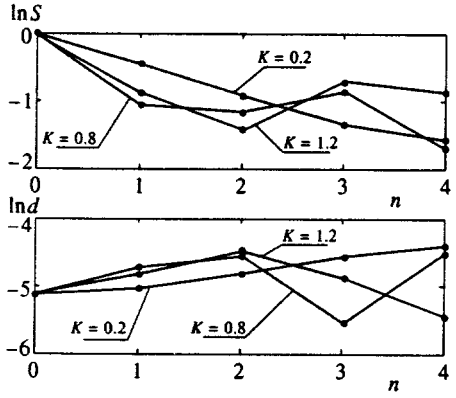


FIG. 3. The same as in Fig. 2 but for $x_0=\pi/2$.

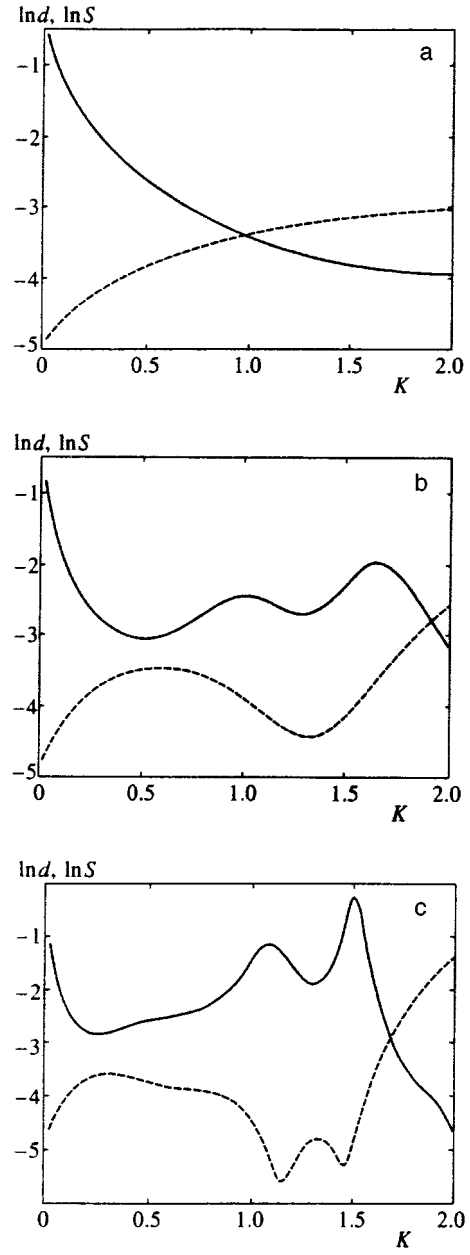


FIG. 4. Logarithm of the squeezing S (solid curves) and $\ln d$ (dashed curves) as functions of the Chirikov parameter K for a fixed number of kicks: (a) $n=3$, (b) $n=4$, and (c) $n=5$; $x_0=\pi$ and $\sigma=0.007$.

Let us study the correlation between squeezing and the degree of local instability in the system for different values of K in greater detail. The K-dependence of the degree of squeezing calculated after a fixed number of kicks at $x_0=\pi$ and $p_0=0.01$ is depicted in Fig. 4. After the third kick the correlation between $\ln S$ and $\ln d$ become very evident (Fig. 4a). However, small discrepancies in this dependence may appear as the number of kicks grows. Such discrepancies become evident, for instance, after the fourth kick for $1.1\lesssim K\lesssim 1.4$ (Fig. 4b). After five kicks, $n=5$, the correlation between $\ln S$ and $\ln d$ is restored (Fig. 4c). Note that this

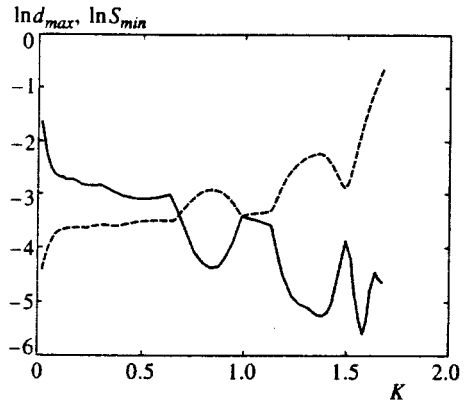


FIG. 5. Logarithm of the minimum squeezing $S_{\min}$ (solid curves) and of the local instability $d_{\max}$ (dashed curves) as functions of the Chirikov parameter K after seven kicks. The parameters and initial conditions are the same as in Fig. 4.

behavior pattern is quite typical. Hence, to establish the correlation between local instability and squeezing more clearly, a certain procedure of coarsening (averaging) these quantities in the given time interval is needed. In our study we determine the minimum squeezing $S_{\min}$ in a time interval during which the packet remains well-localized for most values of K considered here, and hence the maximum $d_{\max}$ in the same time interval. We found that there is a distinct correlation between $S_{\min}$ and $d_{\max}$: the larger the value of $d_{\max}$ the smaller the value of $S_{\min}$, and vice versa. An example of such a dependence is depicted in Fig. 5, where $S_{\min}$ and $d_{\max}$ were calculated after six kicks. Note that the diagrams do not go farther than $K>1.7$ because after six kicks the wave packet becomes delocalized for $K>1.7$ and calculating averages and local instability becomes meaningless.

We also studied the dependence of the dynamics of squeezing on the initial width σ of the wave packet. The results are depicted in Fig. 6. Clearly, the narrower the packet the stronger the squeezing achieved in a fixed time interval. This dependence arises because a narrow wave packet travels farther along its classical trajectory than a wide packet, so that it undergoes stronger deformations related to nonlinear classical dynamics. The exponential decrease is replaced by growth when the wave packet departs from the classical limit and the dynamics is of an essentially quantum nature.

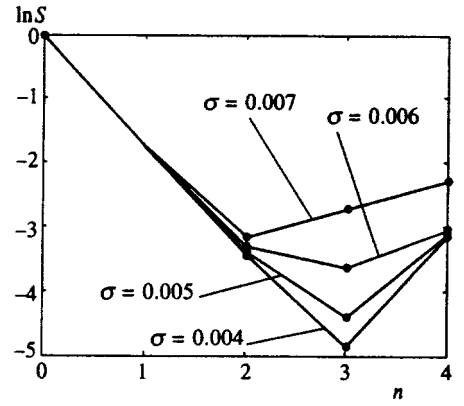


FIG. 6. Time dependence of the logarithm of squeezing S for different initial widths σ of the wave packet at fixed $K=0.8$; $x_0=\pi$.

Now let us examine the problem of stability and observability of squeezing in chaos. The figures mentioned earlier can serve to illustrate the statement that the stronger the chaos the stronger the principal squeezing. However, the definition (19) of principal squeezing is related to fixing the phase, $\theta=\theta^*$. Here θ^* is time-dependent even for exactly integrable systems.[21] When chaos is strong, the time dependence of $\theta^*(t)$ in the classical limit may be extremely complicated. Indeed, in addition to dilation and squeezing, the main feature of chaos in classical systems with a bounded phase space is the multiple formation of folds of the phase volume as chaos evolves.[7] Hence the process of finding the ''minimum width'' of a phase drop, which actually amounts to finding the θ^* vs. t dependence in the quasiclassical limit, becomes unstable for large time intervals.

Basing our reasoning on a similar semiclassical picture, we examined the stability of the time dependence of the optimum phase $\theta^*(t)$ calculated quantum mechanically with a small perturbation of the initial position of the wave packet. More precisely, we found the time dependence of the optimum phase θ_1^* with the initial condition $x_0=\pi$ and, similarly, $\theta_2^*(t)$ with the initial condition $x_0=\pi-0.05$. We denote the difference of these phases by

$$D(t)=\theta_1^*(t)-\theta_2^*(t).$$

Since θ^* is periodic with a period π (see Eq. (18)), it is natural to take $\sin 2D$ as the quantity of interest, since in this way we avoid breaks in the diagrams related to the periodicity of θ^*. The dependence of $\sin 2D$ on the Chirikov parameter K for different fixed numbers of kicks is depicted in Figs. 7a–7c. After two kicks (Fig. 7a) the maximum value of $|\sin 2D|$ does not exceed 0.035 at $K=2$. After three kicks (Fig. 7b) the value of D becomes significant at $K\gtrsim 1.2$. Finally, after four kicks (Fig. 7c) the process of measuring squeezing becomes essentially unstable at $K\gtrsim 1$. Indeed, in these condition with a small perturbation of the initial position of the wave packet, the difference of the optimum phases reaches a value of order π. In Ref. 9 such generation of squeezed states was called unstable squeezing. As Fig. 7 implies, unstable squeezing is observed when chaos is strong and the time intervals are such that semiclassical description is valid. On the other hand, for short time intervals and small K's the squeezing is strong and stable.

To conclude this section we will briefly touch on the problem of the dynamics of disintegration of coherent states in chaos, a problem that is of interest by itself. Figures 8a and 8b depict the dependence of $|\Psi|$ on x and of $|A_k|$ on k (see Eq. (4)). Actually, Fig. 8 gives the shape of the wave function in the coordinate and momentum representations for an initially narrow wave packet with $[\langle \delta x^2\rangle]^{1/2}(t=0)\equiv\sigma=0.006$ and $[\langle \delta p^2\rangle]^{1/2}(t=0)=\frac{1}{12}\times 10^{-3}$. The relatively small value $K=1.2$ makes it possible to examine the fairly long evolution of the wave packet up to the point of its total disintegration.[4)] After six kicks (Fig. 8a) the wave packet spreads out significantly, but on the whole retains its bell-shaped structure. What follows is a

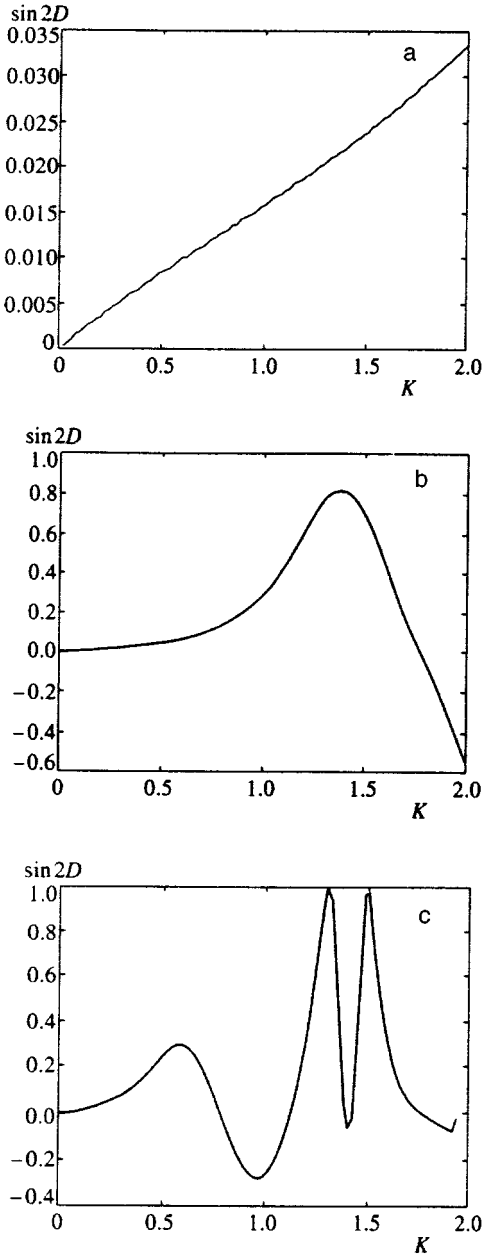


FIG. 7. The difference D of optimum phases as a function of the parameter K at $\sigma=0.006$, $x_0=\pi$, and a fixed number of kicks: (a) $n=2$, (b) $n=3$, and (c) $n=4$.

disintegration of the packet into many small packets, with the characteristic shape of the wave function depicted in Fig. 8b (after 18 kicks). Finally, very soon the wave function becomes so dissected that even 2^{17} Fourier harmonics are insufficient to describe the evolution correctly (for the data of Fig. 8 this happens approximately at the 20th kick). Qualitatively, the same pattern of the evolution of the wave packet was observed at higher values of K: first the broadening, or ''swelling,'' of the wave packet, and then its rapid disintegration into many very small subpackets. The differences in packet disintegration for large values of K in comparison with the case $K\simeq 1$ (Fig. 8) boil down to two facts: first, the swelling of the packet and the disintegration occur very rapidly (it takes only several kicks to complete the process), and second, the emerging subpackets are extremely small. Hence the process of disintegration of wave packets in strong chaos resembles an explosion. On the whole, the pattern being described agrees well with the pattern obtained from the analysis of the behavior of the Wigner function,[18] although we observed some anomalies. In particular, for fairly narrow wave packets ($\sigma=4\times10^{-3}$) we observed the disintegration of the initial packet into two fairly large subpackets. Ripples then appeared on the subpackets, and the two disintegrated into many small packets.

A more detailed description of the disintegration of coherent states in chaos requires further investigations.

5. DISCUSSION AND CONCLUSION

Thus, in this work we have used a numerical experiment to study the dynamics of generation of squeezed states in the evolution of a Gaussian packet in the quasiclassical limit for the model of a quantum rotator excited by kicks. We show that within the time interval where the packet is well-localized the squeezing becomes stronger in the transition to chaos. For strong chaos and in long time intervals the squeezing process becomes unstable. These results, obtained through direct numerical simulation, are in good agreement with the results obtained by perturbation-theoretic techniques and for other models.[8,9,14]

In the final stages of preparing the manuscript for press we became acquainted with two recent papers[29] also devoted to the problem of generating nonclassical states (squeezing and antibunching) in quantum chaos. Rui-Hue Xie and Gong-ou Xu[29] presented the results of numerical experiments on the dynamics of quadrature squeezing in simple quantum models that allow a transition to chaos in the classical limit: the Lipkin–Meshkov–Glick model[30] and the Belobrov–Zaslavskiĭ–Tartakovskiĭ model.[31] In contrast to our approach, Rui-Hue Xie and Gong-ou Xu[29] were interested in the long-time limit, when the wave packets are delocalized and this sense the quantum–classical correspondence is completely violated. They found that quadrature squeezing disappears in the transition to quantum chaos, although to some degree squeezing is always present in regular motion. It must be noted at this point that Rui-Hue Xie and Gong-ou Xu[29] noticed the existence of nonzero squeezing of some sort in the short-time limit and for quantum chaos, but they did not observe the buildup of squeezing described in the present paper, probably because in their numerical experiments[29] the quasiclassicality parameter was not sufficiently large: only several hundred quantum levels participated in the dynamics of the system. Thus, their results do not contradict ours and augment them in another limiting case, the limit of long times of motion. The description of the dynamics of squeezing in the case intermediate between the one described in the present paper and the one studied in Ref. 29 merits a separate investigation.

In conclusion we would like to make several remarks concerning the possibility of experimentally observing squeezing in quantum chaos on a time scale corresponding to a well-defined quantum–classical correspondence. At present essentially all squeezed-light experiments are done in the stationary regime. Squeezing in the transition to quantum

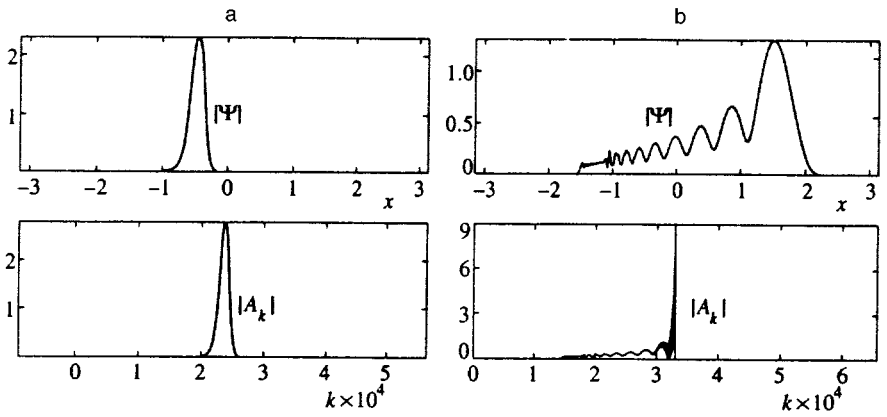


FIG. 8. Dependence of the absolute value of the wave function, $|\Psi|$ on x and dependence of the absolute value of the Fourier transforms, $|A_k|$, on k in the expansion (4) of the wave function at (a) $n=6$ and (b) $n=18$ and fixed $\sigma=0.006$, $K=1.2$, and $x_0=\pi$.

chaos builds up only over finite time intervals and in this sense is a transient dynamical phenomenon. The first experiments in light squeezing in transient regimes are only in the preliminary stage.[32] We hope that the development of effective experimental methods for observing squeezed states of light in transient dynamical regimes will also make it possible to observe the buildup of squeezing in the transition to quantum chaos.

On the other, as noted in the Introduction, it is much simpler to realized the quantum rotator model with kicks in atomic optics.[27] Moreover, it is much simpler to observe transient dynamical regimes in experiments with cooled atoms. Hence we believe that atomic-optics systems of the type discussed in Ref. 27 have great potential for observations of squeezed states in quantum chaos.

We are grateful to Claude Fabre, Jarmo Hietarinta, Zdenek Hradil, Juhani Kurkijärvi, Feodor Kusmartsev, Vlasta Peřinova, and Sergeĭ Turovets for stimulating discussions in stating the problem of squeezing and quantum chaos. We are also grateful to Andreĭ Kolovskiĭ and Jan Peřina for discussions and support at all stages of our work, and to Boris Chirikov for discussing the results and drawing our attention to his paper.[10] Preliminary investigation were done during the visit of one of the authors (K.N.A) to the Department of Physics, University of Illinois at Urbana-Champaign. K.N.A. is grateful to Prof. David Campbell for his hospitality. Partial support for this work was provided by the Russian Fund for Fundamental Research (Grant No. 96-02-16564a), the Krasnoyarsk Regional Scientific Fund (Grant No. 6F0030), INTAS (Grant No. 94-2058), and NATO (Linkage Grant No. 93-1602).

*)E-mail: kna@iph.krasnoyarsk.su

1)The unusual form of the standard map (8) is due to the fact that we find it convenient to take the values of x and p immediately after the nth kick rather than before the nth kick, as is done by the majority of researchers. It must be noted, however, that the properties of the standard map are retained.

2)Note that Lan[23] studied the time dependence of $\langle \delta a^2 \rangle$ and $\langle \delta p^2 \rangle$ for a quantum rotator on a time scale on which quantum–classical correspondence holds (Table I in Ref. 23). Both variances increase uniformly, so that quadrature squeezing (15) is impossible.

3)For a discussion of the problem of reflection and splitting of a wave packet due to the finite range of momentum in the close model of the quantum Arnold cat see Ref. 25.

4)In conducting numerical experiments in the dynamics of the disintegration of wave packets we did not use the procedure (described in Sec. 3) of terminating the counting process when the wave function becomes delocalized.

[1] D. F. Smirnov and A. S. Troshin, Usp. Fiz. Nauk **153**, 233 (1987) [Sov. Phys. Usp. **30**, 851 (1987)]; M. C. Teich and B. E. A. Saleh, Phys. Today **43**, No. 6, 26 (1990).
[2] S. Reynaud, A. Heidmann, E. Giacobino, and C. Fabre, in *Progress in Optics*, Vol. 30, E. Wolf (ed.), North-Holland, Amsterdam (1992), p. 1.
[3] C. Fabre, Phys. Rep. **219**, 215 (1992).
[4] A. Heidmann, J. M. Raimond, and S. Reynaud, Phys. Rev. Lett. **54**, 326 (1985).
[5] L. A. Lugiato, P. Galatola, and L. M. Narducci, Opt. Commun. **76**, 276 (1990).
[6] A. Heidmann, J. M. Raimond, S. Reynaud, and N. Zagury, Opt. Commun. **54**, 189 (1985).
[7] R. Z. Sagdeev, D. A. Usilov, and G. M. Zaslavskiĭ, *Nonlinear Physics: From the Pendulum to Turbulence and Chaos*, Harwood Academic, New York (1988).
[8] K. N. Alekseev, Preprint Kirensky Institute of Physics 674F, Krasnoyarsk (1991).
[9] K. N. Alekseev, Opt. Commun. **116**, 468 (1995).
[10] B. V. Chirikov, in Proceedings of the Second International Workshop on Squeezed States and Uncertainty Relations, Moscow, May 25–29, 1992, NASA Conference Publication 3219, 317 (1993).
[11] L. G. Yaffe, Rev. Mod. Phys. **54**, 407 (1982).
[12] G. P. Berman and G. M. Zaslavsky, Physica A **91**, 450 (1978); M. Berry, N. Balazs, M. Tabor, and A. Voros, Ann. Phys. (N.Y.) **122**, 26 (1979).
[13] K. N. Alekseev and G. P. Berman, Zh. Éksp. Teor. Fiz. **94**, No. 9, 49 (1988) [Sov. Phys. JETP **67**, 1762 (1988)]; **105**, 555 (1994) [**78**, 296 (1994)].
[14] K. N. Alekseev and J. Peřina, Phys. Lett. **231**, 373 (1997); submitted to Phys. Rev. E (1998).
[15] G. Casati, B. V. Chirikov, J. Ford, and F. M. Izrailev, in *Stochastic Behavior in Classical and Quantum Hamiltonian Systems*, G. Casati and J. Ford (eds.), Springer, Berlin (1979), p. 334.
[16] B. V. Chirikov, F. M. Izrailev, and D. L. Shepelyansky, Sov. Sci. Rev. C **2**, 209 (1981).
[17] *Chaos and Quantum Physics*, M. J. Giannoni, A. Voros, and J. Zinn-Justin (Eds.), Les Houches Session *LIL* 1989, Elsevier, Amsterdam (1991).
[18] G. Casati and B. V. Chirikov, Physica D **86**, 220 (1995).
[19] B. V. Chirikov, Phys. Rep. **52**, 263 (1979).
[20] A. Lukš, V. Peřinova, and J. Peřina, Opt. Commun. **67**, 149 (1988).
[21] R. Tanaś, A. Miranowicz, and S. Kielich, Phys. Rev. A **43**, 4041 (1991).
[22] R. F. Fox and B. L. Lan, Phys. Rev. A **41**, 2952 (1990); B. L. Lan and R. F. Fox, Phys. Rev. A **43**, 646 (1991); R. F. Fox and T. C. Elston, Phys. Rev. E **49**, 3683 (1994).
[23] B. L. Lan, Phys. Rev. E **50**, 764 (1994).
[24] R. F. Fox and T. C. Elston, Phys. Rev. E **50**, 2553 (1994).

[25] T. C. Elston and R. F. Fox, Preprint Georgia Inst. of Technol. (1994).
[26] J. Krug, Phys. Rev. Lett. **59**, 2133 (1987); R. E. Prange and S. Fishman, Phys. Rev. Lett. **63**, 704 (1989).
[27] F. L. Moore, J. C. Robinson, C. Bharucha, P. E. Williams, and M. G. Raizen, Phys. Rev. Lett. **73**, 2974 (1994); F. L. Moore, J. C. Robinson, C. F. Bharucha, B. Sundaram, and M. G. Raizen, Phys. Rev. Lett. **75**, 598 (1995).
[28] B. Sundaram and P. W. Milonni, Phys. Rev. E **51**, 1971 (1995).
[29] Rui-Hue Xie and Gong-ou Xu, Phys. Rev. E **54**, 1402 (1996); **54**, 2132 (1996).
[30] H. J. Lipkin, N. Meshkov, and A. J. Glick, Nucl. Phys. **62**, 188 (1965).
[31] P. I. Belobrov, G. M. Zaslavskiĭ, and G. Kh. Tartakovskiĭ, Zh. Éksp. Teor. Fiz. **71**, 1799 (1976) [Sov. Phys. JETP **44**, 945 (1976)].
[32] Claude Fabre, Université P. et M. Curie, Paris, France (private communication).

Translated by Eugene Yankovsky

Dissociation of molecular hydrogen ions by an IR laser pulse

E. A. Volkova, A. M. Popov, and O. V. Tikhonova

D. V. Skobel'tsyn Research Institute for Nuclear Physics, M. V. Lomonosov Moscow State University, 119899 Moscow, Russia

(Submitted 22 May 1997)
Zh. Éksp. Teor. Fiz. **113**, 128–143 (January 1998)

Dissociation dynamics of the simplest molecular systems, such as H_2^+, D_2^+, and HD^+ ions, in an intense IR laser field has been investigated by numerical modeling. An n-term approximation has been developed to describe the molecular system dynamics in an intense electromagnetic field. Calculations by the n-term approximation have been compared to an accurate numerical solution of the two-particle problem. The dissociation probability as a function of the frequency and intensity of radiation for different isotopes in a molecular hydrogen ion is discussed. A quasistatic model of molecule dissociation in an IR field has been suggested, and limits of its applicability have been determined. © *1998 American Institute of Physics.*
[S1063-7761(98)00901-9]

1. INTRODUCTION

Research in the dynamics of simple molecular systems in intense optical fields has attracted a lot of attention in recent years.[1–3] Given the relatively complex structure of molecular systems in comparison with atoms, the range of effects in molecules exposed to powerful laser pulses is quite considerable, in particular, because of the large difference between resonant frequencies of electron and nuclear subsystems in a molecule.

Theoretical description of molecules is usually based on the Born–Oppenheimer adiabatic approximation, which allows one to decouple electron and nuclear variables of a molecule. The possibility of applying this approximation to the dynamics of molecules under intense laser fields, however, has never been sufficiently investigated. Therefore a numerical solution of the problem of laser pulse action on the simplest molecular system, the hydrogen molecular ion H_2^+, beyond the adiabatic approximation is, undoubtedly, very interesting.[4–6] Such calculations performed on modern computers, however, require a lot of CPU time, and therefore cannot be routinely used in modeling molecules, even the simplest ones, under intense electromagnetic fields. For this reason, development of approximate models that would allow us to adequately describe interaction between strong electromagnetic fields and molecules is a topical issue.

In the reported work, we have investigated dissociation of several hydrogen molecular ions, namely H_2^+, D_2^+, and HD^+, by an intense IR laser field. The applicability of the adiabatic approximation to the simplest molecules has been investigated. The dissociation probability of H_2^+, D_2^+, and HD^+ ions as a function of the field amplitude and frequency has been calculated in the n-surface approximation. A quasistatic model of molecule dissociation by an IR field, which could allow us to interpret the intensity dependence of the dissociation probability, has been suggested, and its applicability limits have been determined.

2. MOLECULAR HYDROGEN ION. ONE-DIMENSIONAL MODEL

In this paper, we limit our discussion to a one-dimensional model of a molecular system, in which electron can move only along the molecule axis and the nuclei are allowed only to oscillate in this direction.

In the reference frame with origin at the center of mass, the molecule Hamiltonian H_0 has the form

$$H_0 = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} - \frac{\hbar^2}{2\mu}\frac{\partial^2}{\partial R^2} + V_e(x,R) + \frac{e^2}{R}, \tag{1}$$

where x is the electron coordinate, R is the internuclear distance, $\mu = \xi M$ is the molecule reduced mass, M is the proton mass, ξ is a factor equal to 1/2, 2/3, or 1 for H_2^+, HD^+, or D_2^+ ions, respectively. As in Ref. 6, the expression for the electron potential energy $V_e(x,R)$ was taken in the form

$$V_e(x,R) = -\frac{e^2}{\sqrt{(R/2-x)^2+\alpha^2}} - \frac{e^2}{\sqrt{(R/2+x)^2+\alpha^2}}, \tag{2}$$

where $\alpha = 0.943 a_0$, and a_0 is the Bohr radius.

Using the adiabatic approximation based on the smallness of m/M, one can construct a complete set of eigenfunctions for the Hamiltonian H_0 in the form[7]

$$\Phi_{nv}(x,R) = \phi_{nv}(R)\varphi_n(x,R), \tag{3}$$

where $\varphi_n(x,R)$ $(n = 1,2,3,\ldots)$ is the electron wave function obtained by solving the eigenvalue problem with the electron Hamiltonian:

$$H_e\varphi_n(x,R) = E_e^{(n)}(R)\varphi_n(x,R),$$

$$H_e = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V_e(x,R), \tag{4}$$

and $\phi_{nv}(R)$ is the nuclear wave function in the state corresponding to the nth potential surface (electron term), obtained by solving the equation

$$\left[-\frac{\hbar^2}{2\mu}\frac{\partial^2}{\partial R^2} + V_{\text{eff}}^{(n)}(R)\right]\phi_{nv}(R) = E_{nv}\phi_{nv}(R). \tag{5}$$

 1063-7761/98/010071-08$15.00

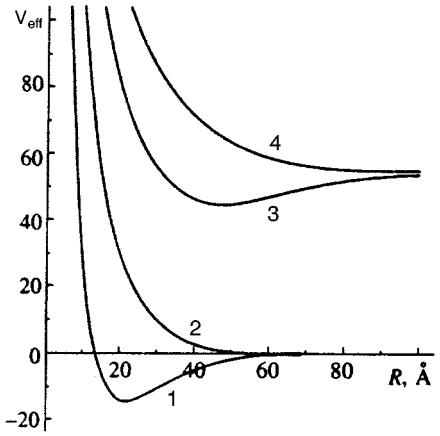


FIG. 1. Effective potential energy of interaction between nuclei in the model hydrogen molecular ion in electronic states $i=1,2,3,4$.

Here

$$V_{\text{eff}}^{(n)}(R)=\frac{e^2}{R}+E_e^{(n)}(R) \tag{6}$$

is the effective potential energy of interaction between nuclei, including the energy of electrons in the system, and $v=0,1,2,\ldots$ is the vibrational quantum number.

Figure 1 shows several curves of $V_{\text{eff}}^{(n)}(R)$ for the lowest potential surfaces calculated using Eqs. (4) and (6), and Table I lists energies E_{nv} of the ground-state potential surface $(n=1)$ for the three molecular ions studied. The full number of vibrational states in the discrete spectrum for this surface is 19, 22, and 25 for H_2^+, D_2^+, and HD^+, respectively.

3. INTERACTION WITH ELECTROMAGNETIC WAVES

In the dipole approximation, the interaction between a molecule and electromagnetic wave is described by the Hamiltonian

$$H=H_0-D\varepsilon(t)\cos\,\omega t, \tag{7}$$

where D is the dipole moment operator, $\varepsilon(t)$ is the wave electric field amplitude, which is a slow function of time, and ω is the wave frequency.

In the general case, the dipole moment D of a molecule can be expressed as

$$D=d_e+d_N, \tag{8}$$

where d_e and d_N are the dipole moments of the electron and nuclear subsystems, respectively.

TABLE I. Energy levels E_{1v} (eV) of modeled H_2^+, HD^+, and D_2^+ molecular ions on the ground-state potential surface.

V	H_2^+	HD^+	D_2^+
0	−1.355	−1.363	−1.373
1	−1.236	−1.259	−1.288
2	−1.120	−1.158	−1.204
3	−1.009	−1.060	−1.123

In the reference frame with origin at the center of mass, in the case of homonuclear molecules (H_2^+ and D_2^+), the nuclear dipole moment is identically zero, and the electric field acts only on the electronic subsystem:

$$D=d_e=-ex, \quad d_N=0. \tag{9}$$

Therefore transitions between various vibrational states of one electron term are forbidden, and higher vibrational states can be populated only through cascade transitions via other electronic states. In the HD^+ heteronuclear molecule, the center of positive charge is displaced with respect to the center of mass, so the nuclear dipole moment is nonzero, and the electric field acts directly on the nuclear subsystem. Taking into account the smallness $m/M\ll 1$, one can easily obtain the dipole moment of the nuclear subsystem for HD^+:

$$d_N=eR/3. \tag{10}$$

In our calculations, we have assumed that the electric field amplitude can be described by the following function of time:

$$\varepsilon(t)=\begin{cases} \varepsilon_0\dfrac{t}{\tau_f}, & t<\tau_f, \\ \varepsilon_0, & \tau_f\leqslant t<\tau+\tau_f, \\ \varepsilon_0\left[1-\dfrac{t-(\tau+\tau_f)}{\tau_f}\right], & \tau+\tau_f\leqslant t<\tau+2\tau_f, \end{cases} \tag{11}$$

where τ_f is the width of the leading and trailing edges of the pulse, and τ is the plateau width.

The photon energy of the electromagnetic wave was varied between $\hbar\omega=0.12$ and 0.96 eV, such that the times τ_f and τ were multiples of the wave period $T=2\pi/\omega$. In particular, for $\hbar\omega=0.12$ eV, which is the photon energy of a CO_2 laser, we took $\tau_f=2T$ and $\tau=5T$.

4. RESULTS AND DISCUSSION

4.1. Approximation of *n* potential surfaces. Comparison with exact calculations

In our model, the dynamics of a molecule in the field of an electromagnetic wave is described by a two-particle wave function $\Psi(x,R,t)$ which can be derived by solving the time-dependent Schrödinger equation

$$i\hbar\frac{\partial\Psi(x,R,t)}{\partial t}=[H_0-D\varepsilon(t)\cos\,\omega t]\Psi(x,R,t). \tag{12}$$

It was shown previously[6] that stationary functions of the system can be approximated to high accuracy by the wave functions obtained in the adiabatic approximation. Therefore the initial condition for Eq. (12) was defined as

$$\Psi(x,R,t=0)=\phi_{1v}(R)\varphi_1(x,R), \tag{13}$$

which corresponds to the electron ground state and vibrational state with index v. In most cases we took $v=0$.

A numerical solution of Eq. (12) with initial condition (13) for laser pulses in the visible and IR, obtained without resorting to the Born–Oppenheimer approximation, was calculated earlier.[8] In the visible, there is competition between

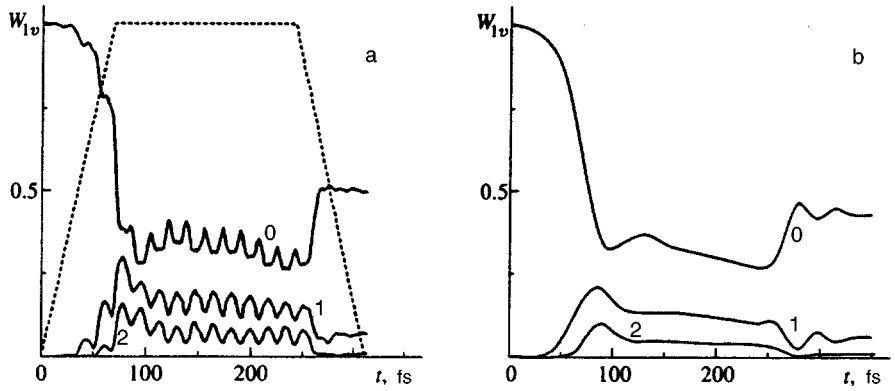


FIG. 2. Populations of various vibrational states of the molecule (ν=0,1,2) on the ground-state potential surface as functions of time during a laser pulse with $P=2\times10^{13}$ W/cm^2, $\hbar\omega=0.12$ eV, $\tau_f=2T$, and $\tau=5T$: a) four-surface approximation; b) exact calculation (high-frequency oscillations are averaged out in exact calculations). The dashed line shows the envelope of the laser pulse.

ionization and dissociation, whereas in the IR radiation there is a range of field amplitudes in which the ionization probability is negligible in comparison with the dissociation probability.

The calculation of an exact solution, however, requires a great deal of CPU time, so it seems reasonable to develop various approximate models. One of them is the n-surface model.

We expand the full wave function of a molecule $\Psi(x,R,t)$ in electron wave functions $\{\varphi_i(x,R)\}$ of different potential surfaces:

$$\Psi(x,R,t)=\sum_i \alpha_i(R,t)\varphi_i(x,R). \tag{14}$$

Then one can easily obtain the following equation system for the expansion coefficients $\alpha_i(R,t)$ in the adiabatic approximation[7]:

$$i\hbar\frac{\partial\alpha_i}{\partial t}=\left[-\frac{\hbar^2}{2\mu}\frac{\partial^2}{\partial R^2}+V_{\text{eff}}^{(i)}(R)-d_N\varepsilon(t)\cos\omega t\right]\alpha_i + \sum_{j\neq 1}\alpha_j W_{ij}, \tag{15}$$

where $W_{ij}=-d_{ij}(R)\varepsilon(t)\cos\omega t$, d_{ij} is the matrix element of the electron dipole moment, which contains the nuclear coordinate as a parameter, $V_{\text{eff}}^{(i)}$ is the effective potential energy of interaction between the nuclei in the electron i-th state defined by Eq. (6), and d_N is the nuclear component of the system dipole moment, which is nonzero only for HD^+.

The set of functions $\{\alpha_i(R,t)\}$ contains nuclear wave functions of nonstationary states for various potential surfaces in the molecule, and the last term in Eq. (15) describes transitions between different electron states of the molecule (transitions between different surfaces) induced by the electromagnetic field. If the summation in Eq. (14) includes integration over electron-continuum states, Eqs. (15) will be equivalent to the initial equation (12) as long as the adiabatic approximation is valid.

The possibility of truncating the system (15) and retaining a small number of electron states in expansion (14) allows us to simplify the problem considerably. It was shown in the earlier study of H_2^+ and D_2^+ ionization and dissociation[8] that at IR intensities of about 10^{13} W/cm^2, the ionization probability is negligible in comparison with the dissociation probability, so transitions to electron-continuum states can be ignored in Eqs. (14) and (15).

The initial conditions for Eq. (15) equivalent to Eq. (13) are expressed as

$$\begin{cases}\alpha_1(R,t=0)=\phi_{1v}(R),\\ \alpha_i(R,t=0)=0, \quad i=2,3,4,\ldots\end{cases} \tag{16}$$

If ionization can be neglected, the following normalization condition holds:

$$\sum_i W_i=1,$$

where

$$W_i=\int|\alpha_i(R,t)|^2dR \tag{17}$$

is the probability of detecting the molecule on the ith potential surface.

In order to determine the number of expansion terms that should be retained in Eq. (14), we have calculated the evolution of the H_2^+ molecule in the field generated by a CO_2 laser ($\hbar\omega=0.12$ eV) at intensities $P=4\times10^{12}-2\times10^{13}$ W/cm^2 in the n-surface approximation ($n=2,3,\ldots$), and using a more accurate model described in Refs. 6 and 8. The comparison between calculations by the exact two-particle model and in the n-surface approximation for $P=2\times10^{13}$ W/cm^2 has shown that at least four terms should be taken into account in Eq. (14). Only in this case did we have reasonable agreement between the results.

An important point is that although the population of excited states of the molecule with $n=3,4$ is low throughout the laser pulse action, they are important for accurate calculations of the populations in states $n=1,2$. Figure 2a shows calculated populations in the lowest vibrational states of the ground potential surface vs. time. The probabilities W_{1v} were calculated by expanding the function $\alpha_1(R,t)$ in eigenfunctions $\phi_{1v}(R)$, $v=0,1,2,\ldots$:

$$W_{1v}(t)=\left|\int\alpha_1^*(R,t)\phi_{1v}(R)dR\right|^2, \tag{18}$$

and they are also in good agreement with similar exact calculations.

TABLE II. Probability of vibrational state population of the H_2^+ molecule at the end of the laser pulse calculated in the n-surface approximation and exact two-particle model at $P=2\cdot10^{13}$ W/cm^2.

V	$n=2$	$n=3$	$n=4$	Exact model
0	0.869	0.641	0.488	0.430
1	4.64(−2)	6.25(−2)	7.43(−2)	7.17(−2)
2	1.85(−3)	2.19(−3)	1.34(−3)	1.41(−3)
3	3.07(−3)	7.34(−3)	7.81(−3)	6.90(−3)
$\Sigma_v W_{1v}$	0.924	0.761	0.595	0.537

Table II lists W_{1v} derived in the exact calculation and in the n-surface approximation $(n=2,3,4)$ for $P=2\times10^{13}$ W/cm^2 and $\hbar\omega=0.12$ eV at the moment when the laser pulse terminates. These data also illustrate insufficiency of the two-(three-)surface approximation for calculating populations of molecule vibrational states. Table 2 clearly shows that the inclusion of the third and fourth potential surfaces leads to a higher probability of molecule dissociation at the end of the laser pulse.

To sum up, in the approximation suggested for molecular ions, a solution of the exact two-particle Schrödinger equation on a two-dimensional mesh can be replaced with a solution of n one-dimensional equations, which makes the problem much easier if n is small. The number of potential surfaces that should be taken into account depends, naturally, on the laser-field parameters. Note, however, that in the studied range of parameters the two-surface approximation is in good agreement with exact calculations at such power densities ($P<7\times10^{12}$ W/cm^2) that the probability of molecule dissociation and its vibrational excitation are small.

Note also that in the n-surface approximation the molecule ionization presents a specific problem that is not discussed in this paper. An important point is that there is a range of field intensities in which ionization processes can be ignored.

4.2. Dissociation probability as a function of field intensity and frequency

Hereinafter dynamics of molecular systems will be analyzed in the four-surface approximation. The dissociation probability was calculated by the formula

$$W_D=1-\sum_v W_{1v},$$

where the sum is calculated over all vibrational states of the ground-state potential surface. Here we have taken into account the exact calculations, which indicate that the probability W_{3v} of populating states of the discrete spectrum of the excited bound-state term is negligible. This fact derives from the Frank–Condon principle and large difference between internuclear separations for the two surfaces (see also Ref. 8).

Curves of $W_D(P)$ for H_2^+, D_2^+, and HD^+ and the quantum energy $\hbar\omega=0.12$ eV are given in Fig. 3. It is clear that the dissociation probability for D_2^+, whose nuclear mass is greater than that of H_2^+, is smaller because of the lower velocity of the nuclear wave packet when the molecule is on the repulsive potential surface. On the other hand, the direct effects of the electromagnetic field on the nuclear subsystem in HD^+ leads to a higher dissociation probability than in H_2^+ and D_2^+ homonuclear molecules. These effects are especially important, since the CO_2 laser frequency $\hbar\omega=0.12$ eV is close to the molecular vibration frequency $\hbar\Omega=E_2-E_1\approx0.104$ eV. However, the strong anharmonicity of the molecule (Table 1) should lead to a reduction in the resonant effect at high vibrational quantum numbers.

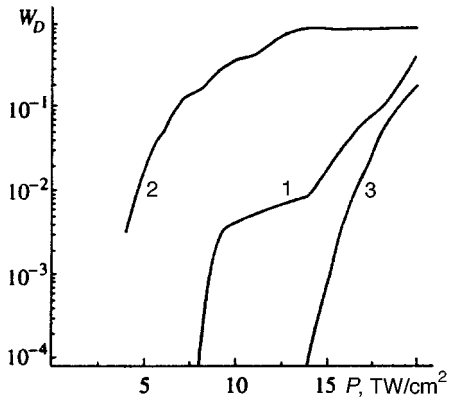


FIG. 3. Dissociation probability of 1) H_2^+, 2) HD^+, and 3) D_2^+ molecules vs. radiation intensity at $\hbar\omega=0.12$ eV.

The importance of the direct effects on the electronic degrees of freedom of HD^+, in addition to the nuclear variables, was demonstrated by omitting the electron dipole moment. This is equivalent to the analysis of the HD^+ dynamics in the one-surface approximation. In this case the dissociation probability is 0.059 at the radiation intensity $P=2\times10^{13}$ W/cm^2, which is even lower than the dissociation probability of D_2^+. Therefore, we can assert that the high dissociation probability of the HD^+ heteronuclear molecule, as compared to that of H_2^+ and D_2^+, is a result of the combined action of the electromagnetic field on the electron and nuclear subsystems.

Note also a characteristic step in the curve of $W_D(P)$ for H_2^+. Its cause will be discussed in the following section in terms of a quasistatic dissociation model.

We now proceed to the dissociation probability as a function of laser frequency. This function was studied at $\hbar\omega=0.12-0.96$ eV, for a constant shape and width of the laser pulse such that the pulse edges and plateau contained an integral number of field oscillations. Calculations for H_2^+ and HD^+ at $P=1.4\times10^{13}$ W/cm^2 are shown in Fig. 4. For both of these molecules, the dissociation probability at $\hbar\omega\geqslant0.5$ eV grows rapidly with laser frequency, which might be due to the smaller number of photons needed for a multiphoton transition between the ground and first excited electron states of the molecule. For HD^+ the dissociation probability also rapidly grows at $\hbar\omega\leqslant0.2$ eV, possibly due to the proximity to the vibration frequency of the nuclear subsystem. In the H_2^+ molecule, W_D also increases in this frequency band, and the population of vibrational states of the ground potential surface is also significant (Fig. 2). Since dipole transitions in the homonuclear molecules H_2^+ and D_2^+ within one potential surface are forbidden, these states can be

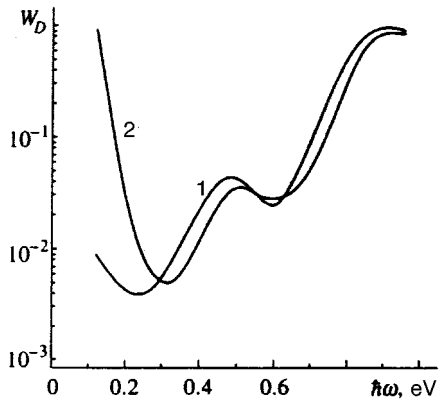


FIG. 4. Dissociation probability of (1) H_2^+ and (2) HD^+ molecules vs. optical photon energy at $P=1.4\times10^{13}$ W/cm^2.

populated only through multiphoton transitions via excited potential surfaces. But at $\hbar\omega=0.12$ eV the number of photons needed for the transition between the two lowest surfaces is about thirty. Given so high a number of photons needed for excitation and dissociation, we need an alternative approach to transitions between potential surfaces in a low-frequency electromagnetic field.

Two regions of intense dissociation of HD^+ molecules in the IR and UV bands were also detected by other authors.[9]

4.3. Quasistatic model of molecule dissociation in IR field

In this section, we discuss a quasistatic model of molecule dissociation in IR field that enables us to considerably simplify calculations of molecular dynamics in a low-frequency IR field and, in particular, interpret features in the dissociation probability of the H_2^+ ion as a function of field intensity and frequency.

Above all, note that the expansion of the molecular wave function $\Psi(x,R,t)$ in the system of electron states unperturbed by the electromagnetic field is not the only way of simplifying the problem (12). From the mathematical viewpoint, this can be done using any set of orthonormal functions, and its selection is a matter of mathematical convenience and physical insight. In intense fields at optical wavelengths, the preferable technique for describing molecular states is based on states modified by the field.[10] This concept yields satisfactory interpretations of various features of the above-threshold dissociation detected in experiments.[11–13]

In a low-frequency IR field, perturbation of potential surfaces of a molecule can be described by a quasistatic approximation based on the smallness of the laser field in comparison with the characteristic frequency of electron motion within the molecule.

The set of molecule electron states in a low-frequency field can be derived by solving the time-independent Schrödinger equation

$$[H_e-d_e\varepsilon(t)\cos\,\omega t]\varphi_n^{(\varepsilon)}(x,R,t)=E_n^{(\varepsilon)}(R,t)\varphi_n^{(\varepsilon)}(x,R,t), \tag{19}$$

where H_e is the electron Hamiltonian of the molecule.

The solution of Eq. (19) is a set of states $\{\varphi_n^{(\varepsilon)}(x,R,t)\}$ and corresponding energies $\{E_n^{(\varepsilon)}(R,t)\}$, which contain both the nuclear coordinate and time as parameters.

In moderate fields, mixing of all eigenstates of the electron Hamiltonian except the lowest two can be neglected in calculations of the molecule's ground potential surface. Then the energies of these states are determined, if we assume that the field is quasistatic, by the expression

$$E_{1,2}^{(\varepsilon)}(R,t)=\frac{1}{2}\,[E_1(R)+E_2(R)]\pm\hbar\Omega_R,$$

where

$$\Omega_R=\frac{1}{\hbar}\,\sqrt{\frac{1}{4}[E_1(R)-E_2(R)]^2+d_{12}^2\varepsilon^2(t)\cos^2\,\omega t}$$

determines the characteristic time of the system $\tau\sim1/\Omega_R$, and $E_1(R)$ and $E_2(R)$ are the energies of electron states at zero field amplitude. The quasistatic condition for the field is equivalent to

$$\omega<\Omega_R. \tag{20}$$

The resulting states $\varphi_n^{(\varepsilon)}(x,R,t)$ are quasistationary. We assume that the field intensity is sufficiently weak that the state decay time is larger than the laser pulse width. In the case under discussion, this condition is satisfied because molecule ionization can be ignored.

As noted above, it is assumed in obtaining the set of functions $\{\varphi_n^{(\varepsilon)}(x,R,t)\}$ that the electric field is a slow function of time, so that the electron wave function tunes up in accordance with the instantaneous strength of the wave electric field. This leads to the condition $\Delta\Omega_R/\Omega_R\ll1$, where $\Delta\Omega_R$ is the change in Ω_R over the characteristic time, and yields, if the laser pulse envelope is a slow function,

$$\omega\ll\frac{\hbar^2\Omega_R^3}{|d_{12}|^2\varepsilon^2(t)}. \tag{21}$$

Note that this condition is a criterion of the validity of the semiclassical approximation with respect to time, and this approximation can be used in solving the problem analytically.

The combination of conditions (20) and (21) in a strong electromagnetic field ($\Delta E/2<|d_{12}|\varepsilon(t)$) yields the condition $\hbar\omega\ll|d_{12}|\varepsilon_0$, i.e., the photon energy should be less than the separation between modified potential surfaces (Fig. 5). But condition (21) does not hold when the field intensity $\varepsilon(t)$ is almost zero. In this case, the quasistatic condition can be interpreted as the smallness of the probabilities of Landau–Zener transitions between surfaces. On the basis of the estimate in Refs. 15 and 16, we obtain

$$\hbar\omega\ll\frac{(\Delta E(R))^2}{|d_{12}(R)|\varepsilon_0}. \tag{22}$$

This estimate limits the range of R in which the quasistatic model holds. In this specific case, condition (22) is satisfied if $R\leqslant5$ Å.

The complete wave function $\Psi(x,R,t)$ of the system can be expanded in the set of functions (19),

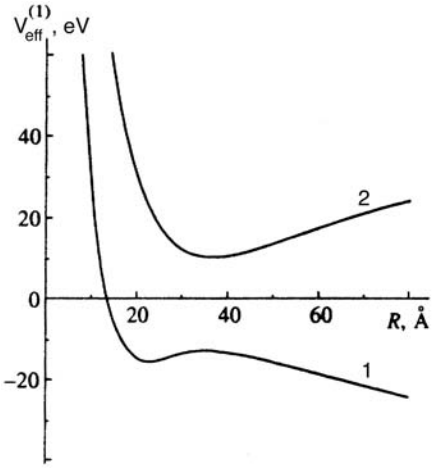


FIG. 5. Two lowest potential surfaces of the H_2^+ molecule in the presence of an electromagnetic field.

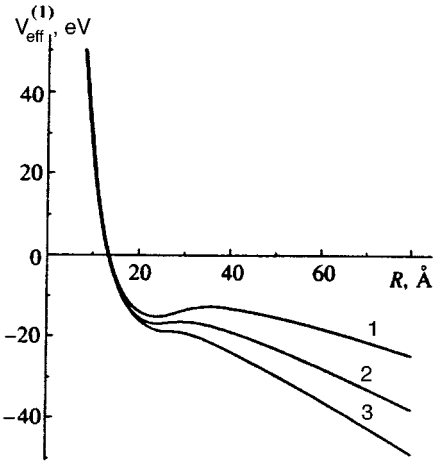


FIG. 6. Ground potential surface of the H_2^+ molecule modified by electromagnetic field of a wave with intensity (in W/cm^2) (1) 5.0×10^{12}; (2) 1.2×10^{13}; (3) 2.0×10^{13}. These curves correspond to maximum electric field intensity.

$$\Psi(x,R,t)=\sum_i \alpha_i^{(\varepsilon)}(R,t)\varphi_i^{(\varepsilon)}(x,R,t), \tag{23}$$

where $\alpha_i^{(\varepsilon)}(R,t)$ is the nuclear wave function of the molecule on the i-th potential surface modified by the slow field $\varepsilon(t)\cos\omega t$. By substituting expansion (23) into Eq. (12) and using the Born–Oppenheimer adiabatic approximation, we obtain the following equation system for functions $\alpha_i^{(\varepsilon)}(R,t)$:

$$i\hbar\frac{\partial\alpha_i^{(\varepsilon)}}{\partial t}+i\hbar\sum_n \alpha_n^{(\varepsilon)}\left\langle\varphi_i^{(\varepsilon)}\left|\frac{\partial\varphi_n^{(\varepsilon)}}{\partial t}\right\rangle\right.$$
$$=\left[-\frac{\hbar^2}{2\mu}\frac{\partial^2}{\partial R^2}+V_{\rm eff}^{(i)}(R,t)-d_N\varepsilon(t)\cos\omega t\right]\alpha_i^{(\varepsilon)}. \tag{24}$$

Here $V_{\rm eff}(R,t)=e^2/R+E_n^{(\varepsilon)}(R,t)$ is the molecule's potential surface modified by the low-frequency field.

Thus, we have obtained the equation describing the motion of a nuclear wave packet in the effective potential modified by the slow electromagnetic field of the wave. The second term on the left of Eq. (24) describes transitions between surfaces induced by the field.

Under certain conditions imposed on the field frequency, the second term on the left of Eq. (24) can be neglected (see Appendix for details). In this case we have

$$i\hbar\frac{\partial\alpha_i^{(\varepsilon)}}{\partial t}=\left[-\frac{\hbar^2}{2\mu}\frac{\partial^2}{\partial R^2}+V_{\rm eff}^{(i)}(R,t)\right.$$
$$\left.-d_N\varepsilon(t)\cos\omega t\right]\alpha_i^{(\varepsilon)}(R,t), \tag{25}$$

i.e., the Schrödinger equation for a nuclear wave packet in a time-dependent potential $V_{\rm eff}^{(i)}$. The elimination of the term mixing potential surfaces in Eq. (24) is similar to the Born–Oppenheimer approximation, since this enables us to decouple the equations of motion for nuclei in the absence of an external field.

The advantage of the approach discussed here is the possibility of replacing Eqs. (15) for the nuclear wave functions with the single equation (25). Indirect evidence for the validity and physical convenience of this approach is provided by the low population of excited electronic states with $n=3$ and 4, and by the fact that the population of all excited states is much lower than that of the molecule's ground-state potential surface by the end of the laser pulse.

The full set of terms $V_{\rm eff}^{(i)}(R,t)$ can be obtained by solving Eq. (19) at different moments in time. The function $V_{\rm eff}^{(i)}(R,t)$ has the form

$$V_{\rm eff}^{(1)}(R,t)=e^2/R+E_1^{(\varepsilon)}(R,t). \tag{26}$$

In a strong field, when $\Delta E\ll|d_{12}\varepsilon|$, we derive from Eq. (26)

$$V_{\rm eff}^{(1)}(R,t)=e^2/R+E_1(R)-|d_{12}\varepsilon(t)\cos\omega t|. \tag{27}$$

In the opposite case, when $|d_{12}\varepsilon|\ll E_2(R)-E_1(R)$, we obtain from the Eq. (26)

$$V_{\rm eff}^{(1)}(R,t)=\frac{e^2}{R}+E_1(R)-\frac{|d_{12}|^2\varepsilon^2(t)\cos^2\omega t}{E_2(R)-E_1(R)}. \tag{28}$$

It is clear that in both cases the effective potential $V_{\rm eff}^{(i)}(R,t)$ oscillates at a frequency that is twice the laser frequency. This effect[16] leads to repulsion of the nuclei during both half-periods of the field optical cycle.

Figure 6 shows the ground-state potential surface of the molecule in the optical electromagnetic field at various intensities $P=c\varepsilon_0^2/8\pi$ calculated by Eq. (26). In the range of small radiation intensities, $P\leqslant9\times10^{12}$ W/cm^2, the curve of $V_{\rm eff}^{(1)}(R,t)$ contains a section corresponding to classical finite motion, i.e., there is a possibility of quasistationary vibrational states of the molecule, which can decay as a result of tunneling across the potential barrier. At an intensity $P\approx1.0\times10^{13}$ W/cm^2 the potential barrier vanishes, and the curve of $V_{\rm eff}^{(1)}(R,t)$ has a plateau (curve *2*), i.e., a region where the force acting between the two nuclei, $F=-\partial V_{\rm eff}^{(1)}/\partial R$, is approximately zero. At higher field intensities the plateau disappears (curve *3*), which means that the nuclei repel one another at all separations R between them.

Figure 7 shows calculations of the dissociation probability of H_2^+ as a function of the radiation intensity obtained in

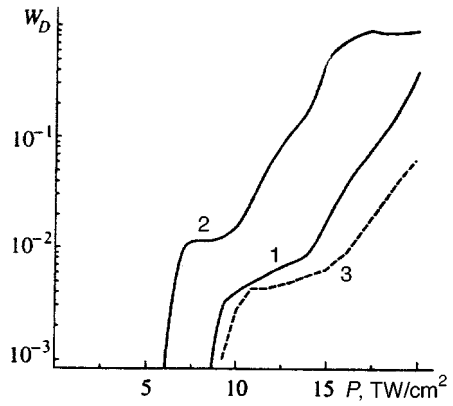


FIG. 7. Dissociation probabilities of H_2^+ in (1,3) ground vibrational state ($v=0$) and (2) excited vibrational state with $v=1$ as functions of radiation intensity. Curves *1* and *2* are obtained in the four-surface approximation, and curve *3* in the quasistatic approximation.

the quasistatic model and in the four-surface model. One can see that these curves are fairly close in the range $P \leqslant 1.5\times 10^{13}$ W/cm^2. At higher intensities the quasistatic model is no longer valid (see Eq. (A6)), and the effect of higher-energy surfaces perturbed by the laser field must be taken into account.

This model allows us to interpret the main features on the curve of $W_D(P)$. For example, at intensities $P \leqslant 10^{13}$ W/cm^2 the H_2^+ molecule dissociates as a result of tunneling of the nuclear wave packet across the potential barrier (curve *1* in Fig. 6). The probability of this process increases with field intensity. In the range $P \sim (1.0-1.5)\times 10^{13}$ W/cm^2 the nuclear wave packet is on the plateau of the $V_{\text{eff}}^{(1)}(R)$ curve, and the dissociation probability depends weakly on the field intensity. At higher intensities the dissociation probability again increases with intensity owing to the growth in the repulsive force $F=-\partial V_{\text{eff}}^{(1)}/\partial R$ on the protons throughout the laser pulse.

If the H_2^+ molecule is initially in an excited vibrational state, tunneling will probably occur in the range of lower field intensities. The effect of the plateau on the $V_{\text{eff}}^{(1)}(R)$ curve should also be less pronounced owing to the greater width of the nuclear wave packet. Such properties of the $W_D(P)$ function were detected in calculations of the molecule vibrational state $v=1$ (Fig. 7).

The model also enables us to understand the shape of the $W_D(\hbar\omega)$ curve for H_2^+ in the low-frequency band ($\hbar\omega \geqslant 0.12$ eV). In particular, the increase in the field frequency leads to shortening of the time during which the nuclear wave packet is in the region of infinite motion, hence the dissociation probability decreases. Thus, the dissociation probability for a molecule should increase with decreasing field frequency.

If the photon energy is higher than $\hbar\omega \sim 0.25$ eV, which corresponds to the minimum in the $W_D(\hbar\omega)$ curve, the quasistatic condition (A6) is violated and the model under discussion no longer holds.

Application of the quasistatic model to the dissociation of the HD^+ molecule under these conditions also yields results that are in good agreement with four-surface calculations. At the same time, the agreement between these two models in the case of D_2^+ is far from satisfactory because the nuclei in this molecule move more slowly, and the quasistatic condition (A6) does not hold.

5. CONCLUSIONS

In this paper, we have discussed dissociation of various isotopes of the molecular hydrogen ion in an intense IR laser field. Exact two-particle calculations have been compared with the results of the n-surface approximation. The dissociation probability has been calculated as a function of the field intensity and frequency. The quasistatic model of dissociation has been discussed. This model provides a fairly accurate description of interaction between the laser field and a molecule in the band of lower frequencies and an interpretation of the curve of dissociation probability versus field intensity.

We thank V. P. Krainov for useful discussions of issues studied in this work. This work was supported by the Russian Fund for Fundamental Research (Grant 96-02-19286).

APPENDIX

We now derive the condition relating parameters of a molecule and laser field such that equation system (24) separates into a system of decoupled equations describing motion on quasistatic potential surfaces modified by the external field. To this end, we estimate the terms with $\partial\alpha_1^{(\varepsilon)}/\partial t$ and $\alpha_1^{(\varepsilon)}\langle\varphi_1^{(\varepsilon)}|\partial\varphi_2^{(\varepsilon)}/\partial t\rangle$ on the left of Eq. (24).

Assuming

$$\frac{\partial\alpha_1^{(\varepsilon)}}{\partial t}\sim\Omega\alpha_1^{(\varepsilon)},$$

where Ω is the molecule oscillation frequency, and taking into account

$$\frac{\partial\varphi_1^{(\varepsilon)}}{\partial t}\sim\frac{\partial\varphi_1^{(\varepsilon)}}{\partial\varepsilon}\,\varepsilon_0\omega,$$

we reformulate the adiabatic condition as

$$\hbar\Omega \gg \hbar\omega\varepsilon_0\left\langle\varphi_1^{(\varepsilon)}\left|\frac{\partial\varphi_2^{(\varepsilon)}}{\partial\varepsilon}\right.\right\rangle. \tag{A1}$$

In the two-surface approximation wave functions $\varphi_{1,2}^{(\varepsilon)}(x,R,t)$ can be expressed as

$$\varphi_{1,2}^{(\varepsilon)}=C_1^{(1,2)}(R,t)\varphi_1(x,R)+C_2^{(1,2)}(R,t)\varphi_2(x,R), \tag{A2}$$

where $\varphi_1(x,R)$ and $\varphi_2(x,R)$ are unperturbed electron wave functions, and the coefficients $C_1^{(1,2)}(R,t)$ and $C_2^{(1,2)}(R,t)$ are given by

$$\begin{aligned}
C_1^{(1)}&=[1+(\alpha/2-\sqrt{\alpha^2/4+1})^2]^{-1/2},\\
C_1^{(2)}&=[1+(\alpha/2+\sqrt{\alpha^2/4+1})^2]^{-1/2},\\
C_2^{(1)}&=(\alpha/2-\sqrt{\alpha^2/4+1})C_1^{(1)},\\
C_2^{(2)}&=(\alpha/2+\sqrt{\alpha^2/4+1})C_1^{(1)}.
\end{aligned} \tag{A3}$$

Here

$$\alpha(R,t)=\frac{E_2(R)-E_1(R)}{d_{12}(R)\varepsilon(t)}=\frac{\Delta E(R)}{d_{12}(R)\varepsilon(t)}, \tag{A4}$$

$E_1(R)$ and $E_2(R)$ are the electron energies of the two lowest unperturbed states of the molecule.

Using Eqs. (A2) and (A3), we can easily obtain the following estimate for $\langle\varphi_1^{(\varepsilon)}|\partial\varphi_2^{(\varepsilon)}/\partial t\rangle$:

$$\left\langle \varphi_1^{(\varepsilon)}\left|\frac{\partial\varphi_2^{(\varepsilon)}}{\partial t}\right.\right\rangle \sim \frac{\omega}{2}\,\frac{\alpha/2}{\alpha^2/4+1} = \frac{\omega}{2}\,\frac{\Delta E(R)/2}{\Delta E^2(R)/4+|d_{12}(R)|^2\varepsilon_0^2}\,d_{12}(R)\varepsilon_0. \tag{A5}$$

After estimating the matrix element defined by Eq. (A5) in the localization region of nuclear wave functions in a molecular bound state, we derive from Eq. (A1) the condition

$$\hbar\omega\ll\hbar\Omega\,\frac{|E_2-E_1|^2}{|d_{12}|\varepsilon}, \tag{A6}$$

which determines the limit on the external field frequency.

It is clear that condition (A6) is more rigorous at higher laser frequencies and masses of atoms in the molecule owing to the lower oscillation photon energy $\hbar\Omega$.

In the case under discussion $\hbar\Omega\approx\hbar\omega\sim 0.1$ eV. Therefore condition (A6) is valid in the localization region of the nuclear wave functions of molecule bound states at intensities up to

$$P\ll P^*\sim\frac{c}{8\pi}\,\frac{|E_2-E_1|^2}{|d_{12}|^2}. \tag{A7}$$

Taking $|E_2-E_1|\sim 1$ eV and $d_{12}\sim ea_0$, we obtain $P^*\sim 10^{13}$ W/cm^2, which is in fair agreement with our numerical calculations.

[1] N. B. Delone and V. P. Krainov, *Multiphoton Processes in Atoms*, Springer-Verlag, New York (1993).
[2] A. D. Bandrauk (ed.), *Molecules in Laser Fields*, Dekker, New York (1994).
[3] A. Giusti-Suzor, F. H. Mies, L. F. DiMauro, E. Charron, and B. Yang, J. Phys. B **28**, 309 (1995).
[4] S. Chelkowski, T. Zuo, O. Atabek, and A. D. Bandrauk, Phys. Rev. A **52**, 2977 (1995).
[5] K. C. Kulander, F. H. Mies, and K. J. Schafer, Phys. Rev. A **53**, 2562 (1996).
[6] E. A. Volkova, A. M. Popov, and O. V. Tikhonova, Zh. Éksp. Teor. Fiz. **110**, 1616 (1996) [JETP **83**, 889 (1996)].
[7] Ju. A. Il'inskii and L. V. Keldysh, *Interaction of Electromagnetic Radiation with Matter* [in Russian], Moscow State University Press, Moscow (1989).
[8] A. M. Popov, O. V. Tikhonova, and E. A. Volkova, Laser Phys. **7**, 843 (1997).
[9] E. Charron, A. Gusti-Suzor, and F. H. Mies, J. Chem. Phys. **103**, 7359 (1995).
[10] M. V. Fedorov, O. V. Kudrevatova, V. P. Makarov, and A. A. Samokhin, Opt. Commun. **13**, 293 (1975).
[11] A. Zavriyev, P. H. Bucksbaum, H. G. Muller, and D. W. Schumacher, Phys. Rev. A **42**, 5500 (1990).
[12] A. Zavriyev, P. H. Bucksbaum, J. Squier, and F. Saline, Phys. Rev. Lett. **70**, 1077 (1993).
[13] H. Rottke, J. Ludwig, and W. Sandner, Phys. Rev. A **54**, 2224 (1996).
[14] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics. Nonrelativistic Theory*, Pergamon Press, New York (1976).
[15] P. Dietrich, M. Yu. Ivanov, F. A. Ilkov, and P. B. Corcum, Phys. Rev. Lett. **77**, 4150 (1996).
[16] V. P. Krainov, M. E. Sukharev, Laser Phys. **7**, 803 (1997).

Translation provided by the Russian Editorial office

Effect of four-photon interactions on coherent population trapping in Λ-systems

B. A. Grishanin, V. N. Zadkov

International Laser Center, M. V. Lomonosov Moscow State University, 119899 Moscow, Russia

D. Meschede

Institut für Angewandte Physik der Universität Bonn, Wegelerstr. 8, D-53115 Bonn, Germany
(Submitted 26 June 1997)
Zh. Éksp. Teor. Fiz. **113**, 144–167 (January 1998)

The resonance fluorescence spectrum of a Λ-system excited by two resonant light fields is calculated using a Markov analysis. Analytical formulas are derived in the strong-field limit within and beyond the rotating wave approximation. It is shown that the resonance fluorescence of the system does not vanish during coherent population trapping. Its spectrum consists of two multiplets which are similar to a triplet in the resonance fluorescence spectrum of a two-level atom and lie at the electronic transition frequencies, together with two triplets located at the frequencies of four-photon processes involving the optical excitation fields. The latter are fundamental in character and impose limits on the lower bound of the dephasing rate for the Raman resonance owing to the effect of radiative decay of the dipole transitions on the dynamics of the ground state. The effect of four-photon dephasing on the absorption spectrum of a Λ-system is analyzed and found to lead to a substantial reduction in the depth of a dip in the absorption spectrum which vanishes as the laser field strength is increased. [S1063-7761(98)01001-4]

1. INTRODUCTION

The interaction of electromagnetic fields with atoms is one of the most fundamental problems in quantum optics. It is known that a much wider range of effects occur in multilevel atoms than in two-level atoms owing to field-induced coherence between the atomic states and quantum interference. The three-level systems realized in Λ-, Ξ-, and V-configurations play an important role in research on these effects, as they are of intermediate complexity between two-level and multilevel atoms. A whole series of new effects have been observed in them; coherent population trapping is one of the most intriguing and has been studied intensely, both experimentally and theoretically. (See the reviews by Agap'ev *et al.*[1] and Arimondo[2] and the references cited there.) Coherent populating trapping shows up most clearly in a three-level system with two close long-lived levels and a third level which lies far from them (Λ- or V-systems) that have been excited by two cw laser fields, so that the distant level is optically coupled to the two others. Tuning the driver fields to resonance with its dipole transitions leads to trapping of the populations of the system in a coherent superposition of the two close levels. In Raman absorption spectra this effect shows up as a very narrow dip against the background of an absorption line and in resonance fluorescence spectra it is observed as the absence of emission, which has led to its being referred to as a ''dark'' (or ''coherent population trapping'') resonance.

In this article we study the question of how four-photon interactions affect the coherent population trapping effect in a Λ-system excited by two cw laser fields, in particular, how ''dark'' the coherent population trapping resonance in the resonance fluorescence spectrum is and how deep the dip in the Raman absorption spectrum is. A rough estimate of the intensity of resonance fluorescence in a Λ-system during coherent population trapping has been made in the rotating wave approximation[3] which yields zero fluorescence intensity for two-level atoms. The same result can be seen in Fig. 5c of Narducci *et al.*,[4] which shows a calculated fluorescence spectrum for a Λ-system. This indicates that during coherent population trapping a Λ-system does not radiate and the dark line is entirely absent in its resonance fluorescence spectrum.

Our calculations, presented in this paper in the asymptotic limit of a strong field, show, however, that the resonance fluorescence of a system does not vanish during coherent population trapping. Its spectrum consists of two multiplets, similar to the triplet in the resonance spectrum of two-level atoms and located at the electronic transition frequencies, together with two triplets located at the frequencies of four-photon processes involving the pump light fields. The latter are fundamental in character and impose limits on the lower bound of the dephasing rate of the Raman resonance owing to the contribution to the dynamics of the ground state from radiative decay of dipole transitions. The effect of the four-photon dephasing mechanism on the absorption spectrum of a Λ-system is analyzed and found to lead to a substantial reduction in the depth of a dip in the absorption spectrum that vanishes as the laser field intensity is increased.

This article is organized as follows: Section 2 is devoted to a description of the complete Liouvillian of an atom interacting with a laser radiation field. The specific features of the excitation of two-level atoms and a Λ-system are analyzed. In Sec. 3 the resonance fluorescence spectrum of a Λ-system

is calculated in the rotating wave approximation, as well as outside the range of validity of this approximation. The effect of four-photon interactions involving the driver light fields on coherent population trapping and their role in the formation of the absorption resonance and in the dispersion of the Λ-system are analyzed in Secs. 4 and 5, respectively. Most of the voluminous mathematical calculations are carried out in the Appendix. In the Conclusion we discuss an experiment for detecting the calculated structure of the resonance fluorescence spectrum of a Λ-system.

2. LIOUVILLIAN OF AN ATOM BEYOND THE RANGE OF VALIDITY OF THE ROTATING WAVE APPROXIMATION; DYNAMICAL TRANSFORMATIONS USED TO CALCULATE THE FLUORESCENCE SPECTRUM

The complete Liouvillian of an atom, which describes changes in the atomic variables according to the equation $d\hat{A}/dt=\mathscr{L}(t)\hat{A}$ in Markov theory, has the form

$$\mathscr{L}(t)=\mathscr{L}_0+\mathscr{L}_\delta+\mathscr{L}_r+\mathscr{L}_L(t). \tag{1}$$

Here $\mathscr{L}_0$ is the unperturbed Liouvillian $(i/\hbar)[\hat{\mathscr{H}}_0,\odot]$, including the free precession of the atom at the laser frequencies according to Eq. (A4) of Appendix A. (The symbol $\odot$ denotes a place for substituting a transformed operator.) $\mathscr{L}_r$ and $\mathscr{L}_\delta$ determine the dynamics of the atom owing, respectively, to relaxation and nonzero detuning of the frequencies of the driver laser fields from the resonance transition frequencies in the atom (resonance excitation is described by $\mathscr{L}_0$), while $\mathscr{L}_L(t)$ describes the laser excitation.

In terms of the interaction representation the transformation $S(0,t)$ corresponding to the Liouvillian (1) takes the form

$$S(0,t)=S_{RWA}(0,t)\tilde{S}_0(0,t)e^{\mathscr{L}_0 t}, \tag{2}$$

where the superoperators

$$S_{RWA}(0,t)=\exp(\mathscr{L}_{RWA}t), \quad \mathscr{L}_{RWA}=\mathscr{L}_\delta+\mathscr{L}_r+\mathscr{L}_p \tag{3}$$

determine the system dynamics in the rotating wave approximation (RWA),

$$\tilde{S}_0(0,t)=\mathrm{T}\exp\left[\int_0^t \delta\mathscr{L}_p(\tau)d\tau\right] \tag{4}$$

is the evolution superoperator for the dynamics of the system owing to the nonresonant excitation component, and

$$\delta\mathscr{L}_p(t)=e^{\mathscr{L}_0 t}\mathscr{L}_L(t)e^{-\mathscr{L}_0 t}-\mathscr{L}_p \tag{5}$$

is the deviation from the average value $\mathscr{L}_p$ of the Liouvillian for the laser excitation.[5] The symbol T used in Eq. (4) denotes the time ordering of the superoperator taken in quantum mechanics.

The deviation $\delta\mathscr{L}_p(\tau)$ in Eq. (5) oscillates at the frequencies of the laser drive fields and their combinations. Thus, we can treat it as a universally small perturbation relative to the resonant excitation contribution for the amplitude of driver waves smaller than the amplitude of the intraatomic field. In a first-order approximation with respect to this parameter, the evolution superoperator has the form

$$S(0,t)=S(0,t)_{RWA}\left[1+\int_0^t \delta\mathscr{L}_p(\tau)d\tau\right]e^{\mathscr{L}_0 t}. \tag{6}$$

Note that this approximation is valid if the value of the integral is less than of order unity.

We now consider the specifics features of the excitation of a two-level atom and a Λ-system.

2.1. Excitation of a two-level atom

For a two-level atom excited by a laser field $E_L\cos(\omega_L t)$, Eq. (5) takes the form

$$\delta\mathscr{L}_p(\tau)=i\frac{\Omega_0}{2}[\hat{\sigma}^+\exp(-2i\omega_L t) + \hat{\sigma}^-\exp(2i\omega_L t),\odot], \tag{7}$$

where Ω_0 is the Rabi frequency and $\hat{\sigma}^\pm$ are the standard Pauli matrices. Applying Eq. (7) to the complex polarization amplitude $\hat{\sigma}^+$ and using Eq. (6), we find that $S(t)\hat{\sigma}^+$ determines the structure of a triplet (because of the presence of the term $S(t)_{RWA}$) in the resonance fluorescence spectrum of a two-level atom at a frequency of $3\omega_L$,[5] which is analogous to the known triplet at the laser excitation frequency.[6,7] Integrating with respect to τ in Eq. (6), we can easily show that the ratio of the corresponding amplitudes of the spectral components at the frequencies of the third harmonic and the laser light is proportional to the small quantity $\Omega_0/2\omega_L$.

2.2. Excitation of a Λ-system

Let us consider a Λ-system consisting of three electronic levels with transition frequencies among them of $\omega_{12}\ll\omega_{13}$, ω_{23} (Fig. 1). Two coherent fields $E\cos(\omega_L t)$ and $E'\cos(\omega_L' t)$ act, respectively, on the transitions $1\leftrightarrow 3$ and $2\leftrightarrow 3$. These fields interact with the complete dipole moment of the system determined by the operator $d_{13}\hat{\sigma}_{13}+d_{23}\hat{\sigma}_{23}$, where $\hat{\sigma}_{13,23}$ are the Pauli matrices $\hat{\sigma}_1$ for the corresponding atomic transitions. As a result, the induced dipole moment of the system oscillates at frequencies $\pm\omega_L$ and $\pm\omega_L'$.

As opposed to the case of two-level atoms, where both driver fields interact with one and the same atomic transition, during excitation of a Λ-system each field interacts with two transitions. Thus, the Liouvillian corresponding to biharmonic laser excitation with a frequency detuning $\Delta=\omega_L'-\omega_L$ takes (according to Eq. (B3)) the form

$$\delta\mathscr{L}_p(\tau)=\frac{i\hbar g_\Lambda}{2}[(\hat{\tau}e^{i\Delta t}+\hat{\tau}^+e^{-i\Delta t}),\odot], \tag{8}$$

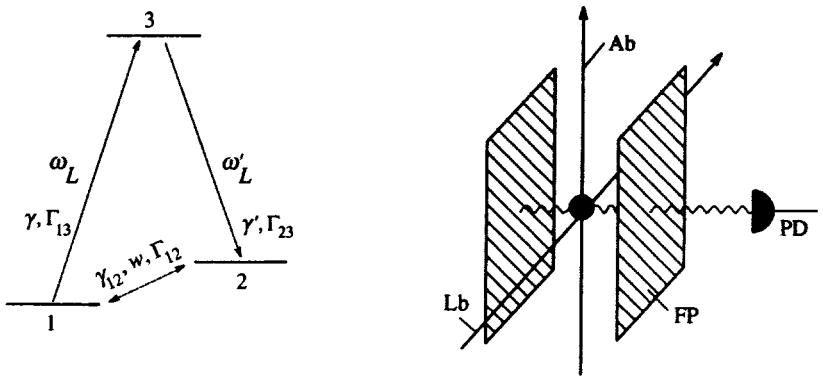


FIG. 1. A Λ-system (a) and a typical arrangement for the experimental measurement of resonance fluorescence induced by two monochromatic laser fields with frequencies ω_L and ω'_L (b). γ, γ', and γ_{12} are the relaxation rates of the populations in the upper levels; Γ_{13}, Γ_{23}, and Γ_{12} are the dephasing rates; and, w is the rate of pumping to level 2. The fluorescence spectrum of the atoms is analyzed using a Fabry-Perot interferometer (FP) and a photodiode (PD) in a direction perpendicular to the directions of the laser (Lb) and atom (Ab) beams.

where $g_\Lambda = \sqrt{g^2+g'^2}$, $g = d_{13}E'$, $g' = d_{23}E$, and the operator $\hat{\tau}$ is defined as

$$\hat{\tau} = g_\Lambda^{-1}(g\hat{\sigma}_{13}^{+} + g'\hat{\sigma}_{23}^{-}). \tag{9}$$

In deriving Eq. (8) we have neglected terms containing sums with higher frequencies.

Equation (8) obviously determines additional spectral components at frequencies $\omega_L \pm \Delta$ and $\omega'_L \pm \Delta$, of which only the components at $\omega_L - \Delta$ and $\omega'_L + \Delta$ are new. The correspond to four-photon processes and should show up for a symmetric Λ-system as a mirror reflection of the virtual levels of the subsystem of lower levels (Fig. 2) owing to modulation of the $1 \leftrightarrow 3$ and $2 \leftrightarrow 3$ transitions by the intrinsic oscillatory frequency $\Delta \approx \omega_{12}$ of the lower level subsystem. It is known that four-photon frequency mixing leads to generation of a coherent signal at the Stokes and anti-Stokes frequencies.[8] Our later calculations show, however, that these nonlinear resonances are also accompanied by sidebands because of incoherent scattering processes.

The above analysis shows that the important difference between exciting a two-level atom (see Eq. (7)) and a Λ-system (Eq. (8)) is that excitation in the case of the Λ-system is mainly determined by the biharmonic frequency detuning Δ. In experiments this detuning is usually much smaller than the frequencies of the laser systems that are exciting the system. This means that the intensity of the additional components in the fluorescence spectrum (fine structure), which is determined by the exponential factors in Eqs. (7) and (8), is substantially higher for a Λ-system than for a two-level atom.

3. CALCULATING THE FLUORESCENCE SPECTRUM OF A Λ-SYSTEM

The spectral density of the emission from an excited atom (resonance fluorescence spectrum) is determined by the normally ordered two-time correlation function of the light emitted by the atom.[9,10] Assuming that the atomic fluctuations are Markovian, i.e., they are independent of one another at times t and $t+\tau$, we can write down the correlation function for the atomic fluorescence in the form

$$\mathcal{K}(\tau) = \langle \hat{\rho}_0 S(0,t) | \hat{\sigma}^{-}(t)[S(t,t+\tau)\hat{\sigma}^{+}(t+\tau)] \rangle, \tag{10}$$

where $\hat{\sigma}^{\pm}(t)$ are the Heisenberg positive (negative) frequency operators. These operators have a time dependence only in the form of high-frequency oscillations at optical frequencies. The superoperators $S(0,t)$ and $S(t,t+\tau)$ describe the relaxation and interaction of the atoms with the exciting laser fields during the time intervals $(0,t)$ and $(t,t+\tau)$, and $\hat{\rho}_0 S(0,t)$ is the density matrix $\hat{\rho}(t)$ at time t. It follows from Eq. (3) that the superoperators $S(0,t)$ and $S(t,t+\tau)$ are simple exponentials of the form $\exp[\mathcal{L}_{RWA}t]$, according to the rotating wave approximation.

The term $\hat{\sigma}^{-}(t)[S(t,t+\tau)\hat{\sigma}^{+}(t+\tau)]$ in Eq. (10) is simply the product of the two operators $\hat{\sigma}^{-}(t)$ and $\hat{\sigma}^{+}(t+\tau)$ averaged over the fluctuations in the time interval $(t,t+\tau)$. This averaging is carried out with the aid of the transformation $S(t,t+\tau)$, which determines the conditional atomic quantum mechanical probability distribution function at time $t+\tau$ relative to time t. The emission spectrum of the atom can then be calculated as the Fourier transform of the correlation function (10).

In the stationary case the density matrix in the vector representation is simply the zero vector $\langle 0|$ of the matrix of the evolution superoperator $\mathcal{L}_{RWA}$. Then we can obtain the stationary correlation function from Eq. (10) by averaging it over the temporal oscillations. This averaging leads to the replacement of the bilinear combination of the complete operators $\hat{\sigma}^{\pm}(t)$ by two combinations:

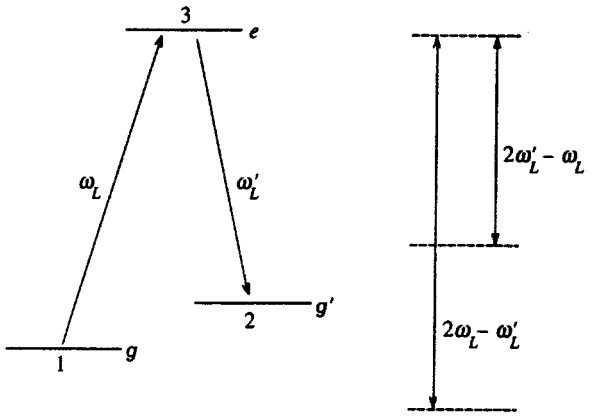


FIG. 2. A Λ-resonance and additional resonances which determine the fine structure of the resonance fluorescence spectrum.

$$\hat{\sigma}^-(t)\times\hat{\sigma}^+(t+\tau)\to\hat{\sigma}_{13}^-\times\hat{\sigma}_{13}^+(\tau)\oplus\hat{\sigma}_{23}^-\times\hat{\sigma}_{23}^+(\tau).$$

Similarly, we have

$$\mathscr{K}(\tau)\to\mathscr{K}_1(\tau)+\mathscr{K}_2(\tau).$$

When we calculate the resonance fluorescence spectrum of the atom in the rotating wave approximation the correlation functions $\mathscr{K}_{1,2}(\tau)$ correspond to the frequencies $\omega_{13}\approx\omega_L$ and $\omega_{23}\approx\omega_L'$ and when we calculate the fine structure of the spectrum beyond the range of validity of this approximation they correspond, respectively, to the frequencies $\omega_L-\Delta=2\omega_L-\omega_L'$ and $\omega_L'+\Delta=2\omega_L'-\omega_L$. (See Sec. 2.2.)

On describing the Λ-system with the aid of the Liouvillian in the rotating wave approximation and expanding it in terms of the eigen-projectors, we obtain the following relatively simple expression:

$$\mathscr{K}(\tau)=\sum_{k=0}^{8}\{\langle 0|\hat{\sigma}_{13}^-\cdot|k\rangle\langle k|\hat{\sigma}_{13}^+\rangle\exp[(\lambda_k-i\omega_L)\tau]+\langle 0|\hat{\sigma}_{23}^-\cdot|k\rangle\langle k|\hat{\sigma}_{23}^+\rangle\exp[(\lambda_k-i\omega_L')\tau]\},\quad(11)$$

where the symbol ''$\cdot$'' means that the operators are multiplied in accordance with the multiplication rules for operators and the result is presented in the form of a ket-vector; λ_k, $|k\rangle$, and $\langle k|$ are the eigenvalues of the matrix and the eigenvectors of the Liouvillian $\mathscr{L}_{RWA}$.

Using Eq. (11) together with Appendix B we can obtain the following expression for the correlation function that describes the structure of the resonance fluorescence spectrum of the atom outside the range of validity of the rotating wave approximation:

$$\mathscr{K}(\tau)=\frac{g_\Lambda^2}{4\omega_{12}^2}\sum_{k=0}^{8}\langle 0|\hat{\sigma}_{12}^-\cdot|k\rangle\langle k|\hat{\sigma}_{12}^+\rangle\{\exp[-i(\omega_L-\Delta)\tau]+\exp[-i(\omega_L'+\Delta)\tau]\}\exp(\lambda_k\tau),\quad(12)$$

where the $\hat{\sigma}_{12}^{\pm}$ are the complex conjugate amplitude of the subsystem of lower levels which modulate the dipole moment of the transition. This modulation gives rise to new spectral components in the fluorescence spectrum.

Recall that Eq. (12) describes only the basic structure of the fluorescence spectrum, which is determined by the parameter g_Λ/ω_{12}, which, in turn, we assume to be small. Here we have neglected the higher order contribution which makes a nonzero contribution to the coherent component of the response in the rotating wave approximation, which component equals zero when this correction is neglected in a strong field in a first-order approximation with respect to the parameter Γ/g_Λ for detunings δ, $\delta_R\approx 0$ (by analogy with the two-level atom[7]).

3.1. Fluorescence spectrum in the rotating field approximation

The fluorescence spectrum determined by Eq. (11) is the sum of Lorentz spectrum lines whose total spectral powers are proportional to the coefficient in front of the corresponding exponent in Eq. (11), while their width and frequency shift are determined by the real and imaginary parts of the eigenvalues λ_k. In general, the fluorescence spectrum can be calculated numerically. In an asymptotically strong field, however, as we shall show below (and has been demonstrated previously[4] for a special case) an analytic solution can also be obtained.

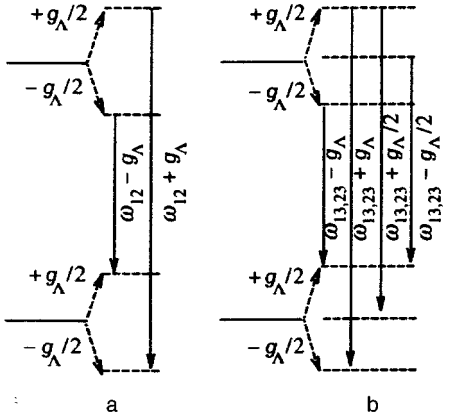


FIG. 3. The formation of Rabi nutations in a two-level atom (a) and in a Λ-system (b). Only a minimum set of transitions between the quasienergy states, corresponding to the set of all possible lines in the fluorescence spectrum, is shown,

For simplicity let us consider a Λ-system excited by two high-power laser fields of equal intensity. In this case we can average the relaxation of the system over the Rabi nutations, while the Hamiltonian corresponding to the laser-induced precession operator $\mathscr{L}_p=(i/\hbar)[\hat{\mathscr{H}}_p,\odot]$ takes the form

$$\hat{\mathscr{H}}_p=\hbar\frac{g}{2}\hat{\sigma}=\hbar\frac{g_\Lambda}{2\sqrt{2}}\begin{pmatrix}0&1&1\\1&0&0\\1&0&0\end{pmatrix}.$$

This Hamiltonian corresponds to quasi-energy states with energies that are shifted with respect to the eigenvalues of the Hamiltonian $\hat{\mathscr{H}}_p$, which equal $\{0,\pm g_\Lambda/2\}$.[11] (For a two-level atom the eigenvalues of the Hamiltonian are equal to $\{\pm g_\Lambda/2\}$). The temporal dynamics of these mixed quasi-energy states cause oscillations in the expected values of the physical variables at two different frequencies g_Λ and $g_\Lambda/2$. The physical significance of these nutations in terms of the quasienergy levels is illustrated in Fig. 3.

The Rabi nutations between the quasi-energy levels are described by a Liouvillian which, in the operator basis $\{\hat{n}_3,\hat{n}_1,\hat{n}_2,\hat{\sigma}_{12}^c,\hat{\sigma}_{12}^s,\hat{\sigma}_{13}^c,\hat{\sigma}_{13}^s,\hat{\sigma}_{23}^c,\hat{\sigma}_{23}^s\}$ (the indices c and s denote the cosine and sine components, respectively), takes the form

$$\mathscr{L}_p=\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & \frac{g_\Lambda}{2} & 0 & \frac{g_\Lambda}{2} \\ 0 & 0 & 0 & 0 & 0 & 0 & -\frac{g_\Lambda}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\frac{g_\Lambda}{2} \\ 0 & 0 & 0 & 0 & 0 & 0 & -\frac{g_\Lambda}{2\sqrt{2}} & 0 & -\frac{g_\Lambda}{2\sqrt{2}} \\ 0 & 0 & 0 & 0 & 0 & \frac{g_\Lambda}{2\sqrt{2}} & 0 & -\frac{g_\Lambda}{2\sqrt{2}} & 0 \\ 0 & 0 & 0 & 0 & -\frac{g_\Lambda}{2\sqrt{2}} & 0 & 0 & 0 & 0 \\ -\frac{g_\Lambda}{2} & \frac{g_\Lambda}{2} & 0 & \frac{g_\Lambda}{2\sqrt{2}} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{g_\Lambda}{2\sqrt{2}} & 0 & 0 & 0 & 0 \\ -\frac{g_\Lambda}{2} & 0 & \frac{g_\Lambda}{2} & \frac{g_\Lambda}{2\sqrt{2}} & 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{13}$$

Its eigenvalues λ_k ($k=0, \ldots, 8$) are equal to 0, 0, 0, $-ig_\Lambda/2$, $-ig_\Lambda/2$, $ig_\Lambda/2$, $ig_\Lambda/2$, $-ig_\Lambda$, ig_Λ, while the corresponding set of eigenvectors is defined as

$$\{\psi_k\}=\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & \sqrt{1/2} & 0 & \sqrt{1/2} & 0 \\ 0 & -1/2 & -1/2 & \sqrt{1/2} & 0 & 0 & 0 & 0 & 0 \\ \sqrt{1/2} & 2^{-3/2} & 2^{-3/2} & 1/2 & 0 & 0 & 0 & 0 & 0 \\ 0 & i/2 & -i/2 & 0 & 0 & 0 & -1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & \sqrt{1/2} & -i/2 & 0 & i/2 & 0 \\ 0 & 0 & 0 & 0 & \sqrt{1/2} & i/2 & 0 & -i/2 & 0 \\ 0 & -i/2 & i/2 & 0 & 0 & 0 & -1/2 & 0 & 1/2 \\ 1/2 & -1/4 & -1/4 & -2^{-3/2} & 0 & 0 & -i/2 & 0 & -i/2 \\ 1/2 & -1/4 & -1/4 & -2^{-3/2} & 0 & 0 & i/2 & 0 & i/2 \end{pmatrix}.$$

Let us now discuss the physical significance of the dynamical variables corresponding to the eigenvectors ψ_k.

The eigenvector ψ_0 describes the stationary excitation of a system by two laser fields of equal intensity acting on the $1\leftrightarrow 3$ and $2\leftrightarrow 3$ transitions, respectively. The eigenvectors ψ_1 and ψ_2 describe a two-dimensional stationary excitation space, a combination of the polarization of the ground state and the populations of all three levels. The eigenvectors ψ_3 and ψ_6 describe excitation which involves a combination of the populations of the subsystems of the lower levels and an independent combination of the polarizations $1\leftrightarrow 3 \oplus 2\leftrightarrow 3$; they oscillate at half the Rabi frequency, $g_\Lambda/2$. The eigenvectors ψ_4 and ψ_5 describe excitations which are a combination of three polarizations ($1\leftrightarrow 2$ and an independent combination $1\leftrightarrow 3 \oplus 2\leftrightarrow 3$) which also oscillate at a frequency of $g_\Lambda/2$. The last two eigenvectors, ψ_7 and ψ_8, describe the excitation of the bound state together with the populations of all three levels and the polarization of the ground state, (all) oscillating at a frequency of g_Λ. Therefore, the Rabi nutations of the eigenexcitations of the system for combinations of the unbound levels take place at a frequency of $g_\Lambda/2$, while the bound states oscillate at a frequency of g_Λ. (See Eq. (A5)).

Using Eq. (13) for the nutation operator $\mathscr{L}_P$, we can average the Liouvillian $\mathscr{L}_\delta + \mathscr{L}_r$ in Eq. (1) over the nutations and write it in the form of a sum of 3×3 matrices, two 2×2, and two 1×1. (The last two are diagonal elements.) As a result, we can obtain a simplified expression for the last three terms in Eq. (1), which describe the overall dynamics of the system in the interaction representation, of the form

$$\mathscr{L}_{RWA}=\begin{pmatrix} -(\Gamma_{13}+\Gamma_{23})/2 & 0 & 0 \\ 0 & -\Gamma_{12}/2 & -(\gamma+\gamma'+\Gamma_{12})/2\sqrt{2} \\ 0 & -\Gamma_{12}/2\sqrt{2} & -(\gamma+\gamma'+\Gamma_{12})/4 \end{pmatrix}$$
$$\oplus\begin{pmatrix} -(\gamma_{12}+w)/2-\Gamma_{13}/4-\Gamma_{23}/4-ig_\Lambda/2 & -i\delta_s/2 \\ -i\delta_s/2 & -\Gamma_{12}/2-\Gamma_{13}/4-\Gamma_{23}/4-ig_\Lambda/2 \end{pmatrix}$$
$$\oplus\begin{pmatrix} -\Gamma_{12}/2-\Gamma_{13}/4-\Gamma_{23}/4+ig_\Lambda/2 & i\delta_s/2 \\ i\delta_s/2 & -(\gamma_{12}+w)/2-\Gamma_{13}/4-\Gamma_{23}/4+ig_\Lambda/2 \end{pmatrix}$$
$$\oplus(-3\gamma/8-3\gamma'/8-\Gamma_{12}/8-\Gamma_{13}/4-\Gamma_{23}/4-ig_\Lambda)$$
$$\oplus(-3\gamma/8-3\gamma'/8-\Gamma_{12}/8-\Gamma_{13}/4-\Gamma_{23}/4+ig_\Lambda),$$

where the total detuning is $\delta_s=2\delta+\delta_R=\omega_L+\omega_L'-\omega_{13}-\omega_{23}$. The corresponding eigenvalues are given by

$$\{\lambda_\kappa\}=\begin{Bmatrix} 0 \\ -(\gamma+\gamma'+3\Gamma_{12})/4 \\ -(\Gamma_{13}+\Gamma_{23})/2 \\ \mu_1 \\ \mu_2 \\ \mu_1^* \\ \mu_2^* \\ (-3\gamma-3\gamma'-\Gamma_{12}-2\Gamma_{13}-2\Gamma_{23}-8ig_\Lambda)/8 \\ (-3\gamma-3\gamma'-\Gamma_{12}-2\Gamma_{13}-2\Gamma_{23}+8ig_\Lambda)/8 \end{Bmatrix}, \tag{14}$$

where

$$\mu_{1,2}=\frac{1}{4}\,[-\gamma_{12}-w-\Gamma_{12}\pm i\sqrt{4\delta_s^2-(\gamma_{12}+w+\Gamma_{12})^2}$$
$$-\Gamma_{13}-\Gamma_{23}-2ig_\Lambda].$$

Here γ_{12} and w are the rates of relaxation and pumping of the lower level system, Γ_{12} is the dephasing rate in this system, γ and γ' are the rates of relaxation from the excited states, and Γ_{13} and Γ_{23} are the corresponding dephasing rates.

Let us now discuss the eigenvalues λ_k in detail.

Note that because the relaxation operator is not self-adjoint, each eigenvalue corresponds to two eigenvectors, one of which describes the operators acting on the physical variables, while the other describes the density matrix. The eigenvalue $\lambda_0=0$ corresponds to the stationary state $\hat{\rho}_{st}\to\langle 0|$ and the operator $\hat{I}\to|0\rangle$, which has no dynamical significance. This eigenvalue determines the coherent line in the fluorescence spectrum. The eigenvalues $\lambda_{1,2}$ describe the nonoscillatory dynamics of the system and determine the Rayleigh scattering of the fields which excite the system. The four eigenvalues $\lambda_{3,4,5,6}$ determine oscillations at half the Rabi frequency $g_\Lambda/2=g/\sqrt{2}$ and describe the contribution of field-induced resonances to the fluorescence spectrum. The last two eigenvalues $\lambda_{7,8}$ determine oscillations at the Rabi frequency g_Λ and describe the ordinary $1\leftrightarrow 3$ and $2\leftrightarrow 3$ nutations in the weak-field limit.

The preceding analysis shows that in a strong field only the total detuning $\delta_s=2\delta+\delta_R$ appears among the eigenfrequencies of the system, and not the Raman detuning δ_R. Here for $\delta=0$, relaxation in the system of lower levels does not contribute to the oscillations at half the Rabi frequency:

$$\mu_{1,2}=-\Gamma_{13}/4-\Gamma_{23}/4-ig_\Lambda/2.$$

This effect can, in principle, be used to study the contribution of the lower level system to the fluorescence spectrum in an experiment where the spectra are measured as a function of the detuning δ for different intensities of the laser line. The measured width of the spectral components located at half the Rabi frequency is then determined directly by the relaxation rate in the lower level system.

For the case of an exact resonance ($\delta,\delta_R=0$), we can obtain an analytic expression for the fluorescence spectrum in the rotating wave approximation. The major difference compared to the spectra from two independent two-level systems, however, is that in the case of the Λ-system the general coefficient in Eq. (11), which determines the intensity of the spectral components, differs from the corresponding coefficient for the case of a two-level atom, which is simply proportional to γ. In coherent population trapping, this coefficient for a Λ-system and, therefore, the intensity of the spectral components decrease by roughly a factor of Γ_{12}/γ, which is a small parameter. For the cesium atom,[12] as an example, it is $\approx 1.6\times 10^{-3}$, while for sodium[2] it can be estimated to be $\approx 4.9\times 10^{-3}$ using published parameters.[13]

3.2. Fine structure of the fluorescence spectrum

For simplicity let us again consider the case of an exact resonance. Using the equations from Sec. 3.1 together with Eq. (12), we obtain the following expressions for the coefficients c_k in front of the exponential factors:

$$c_0=\frac{\gamma^2/\Gamma_{12}^2}{\sqrt{(3+2\gamma/\Gamma_{12})^3}},\quad c_1=\frac{9}{8}\,\frac{1+2\gamma/\Gamma_{12}}{\sqrt{(3+2\gamma/\Gamma_{12})^3}},$$

$$c_{2,3,5}=0,\quad c_{4,6}=\frac{1}{4}\,\frac{1+\gamma/\Gamma_{12}}{\sqrt{3+2\gamma/\Gamma_{12}}},$$

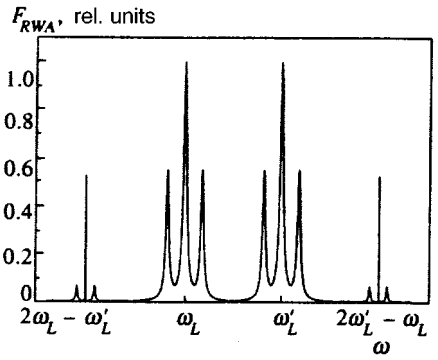


FIG. 4. The resonance fluorescence spectrum of a Λ-system (the $^2S_{1/2} \rightarrow {}^2P_{3/2}$ transition in the cesium atom) excited by two intense laser fields into a coherent population trapping state.

$$c_{7,8} = \frac{1}{16} \frac{1}{\sqrt{3 + 2\gamma/\Gamma_{12}}}, \tag{15}$$

These coefficients multiplied by the common factor $\gamma g_\Lambda^2 / 4\omega_{12}^2$ determine the intensity of the fine structure components of the fluorescence spectrum (12).

The complete fluorescence spectrum of a Λ-system, including the structure in the rotating wave approximation as well as the fine structure calculated above, is shown in Fig. 4. Equation (15) implies that for the typically large values of the ratio γ/Γ_{12} only the coefficients c_0 and $c_4 = c_6$ are proportional to the large values of order $\sqrt{\gamma/\Gamma_{12}}$. As a result, only three lines show up in each of the two fine structure features of the fluorescence spectrum (Fig. 4). One of them is coherent (i.e., has zero width) with an intensity proportional to c_0, while the other two are broadened lines with an intensity proportional to c_4 shifted to the left and right of the coherent line at half the Rabi frequency $g_\Lambda/2$.

4. EFFECT OF FOUR-PHOTON INTERACTIONS ON COHERENT POPULATION TRAPPING

Using Appendix B we can easily calculate the contribution to the relaxation of the ground state from the relaxation contributions of the dipole transitions. The nature of the relaxation processes involves an interaction of the dipole moments of the $1 \leftrightarrow 3$ and $2 \leftrightarrow 3$ transitions with the vacuum fluctuations of the electromagnetic field. These interactions are described by the Hamiltonian

$$\hat{\mathscr{H}}_\xi = \frac{\hbar}{2} \left(\hat{\xi}_1^- \hat{\sigma}_{13}^+ + \hat{\xi}_2^- \hat{\sigma}_{23}^+ \right) + \text{H.c.}, \tag{16}$$

where

$$\hat{\xi}_1^- = \frac{1}{\hbar} \int \mathbf{d}_{13}(\mathbf{r}) \hat{\mathbf{E}}_0(\mathbf{r}), \quad \hat{\xi}_2^- = \frac{1}{\hbar} \int \mathbf{d}_{23}(\mathbf{r}) \hat{\mathbf{E}}_0(\mathbf{r})$$

are the components of the vacuum electromagnetic field $\hat{\mathbf{E}}_0(\mathbf{r})$ with negative frequency amplitudes integrated over the spatial distribution $\mathbf{d}_{k3}(\mathbf{r})$ of the dipole moments $(k = 1,2)$.

Isolating the contribution to $\hat{\sigma}_{13}^+$ and $\hat{\sigma}_{23}^+$ owing to four-photon interactions outside the range of applicability of the rotating wave approximation, we can write Eq. (16) in the form

$$\hat{\mathscr{H}}_\xi = \hat{\mathscr{H}}_\xi^0 + \hat{\mathscr{H}}_\xi^{Aph},$$

$$\hat{\mathscr{H}}_\xi^{Aph} = \frac{\hbar}{2} \left(\hat{\xi}_1^- \, \delta S_0 \hat{\sigma}_{13}^+ + \hat{\xi}_2^- \, \delta S_0 \hat{\sigma}_{23}^+ \right) + \text{H.c.},$$

where $\hat{\mathscr{H}}_\xi^0$ determines the standard interaction, $\hat{\mathscr{H}}_\xi^{Aph}$ describes the additional contribution owing to four-photon interactions, and δS_0 is the four-photon contribution to the dynamic transformation of the Λ-system. Using the final formulas of Appendix B for the transformed operators $\hat{\sigma}_{13}^+$ and $\hat{\sigma}_{23}^+$, we obtain the following formula for the four-photon contribution to the Hamiltonian:

$$\hat{\mathscr{H}}_\xi^{Aph} = \frac{g_\Lambda}{2\Delta} \frac{\hbar}{2} \{ \hat{\xi}_1^- [\hat{\tau}, \hat{\sigma}_{13}^+] \exp[-i(\omega_L - \Delta)t] + \hat{\xi}_2^- [\hat{\tau}^+, \hat{\sigma}_{23}^+] \exp[-i(\omega'_L + \Delta)t] \} + \text{H.c.} \tag{17}$$

By calculating the commutators in Eq. (17), we can write down the four-photon contribution with the aid of the transition operators for the low level subsystem as

$$\hat{\mathscr{H}}_\xi^{Aph} = \frac{\hbar}{4\Delta} \{ g' \hat{\xi}_1^-(t) \exp[-i(\omega_L - \Delta)t] \hat{\sigma}_{12}^- + g \hat{\xi}_2^-(t) \exp[-i(\omega'_L + \Delta)t] \hat{\sigma}_{12}^+ \} + \text{H.c.} \tag{18}$$

This implies that the vacuum electromagnetic field interacts with the lower level subsystem through four-photon processes. The efficiency of this interaction depends on the interaction constant of the laser fields with the dipole transitions of the Λ-system. The distinctive feature of this interaction is that the emission of a vacuum photon by the $1 \leftrightarrow 3$ transition is accompanied by the absorption of a $1 \leftrightarrow 2$ transition photon, while the emission of a vacuum photon by the $2 \leftrightarrow 3$ transition is accompanied by the emission of a $1 \leftrightarrow 2$ transition photon. Conservation of energy in these processes is ensured by the four-photon interaction of the vacuum field with the laser fields, and this is reflected in the exponential terms in Eq. (18).

Following Ref. 10, we can write the relaxation operator for the low level subsystem, which corresponds to Eq. (18), in its customary form in terms of the operator basis $\{\hat{n}_1, \hat{n}_2, \hat{\sigma}_1, \hat{\sigma}_2\}$:

$$\mathscr{L}_{12} = \begin{pmatrix} -\gamma_{12} & \gamma_{12} & 0 & 0 \\ w_{12} & -w_{12} & 0 & 0 \\ 0 & 0 & -(\gamma_{12} + w_{12})/2 & 0 \\ 0 & 0 & 0 & -(\gamma_{12} + w_{12})/2 \end{pmatrix},$$

where γ_{12} is the relaxation rate of the lower level subsystem. Given that the term $g'\hat{\xi}_1^-\hat{\sigma}_{12}^-$ in Eq. (18) describes a relaxation transition from level 1 to level 2 and the term $g\hat{\xi}_2^-\hat{\sigma}_{12}^+$ describes a transition in the opposite direction, we obtain the following expression for the relaxation owing to the contribution of four-photon processes:

$$\mathscr{L}_{12}^{4ph}=\frac{g_\Lambda^2}{4\Delta^2}\begin{pmatrix} -c^2\gamma_{13} & c^2\gamma_{13} & 0 & 0 \\ c'^2\gamma_{23} & -c'^2\gamma_{23} & 0 & 0 \\ 0 & 0 & -(c^2\gamma_{13}+c'^2\gamma_{23})/2 & 0 \\ 0 & 0 & 0 & -(c^2\gamma_{13}+c'^2\gamma_{23})/2 \end{pmatrix}, \tag{19}$$

where $c=g/g_\Lambda$, $c'=g'/g_\Lambda$, and $c^2+c'^2=1$.

Equation (19) implies that the contribution of four-photon processes to the relaxation rate constant of the lower level subsystem, Γ_{12} is given by

$$\Gamma_{12}^{4ph}=\frac{g_\Lambda^2}{8\Delta^2}\,(c'^2\gamma_{13}+c^2\gamma_{23})\approx\frac{g_\Lambda^2}{4\Delta^2}\,\gamma_{13}/2.$$

This contribution leads to a fundamental lower bound on Γ_{12}. As an example, for the Cs atom and laser field intensities of ~ 1 W/cm^2, we have $\Gamma_{12}^{4ph}\sim 10^{-3}\gamma_{13}/2\approx 10^4$ s^{-1}.

5. THE ROLE OF FOUR-PHOTON INTERACTIONS IN THE FORMATION OF AN ABSORPTION RESONANCE AND DISPERSION

The simplest experimental possibility for observing the dark resonance is to measure the transmission and/or dispersion (in atomic vapor) of exciting laser waves, whose independent detection is made easier by the relatively large difference Δ of the corresponding frequencies ω_L and ω_L' in the neighborhood of the resonance: $\Delta\gg\Gamma$. The real and imaginary parts of the corresponding refractive indices are expressed in an obvious way in terms of the operators for the corresponding dipole transitions (assuming that macroscopic volume averaging is valid) as

$$n_k''=\frac{\hbar c g_k N_0}{I_k}\,\mathrm{Im}\langle|k\rangle\langle 3|\rangle, \tag{20}$$

and

$$n_k'-1=\frac{\hbar c g_k N_0}{I_k}\,\mathrm{Re}\langle|k\rangle\langle 3|\rangle, \tag{21}$$

where g_k and I_k are the corresponding Rabi frequencies and intensities of the fields, with $k=1$ corresponding to frequency ω_L and $k=2$ to ω_L' (here $g_1=g$ and $g_2=g'$). To calculate Eqs. (20) and (21) in the stationary case it is necessary to find the stationary density matrix in the rotating wave approximation; this matrix is represented by the corresponding zero eigenvector $\langle 0|$ determined from the equation

$$\langle 0|\mathscr{L}_{RWA}=0.$$

When the contribution of four-photon processes to resonance dephasing is taken into account, the relaxation operator in the expression for $\mathscr{L}_{RWA}$ must include the field-dependent correction (19). Then the results of averaging in Eqs. (20) and (21) with a natural choice of basis for the vector representation of the density matrix (see Appendix A1) are described simply by the corresponding components of the vector $\langle 0|$.

Performing the corresponding analytic calculations and expressing the concentration of active atoms in terms of the pressure p, we obtain

$$n_k'-1=-0.0289p\lambda^3\,\frac{\tilde{\gamma}^2}{\tilde{g}_k^2}\,n_3\left[\tilde{\delta}-\frac{(\tilde{g}_1^2-\tilde{g}_2^2)\tilde{\Gamma}_{12}\tilde{\delta}+\tilde{g}_\Lambda^2\tilde{\delta}_R}{\tilde{g}_\Lambda^2\tilde{\Gamma}_{12}+4\tilde{\Gamma}_{12}^2+4\tilde{\delta}_R^2}\right], \tag{22}$$

and

$$n_k''=0.0289p\lambda^3\,\frac{\tilde{\gamma}^2}{\tilde{g}_k^2}\,n_3. \tag{23}$$

Here n_3 describes the population of the excited state calculated according to the formula

$$n_3=\left[3+\frac{2\tilde{\gamma}\tilde{g}_\Lambda^2}{\tilde{g}_1^2\tilde{g}_2^2}\,(1+\tilde{\delta}^2)+\frac{\tilde{\gamma}}{2}\right.$$
$$\left.\times\frac{\tilde{g}_\Lambda^4\tilde{\Gamma}_{12}+\tilde{g}_\Lambda^6/4-(\tilde{g}_2^2-\tilde{g}_1^2)^2\tilde{\Gamma}_{12}\tilde{\delta}^2+2\tilde{g}_\Lambda^2(\tilde{g}_2^2-\tilde{g}_1^2)\tilde{\delta}_R\tilde{\delta}}{\tilde{g}_1^2\tilde{g}_2^2(\tilde{\delta}_R^2+\tilde{\Gamma}_{12}^2+\tilde{\Gamma}_{12}\tilde{g}_\Lambda^2/4)}\right]^{-1},$$

where the tilde means that the corresponding variables are normalized to Γ. The arguments δ and δ_R depend on the velocity of the atom owing to the single-photon and residual Doppler effect, while the damping Γ_{12} in the lower level system is determined by the reciprocal time of flight of the atom (for a cuvette with pure vapor). Thus, in order to obtain computational data which model the experimental situation, Eqs. (22) and (23) must be averaged over a Maxwellian velocity distribution, which is done numerically.

Figure 5 shows calculated resonance absorption curves for $^2S_{1/2}\rightarrow{}^2P_{3/2}$ transitions in cesium and potassium for intense pump and weak probe fields. Although four-photon dephasing is not very important for cesium, in the case of

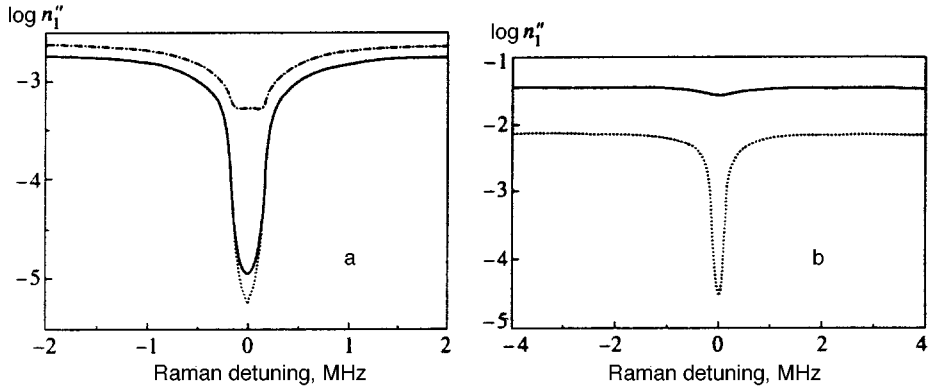


FIG. 5. Absorption resonances in cesium (a) and potassium (b) vapor including (smooth curve) and neglecting (dotted line) four-photon dephasing. The dot-dashed curve in Fig. a corresponds to a calculation taking four-photon dephasing into account where the separation between the lower levels was specially reduced by a factor of 10. The field intensities in the calculations were $I_1 = 0.01$ mW/cm^2 and $I_2 = 10$ mW/cm^2.

potassium, for which the splitting of the ground state for the isotope ^{41}K is only 0.25 GHz, i.e., almost two orders of magnitude smaller than the splitting in cesium, with a pump field intensity of 10 mW/cm^2 the resonance in the absence of a magnetic field is essentially unobservable. Thus, the four-photon mechanism for dephasing of the Λ-resonance establishes a fundamental limit of ''observability'' for the absorption resonance in strong fields, by imposing a limit either on the pump field intensity or on the magnitude of the splitting in the ground state.

6. CONCLUSION

Four-photon interactions, therefore, play a fundamental role in the formation of fluorescence spectra, as well as of the absorption spectra and/or dispersion of a resonance Λ-system during coherent population trapping.

A typical arrangement of a possible experiment for detecting a resonance fluorescence spectrum employing an atomic beam is shown in Fig. 1b. Experiments of this sort using an atomic beam and an atom trap have been described in detail by Gauthier *et al.*[14] and Stalgies *et al.*,[15] respectively. The directions of the atomic and laser beams are chosen to be mutually perpendicular so as to avoid the ordinary Doppler effect. The fluorescence spectrum is analyzed with a Fabry-Perot interferometer. Calculations for the fluorescence spectrum of two-level atoms show that for an atomic beam with 10^9 atoms/s·mm^2, using a 5-millimeter Fabry-Perot cavity with $Q \sim 10^4$ one can expect more than 10^5 photons/s from a volume of diameter ~ 100 μm. For a Λ-system, four-photon interactions, on one hand, reduce the fluorescence intensity relative to that of a two-level atoms by a factor of $(g_\Lambda/2\omega_{12})^2$, and, on the other, increase it by a factor of $(\gamma/\Gamma_{12})^{1/2}$. As an example, for the Cs atom a saturation intensity of 1.1 mW/cm and the corresponding parameter $g_\Lambda = 10^2\gamma$ are already achieved for a laser power of 30 mW at a wavelength of 852 nm focussed into a spot with a diameter of about 1 mm. For these experimentally easily realized parameters, we may expect, as the calculations show, a reduction in the scattering intensity of the Λ-system compared to two-level atoms by a factor of 2×10^3 and, therefore, to detect fewer than ~ 100 photons/s, which is not a problem for modern detection systems.

The authors thank D. N. Klyshko, A. Schentzle, and R. Wynands for fruitful discussions. One of the authors (V.N.Z.) is indebted to the Humboldt Foundation (Germany) for support.

Thus work was supported in part by the Volkswagen-Stiftung (Grant No. I/72944) and the Russian Fund for Fundamental Research (Grant No. 96-03-032867).

APPENDIX A

Dynamic superoperator of a Λ-system

Let us consider a Λ-configuration of the quantum mechanical levels of an atom (Fig. 1) acted on by two laser fields with frequencies close to a Raman resonance which is described by a Hamiltonian of the form

$$\hat{\mathscr{H}}_\Lambda = \hat{\mathscr{H}}_a + \hat{\mathscr{H}}_I, \tag{A1}$$

where

$$\hat{\mathscr{H}}_a = -\hbar\omega_{12}|2\rangle\langle 2| + \hbar\omega_{13}|3\rangle\langle 3|$$

is the intrinsic Hamiltonian of the atom (the energy of level 1 is taken to be zero, so that the projection operator $|1\rangle\langle 1|$ is absent in the Hamiltonian) and

$$\hat{\mathscr{H}}_I = \hbar g\ \cos(\omega_L t+\varphi)(|1\rangle\langle 3| + |3\rangle\langle 1|) + \hbar g' \times \cos(\omega_L' t+\varphi')(|2\rangle\langle 3| + |3\rangle\langle 2|)$$

is the Hamiltonian of the interaction of the atomic system with two light fields having frequencies ω_L and ω_L', including the dependence of the excitation on the phase of the field. The interaction constants, i.e., the Rabi frequencies, depend on the amplitudes A_{ω_L} and $A_{\omega_L'}$ of the external field an on the dipole matrix elements d_{13} and d_{23}:

$$g = \frac{1}{\hbar}\, d_{13}A_{\omega_L}, \quad g' = \frac{1}{\hbar}\, d_{23}A_{\omega_L'}. \tag{A2}$$

Only the case when a single-photon resonance is present is of interest, i.e., when ω_L and ω_L' are close, respectively, to ω_{13} and ω_{23}. We can rewrite $\hat{\mathscr{H}}_a$ in the form

$$\hat{\mathscr{H}}_a = \hat{\mathscr{H}}_0 + \hat{\mathscr{H}}_\delta, \tag{A3}$$

where the ''unperturbed'' Hamiltonian

$$\hat{\mathscr{H}}_0=\hbar(\omega_L|3\rangle\langle 3|-\Delta|2\rangle\langle 2|), \tag{A4}$$

including the biharmonic frequency detuning $\Delta=\omega_L'-\omega_L\approx\omega_{12}$, describes free precession with the two laser frequencies. The ''perturbing'' Hamiltonian $\hat{\mathscr{H}}_b$ can be written in the form

$$\hat{\mathscr{H}}_\delta=-\hbar\delta|3\rangle\langle 3|+\hbar\delta_R|2\rangle\langle 2|,$$

where

$$\delta=\omega_L-\omega_{13},\quad \delta_R=\omega_L'-\omega_L-\omega_{12}$$

describe the single-photon detuning for the $|1\rangle\rightarrow|3\rangle$ transition and the two-photon Raman detuning, respectively. Both detunings can be zero with a suitable choice of laser frequencies.

The dynamics of an atomic system with the Hamiltonian (A3) can be characterized as a combination of fast (owing to $\hat{\mathscr{H}}_0$) and slow (owing to $\hat{\mathscr{H}}_\delta$) precessions, so it is appropriate to shift to a representation of the interaction with the unperturbed unitary transformation

$$\mathscr{U}_0(t)=\exp\left[-\frac{i}{\hbar}\hat{\mathscr{H}}_0 t\right].$$

In the rotating wave approximation[6] we can neglect the rapidly oscillating terms, so that the Hamiltonian (A1) takes the form

$$\hat{\mathscr{H}}_\Lambda=\hat{\mathscr{H}}_\delta+\hat{\mathscr{H}}_p=\hbar[-\delta|3\rangle\langle 3|+\delta_R|2\rangle\langle 2|+(g_\Lambda/2)(|c\rangle\times\langle 3|+\text{H.c.})] \tag{A5}$$

and is the effective Hamiltonian in this approximation. Here we have introduced the bound state ($|c\rangle$) and the unbound state ($|n\rangle$) orthogonal to it:

$$|c\rangle=g_\Lambda^{-1}(ge^{-i\varphi}|1\rangle+g'e^{-i\varphi'}|2\rangle),$$

$$|n\rangle=-g_\Lambda^{-1}(g'e^{-i\varphi}|1\rangle-ge^{-i\varphi'}|2\rangle). \tag{A6}$$

The state $|c\rangle$ is associated with excitation of the level e with an effective coupling constant of $g_\Lambda=\sqrt{g^2+g'^2}$. For zero Raman detuning ($\delta_R=0$), it is easy to see that the Hamiltonian (A5) describes a two-level system. This can be demonstrated most clearly by substituting the expression $|2\rangle\langle 2|$ corresponding to the inverse transformation (A6) in Eq. (A5):

$$|1\rangle=g_\Lambda^{-1}(g'e^{i\varphi'}|c\rangle+ge^{-i\varphi}|n\rangle),$$

$$|2\rangle=g_\Lambda^{-1}(ge^{i\varphi}|c\rangle-g'e^{-i\varphi'}|n\rangle),$$

which leads to the result

$$\hat{\mathscr{H}}_\Lambda=\hbar\left[\left(\delta+\frac{\delta_R g^2}{g_\Lambda^2}\right)|c\rangle\langle c|+\left(\delta+\frac{\delta_R g'^2}{g_\Lambda^2}\right)|n\rangle\langle n|\right]+\hbar\delta_R\frac{gg'}{g_\Lambda^2}[e^{i(\varphi-\varphi')}|c\rangle\langle n|+\text{H.c.}]+\frac{\hbar g_\Lambda}{2}(|c\rangle\langle 3|+\text{H.c.}).$$

In the basis $\{|3\rangle,|c\rangle,|n\rangle\}$ the corresponding matrix has the form

$$\hat{\mathscr{H}}_\Lambda=\begin{pmatrix}0 & \hbar g_\Lambda/2 & 0\\ \hbar g_\Lambda/2 & \hbar\delta & 0\\ 0 & 0 & \hbar\delta\end{pmatrix}+\frac{\hbar\delta_R}{g_\Lambda^2}\begin{pmatrix}0 & 0 & 0\\ 0 & g^2 & gg'e^{i(\varphi-\varphi')}\\ 0 & gg'e^{-i(\varphi-\varphi')} & g'^2\end{pmatrix}$$

and, for $\delta_R=0$, can quickly be expanded in terms of the 2×2 matrix of a two-level system ''dressed'' with the atomic field and the 1×1 matrix of a single unbound state, i.e., the excited and bound states form an effective two-level system $|e\rangle\oplus|c\rangle$. For simplicity we redefine $|1\rangle$ as $\exp(-i\varphi)|1\rangle$ and $|2\rangle$ as $\exp(-i\varphi')|2\rangle$, so that we can rewrite Eq. (A6) in the form

$$|c\rangle=g_\Lambda^{-1}(g|1\rangle+g'|2\rangle),\quad |n\rangle=-g_\Lambda^{-1}(g'|1\rangle-g|2\rangle),$$

which does not contain the phase factors explicitly.

With this representation of the Hamiltonian in the rotating wave approximation, the corresponding dynamic part of the Liouvillian has the form

$$\mathscr{L}_\Lambda=\frac{i}{\hbar}[\hat{\mathscr{H}}_\Lambda,\odot]. \tag{A7}$$

The complete Liouvillian $\mathscr{L}_{RWA}$ also contain a relaxation operator which is specified phenomenologically here.

A1. Transformation of the Liouvillian in the rotating wave approximation

The initial representation of the Liouvillian in the rotating wave approximation is a matrix in the nonhermitian basis $\{\hat{e}_k\}=\hat{P}_{\alpha,\beta}=|\alpha\rangle\langle\beta|$, where $k=(\alpha,\beta)$ and $\alpha,\beta=1,2,3$, which can be represented by the following complex matrix elements using Eqs. (A7):

$$\mathscr{L}_{RWA}=\begin{pmatrix}
-\gamma-\gamma' & \gamma & \gamma' & 0 & 0 & -\frac{ig}{2} & \frac{ig}{2} & -\frac{ig'}{2} & \frac{ig'}{2} \\
0 & -\gamma_{12} & \gamma_{12} & 0 & 0 & \frac{ig}{2} & -\frac{ig}{2} & 0 & 0 \\
0 & w & -w & 0 & 0 & 0 & 0 & \frac{ig'}{2} & -\frac{ig'}{2} \\
0 & 0 & 0 & i\delta_R-\Gamma_{12} & 0 & \frac{ig'}{2} & 0 & 0 & -\frac{ig}{2} \\
0 & 0 & 0 & 0 & -i\delta_R-\Gamma_{12} & 0 & -\frac{ig'}{2} & \frac{ig}{2} & 0 \\
-\frac{ig}{2} & \frac{ig}{2} & 0 & \frac{ig'}{2} & 0 & -i\delta-\Gamma_{13} & 0 & 0 & 0 \\
\frac{ig}{2} & -\frac{ig}{2} & 0 & 0 & -\frac{ig'}{2} & 0 & i\delta-\Gamma_{13} & 0 & 0 \\
-\frac{ig'}{2} & 0 & \frac{ig'}{2} & 0 & \frac{ig}{2} & 0 & 0 & -i\delta-\Gamma_{23} & 0 \\
\frac{ig'}{2} & 0 & -\frac{ig'}{2} & -\frac{ig}{2} & 0 & 0 & 0 & 0 & i\delta-\Gamma_{23}
\end{pmatrix}.$$

For converting to the more convenient Hermitian bases, we can introduce two transformations V_c and V_{c1} of the form

$$V_c=\begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & g/g_\Lambda & 0 & g'/g_\Lambda & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & g/g_\Lambda & 0 & g'/g_\Lambda \\
0 & 0 & 0 & 0 & 0 & -g'/g_\Lambda & 0 & g/g_\Lambda & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -g'/g_\Lambda & 0 & g/g_\Lambda
\end{pmatrix},$$

$$V_{c1}=\begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1/\sqrt{2} & 1/\sqrt{2} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -i/\sqrt{2} & i/\sqrt{2} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1/\sqrt{2} & 1/\sqrt{2} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -i/\sqrt{2} & i/\sqrt{2} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/\sqrt{2} & 1/\sqrt{2} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -i/\sqrt{2} & i/\sqrt{2}
\end{pmatrix}. \tag{A8}$$

The transformation V_c introduces two pairs of polarization operators $\hat{P}_c, \hat{P}_c^+$ and $\hat{P}_n, \hat{P}_n^+$ for transitions to the excited level from the bound and unbound states, while V_{c1} introduces the Hermitian cosine-sine operators (analogs of the coordinates and momenta or the Pauli matrices $\hat{\sigma}_1$, $\hat{\sigma}_2$ in a two-level system):

$$\hat{q}_g=(\hat{P}_{12}+\hat{P}_{21})/\sqrt{2}, \quad \hat{p}_g=-i(\hat{P}_{12}-\hat{P}_{21})/\sqrt{2},$$
$$\hat{q}_c=(\hat{P}_c+\hat{P}_c^+)/\sqrt{2}, \quad \hat{p}_c=-i(\hat{P}_c-\hat{P}_c^+)/\sqrt{2},$$
$$\hat{q}_n=(\hat{P}_n+\hat{P}_n^+)/\sqrt{2}, \quad \hat{p}_n=-i(\hat{P}_n-\hat{P}_n^+)/\sqrt{2}. \tag{A9}$$

Here the subscripts g, c, and n correspond to the ground $(1\leftrightarrow 2)$, bound $(c\leftrightarrow 3)$, and unbound $(n\leftrightarrow 3)$ subsystems.

After applying the transformations (A8) to the Liouvillian $\mathscr{L}_{RWA}$ in the rotating wave approximation, for the transformed operator $L_{RWA}=V_{c1}V_cL_{RWA}V_c^{-1}V_{c1}^{-1}$ we obtain

$$\mathscr{L}_{RWA}=\left(\begin{array}{ccc|cc|cc|cc}
-\gamma-\gamma' & \gamma & \gamma' & 0 & 0 & 0 & \frac{g_\Lambda}{\sqrt{2}} & 0 & 0 \\
0 & -\gamma_{12} & \gamma_{12} & 0 & 0 & 0 & -\frac{\xi^2 g_\Lambda}{\sqrt{2}} & 0 & \frac{\xi\eta g_\Lambda}{\sqrt{2}} \\
0 & w & -w & 0 & 0 & 0 & -\frac{\eta^2 g_\Lambda}{\sqrt{2}} & 0 & -\frac{\xi\eta g_\Lambda}{\sqrt{2}} \\
\hline
0 & 0 & 0 & -\Gamma_{12} & -\delta_R & 0 & -\xi\eta g_\Lambda & 0 & -(\xi^2-\eta^2)\frac{g_\Lambda}{2} \\
0 & 0 & 0 & \delta_R & -\Gamma_{12} & 0 & 0 & -\frac{g_\Lambda}{2} & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & -\Gamma_c & \delta & \Delta\Gamma & 0 \\
-\frac{g_\Lambda}{\sqrt{2}} & \frac{\xi^2 g_\Lambda}{\sqrt{2}} & \frac{\eta^2 g_\Lambda}{\sqrt{2}} & \xi\eta g_\Lambda & 0 & -\delta & -\Gamma_c & 0 & \Delta\Gamma \\
\hline
0 & 0 & 0 & 0 & \frac{g_\Lambda}{2} & \Delta\Gamma & 0 & -\Gamma_n & \delta \\
0 & -\frac{\xi\eta g_\Lambda}{\sqrt{2}} & \frac{\xi\eta g_\Lambda}{\sqrt{2}} & (\xi^2-\eta^2)\frac{g_\Lambda}{2} & 0 & 0 & \Delta\Gamma & -\delta & -\Gamma_n
\end{array}\right), \tag{A10}$$

where we have used the following notation

$$\xi=g/g_\Lambda, \quad \eta=g'/g_\Lambda,$$

$$\Gamma_c=\xi^2\Gamma_{13}+\eta^2\Gamma_{23}, \quad \Gamma_n=\eta^2\Gamma_{13}+\xi^2\Gamma_{23},$$

$$\Delta\Gamma=\xi\eta(\Gamma_{13}-\Gamma_{23}).$$

The block structure of the transformed dynamic superoperator L_{RWA} indicated by the continuous lines in Eq. (A10) is discussed in more detail in Appendix A2.

As opposed to the initial complex representation, the transformed operator $\mathscr{L}_{RWA}$ has real matrix elements, since it corresponds to the Hermitian basis $\{\hat{e}_k\}$.

For a symmetric Λ-system, with $\Gamma_{13}=\Gamma_{23}$, $g=g'$, and, therefore, $\Delta\Gamma=0$ and $\xi=\eta$, Eq. (A10) takes the form

$$L_{RWA}=\left(\begin{array}{ccc|cc|cc|cc}
-\gamma-\gamma' & \gamma & \gamma' & 0 & 0 & 0 & g_\Lambda/\sqrt{2} & 0 & 0 \\
0 & -\gamma_{12} & \gamma_{12} & 0 & 0 & 0 & -g_\Lambda/2\sqrt{2} & 0 & g_\Lambda/2\sqrt{2} \\
0 & w & -w & 0 & 0 & 0 & -g_\Lambda/2 & 0 & -g_\Lambda/2\sqrt{2} \\
\hline
0 & 0 & 0 & -\Gamma_{12} & -\delta_R & 0 & -g_\Lambda/2 & 0 & 0 \\
0 & 0 & 0 & \delta_R & -\Gamma_{12} & 0 & 0 & -g_\Lambda/2 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & -\Gamma & \delta & 0 & 0 \\
-g_\Lambda/\sqrt{2} & g_\Lambda/2\sqrt{2} & g_\Lambda/2 & g_\Lambda/2 & 0 & -\delta & -\Gamma & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & g_\Lambda/2 & 0 & 0 & -\Gamma & \delta \\
0 & -g_\Lambda/2\sqrt{2} & g_\Lambda/2\sqrt{2} & 0 & 0 & 0 & 0 & -\delta & -\Gamma
\end{array}\right).$$

A2. The block structure of the dynamic superoperator in the rotating wave approximation

The physical significance of the superoperator L_{RWA} defined by Eq. (A10) becomes most transparent on examining its block structure. It is convenient to break the matrix (A10) up into blocks in accordance with a definite set of polarization variables (A9). By analogy, the matrix blocks can be numbered with a subscript p corresponding to the populations and by the subscripts g, c, and n corresponding to the polarizations of ground $(1\leftrightarrow 2)$, bound $(1\leftrightarrow 3)$, and unbound $(1\leftrightarrow 2)$ subsystems. In this notation, the matrix (A10) appears as

$$L_{RWA}=\begin{pmatrix} L_{pp} & 0 & -L_{cp}^{T} & -L_{np}^{T} \\ 0 & L_{gg} & -L_{cg}^{T} & -L_{ng}^{T} \\ L_{cp} & L_{cg} & L_{cc} & -L_{nc}^{T} \\ L_{np} & L_{ng} & L_{nc} & L_{nn} \end{pmatrix}. \tag{A11}$$

It consists of nine nonzero independent blocks. The diagonal block L_{pp} describes the dynamics of the populations n_3, n_c, and n_n, and the blocks L_{gg}, L_{cc}, and L_{nn}, the polarization dynamics, respectively, of the ground state and of the bound and unbound subsystems. The five nonzero nondiagonal matrices describe the dynamics of the Λ-system owing to coupling among the above basis variables. The antisymmetry of these five blocks is a consequence of the purely oscillatory character of the dynamics resulting from the interaction with the external field, while the inner dynamics includes relaxation so it is also represented by the matrix elements which yield nonzero real components in the eigenvalues of the matrix $\mathscr{L}_{RWA}$.

It is easy to see from the block structure of (A11) that there is no connection between the populations and polarizations of the ground state, since $\mathscr{L}_{gp}=0$. This reflects the fact that the exciting field acts directly only on transitions into the excited state, while single-photon excitation of the ground state is absent.

The block representation given here for the dynamic superoperator in the rotating wave approximation is convenient for qualitative discussions of the effect of the parameters of the Λ-system on its dynamics, since it reduces to changes in only the inner structure of the blocks in the representation (A11).

APPENDIX B

Superoperator calculation of the general formula for the fine structure of the spectrum

Let us calculate the two-time correlation function (10) that determines the atomic fluorescence spectrum:

$$\mathscr{K}(\tau)=\langle\hat{\rho}_0 S(0,t)|\hat{\sigma}^{-}(t)[S(t,t+\tau)\hat{\sigma}^{+}(t+\tau)]\rangle. \tag{B1}$$

Here the total evolution superoperator (see Eq. (2)) has the form

$$S(0,t)=S_{RWA}(t)\widetilde{S}_0(t), \tag{B2}$$

where $S_{RWA}(t)$ is the superoperator in the rotating wave approximation and $\widetilde{S}_0(t)$ is the superoperator for the perturbed evolution owing to the nonresonant interaction. The superoperator $\widetilde{S}_0(t)$ describes the transformation of an initial system Hamiltonian $\hat{\mathscr{H}}(t)$ of the form

$$\hat{\mathscr{H}}(t)\to\hat{\mathscr{H}}_0+\delta\hat{\mathscr{H}}(t)=\hat{\mathscr{H}}_0+\frac{\hbar g_\Lambda}{2}(\hat{\tau}e^{i\Delta t}+\hat{\tau}^{+}e^{-i\Delta t}), \tag{B3}$$

where the operator $\hat{\tau}$ is defined by Eq. (9). In first-order perturbation theory we can introduce a superoperator $\widetilde{S}_0(t)$, corresponding to the Hamiltonian (B3), in the form

$$\widetilde{S}_0(t)=S_0(t)+\delta S_0(t)=S_0(t)+\int_0^t\delta\mathscr{L}_p(\tau)d\tau S_0(t),$$

where the perturbation Liouvillian has the form

$$\delta\mathscr{L}_p(\tau)=\frac{i}{\hbar}[\delta\hat{\mathscr{H}}(\tau),\odot]. \tag{B4}$$

Integrating $\delta S_0(t)$ with respect to τ and using Eq. (B4) together with Eq. (B3), we obtain

$$\delta S_0(t)=\frac{g_\Lambda}{2\Delta}[\hat{\tau}e^{i\Delta t}-\hat{\tau}^{+}e^{-i\Delta t},\odot]S_0(t), \tag{B5}$$

which describes oscillations at a frequency Δ.

The superoperator $S_0(t)$ in Eq. (B5) describes the unperturbed dynamics represented by the Hamiltonian (A4), which accounts for the free precession of all the Λ-system transitions. The latter is represented in the form

$$\begin{aligned} S_0(t)=&S_{12}^{+}e^{-i\Delta t}\oplus S_{12}^{+}e^{i\Delta t}\oplus S_{13}^{+}e^{-i\omega_L t}\oplus S_{13}^{+}e^{i\omega_L t} \\ &\oplus S_{23}^{+}e^{-i\omega_L' t}\oplus S_{23}^{+}e^{i\omega_L' t}\oplus P_0, \end{aligned} \tag{B6}$$

where the matrices $S_{kl}^{\pm}$ for the corresponding superoperators are the one-dimensional eigen-projectors on the corresponding intrinsic precession of the variables and P_0 is the projector on the three-dimensional subspace of the nonoscillatory variables, i.e., the populations. After substituting Eq. (B6) in Eq. (B5), we obtain

$$\begin{aligned} \delta S_0(t)=\frac{g\Lambda}{2\Delta}\{&[\hat{\tau},\odot]S_{13}^{+}\exp[-i(\omega_L-\Delta)t] \\ &-[\hat{\tau}^{+},\odot]S_{23}^{+}\exp[-i(\omega_L'+\Delta)t] \\ &-[\hat{\tau}^{+},\odot]S_{13}^{-}\exp[i(\omega_L-\Delta)t] \\ &+[\hat{\tau},\odot]S_{23}^{-}\exp[i(\omega_L'+\Delta)t]\}. \end{aligned} \tag{B7}$$

Then substituting Eq. (B7) into Eq. (B2) and using Eq. (B1) together with the relation $\hat{\rho}_0 S(0,t)\to\langle 0|$ for $t\to\infty$ because of the damping of all the eigen-oscillations corresponding to the nonzero eigenvalues, we can finally write the correlation function in the form

$$\mathscr{K}(t,t+\tau)=\langle 0|\delta S_0(t)\hat{\sigma}^{-}\cdot\exp(\mathscr{L}_{RWA}\tau)\delta S_0(t+\tau)\hat{\sigma}^{+}\rangle, \tag{B8}$$

where the symbol ''$\cdot$'' denotes the product of transformed operators and $\sigma^{\pm}=\sigma_{13}^{\pm}+\sigma_{23}^{\pm}$ is the sum of the complex amplitudes oscillating at the optical frequencies.

If we then apply δS_0 to $\hat{\sigma}^{\pm}$ and recall that $\hat{\sigma}_{13}^{\pm}$ and $\hat{\sigma}_{23}^{\pm}$ are eigenvectors for the eigen-projectors $S_{13}^{\pm}$ and $S_{23}^{\pm}$, we obtain

$$\begin{aligned} \delta S_0(t)\hat{\sigma}^{-}=-\frac{g_\Lambda}{2\Delta}\{&[\hat{\tau}^{+},\hat{\sigma}_{13}^{-}]\exp[i(\omega_L-\Delta)t] \\ &+[\hat{\tau},\hat{\sigma}_{23}^{-}]\exp[i(\omega_L'+\Delta)t], \end{aligned} \tag{B9}$$

and

$$\begin{aligned} \delta S_0(t+\tau)\hat{\sigma}^{+}=\frac{g_\Lambda}{2\Delta}\{&[\hat{\tau},\hat{\sigma}_{13}^{+}]\exp[-i(\omega_L-\Delta)(t+\tau)] \\ &+[\hat{\tau}^{+},\hat{\sigma}_{23}^{+}]\exp[-i(\omega_L'+\Delta)(t+\tau)]\}, \end{aligned} \tag{B10}$$

where the commutators of the complex amplitudes of the $1\leftrightarrow 3$ and $2\leftrightarrow 3$ transitions with $\hat{\tau}$ and $\hat{\tau}^+$ are given by

$$[\tau^+,\hat{\sigma}_{13}^-]=-\frac{\hat{\sigma}_{12}^-}{\sqrt{2}}, \quad [\tau,\hat{\sigma}_{23}^-]=-\frac{\hat{\sigma}_{12}^+}{\sqrt{2}},$$

$$[\tau,\hat{\sigma}_{13}^+]=\frac{\hat{\sigma}_{12}^+}{\sqrt{2}}, \quad [\tau^+,\hat{\sigma}_{23}^+]=\frac{\hat{\sigma}_{12}^-}{\sqrt{2}}.$$

After substitution of Eqs. (B9) and (B10) in Eq. (B8) and leaving out terms which oscillate relative to t, with the dynamical representation in the rotating wave approximation in terms of the eigenvectors and corresponding eigenvalues Eq. (B8) finally takes the form of Eq. (12).

APPENDIX C

Transformation of the time evolution superoperator in the rotating wave approximation

For a Λ-system with a time-independent Hamiltonian $\hat{\mathscr{H}}$, the time evolution superoperator is unitary and is given by an exponential $\mathscr{S}(t)=\exp(\mathscr{L}_{\mathscr{H}}t)$ with a purely dynamic Liouvillian of the type (A7) and can be written in the form

$$\mathscr{L}(t)=\mathscr{U}(t)\odot\mathscr{U}^{-1}(t)=\sum_{\alpha,\beta}\exp[-i(\omega_\alpha-\omega_\beta)t]|\alpha\rangle\langle\alpha|\odot|\beta\rangle\langle\beta|, \tag{C1}$$

where the ω_α and $|\alpha\rangle$ are the Bohr eigenfrequencies and the corresponding eigenvectors of the Hamiltonian, while the unitary transformation $\mathscr{U}(t)$ is specified by the relation

$$\mathscr{U}(t)=\exp\left[-\frac{i}{\hbar}\hat{\mathscr{H}}t\right].$$

Subsequently, we can use the interaction representation for the superoperator $\mathscr{L}_{\mathscr{H}}+\mathscr{L}_r$, which differs from the purely dynamical Liouvillian corresponding to $\mathscr{S}(t)$ in that it includes the relaxation superoperator $\mathscr{L}_r$, and treat $\mathscr{L}_{\mathscr{H}}$ as the Liouvillian for the unperturbed time evolution $\mathscr{S}(t)$. In the interaction representation the time dependent relaxation superoperator has the form

$$\mathscr{L}_r^I(t)=\sum_{\alpha,\beta;\mu,\nu}\exp\{i[(\omega_\alpha-\omega_\beta)-(\omega_\mu-\omega_\nu)]t\}L_{\alpha\beta,\mu\nu}|\alpha\rangle\langle\mu|\odot|\nu\langle\beta|, \tag{C2}$$

where

$$L_{\alpha\beta,\mu\nu}=\sum_{k,l;m,n}\langle\alpha|k\rangle\langle l|\beta\rangle L_{kl,mn}\langle m|\mu\rangle\langle\nu|n\rangle$$

denotes the matrix elements of the relaxation superoperator in terms of the eigen-basis of the Hamiltonian $\hat{\mathscr{H}}$.

Let us consider the important special case in which the oscillations in Eq. (C.2) are fast compared to the rates of all the relaxation processes, so that it is possible to average over these oscillations. Then the resulting effective (''reduced'') relaxation operator has the form

$$\mathscr{L}_{re}=\sum_{\alpha,\beta}L_{\alpha\alpha,\beta\beta}|\alpha\rangle\langle\beta|\odot|\beta\rangle\langle\alpha| + \sum_{\alpha\neq\beta}L_{\alpha\beta,\alpha\beta}|\alpha\rangle\langle\alpha|\odot|\beta\rangle\langle\beta|, \tag{C3}$$

where it is assumed that all the frequencies $\omega_{\alpha\beta}$ corresponding to the $\beta\rightarrow\alpha$ atomic transitions ($\alpha\neq\beta$) are different. The first term in Eq. (C.3) describes the relaxation of the populations owing to $\beta\rightarrow\alpha$ transitions from other levels ($\beta\neq\alpha$) and radiative decay ($\beta=\alpha$). The second term describes the relaxation of the polarization variables. The corresponding matrix is $n\times n$, where $n=3$ is the number of levels in a Λ-system.

The superoperator (C3) commutes with the dynamical Liouvillian, since they have an eigenbasis in common. Given this circumstance, the relaxation of the atomic oscillations is simply described by the corresponding damping rates

$$\Gamma_{\alpha\beta}=-\mathrm{Re}\,L_{\alpha\beta,\alpha\beta}.$$

If these quantities are all nonzero, then the stationary (zero-) vector $\langle\hat{\rho}^{st}|$ has nonzero components only in population relaxation space and is actually described by the n-component zero-vector ρ_α^{st} of the $n\times n$ submatrix $L_{\alpha\alpha,\beta\beta}$.

[1] B. D. Agap'ev, M. B. Gornyĭ, B. G. Matisov, and Yu. V. Rozhdestvenskiĭ, Usp. Fiz. Nauk **163**, 1 (1993).
[2] E. Arimondo, Progress in Optics **35**, 257 (1996).
[3] M. R. Ferguson, Z. Ficek, and B. J. Dalton, Phys. Rev. A **54**, 2379 (1996).
[4] L. M. Narducci, M. O. Scully, G.-L. Oppo, R. Pu, and J. R. Tredicce, Phys. Rev. A **42**, 1630 (1990).
[5] B. A. Grishanin, Kvant. Elektrodinamika **6**, 1649 (1979).
[6] L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics*, Cambridge Univ. Press (1995).
[7] B. R. Mollow, Phys. Rev. **188**, 1969 (1969).
[8] I. R. Shen, *The Principles of Nonlinear Optics*, Wiley, New York (1984).
[9] R. Glauber, in *Quantum Optics and Electronics*, Gordon and Breach, N.Y. (1965), p. 63.
[10] B. A. Grishanin, *Quantum Electrodynamics for Radio Physicists* [in Russian], Izd-vo. Mosk. Un-ta., Moscow (1981).
[11] Van M. Lu, G. Mainfray, C. Manus, and I. Tugov, Phys. Rev. A **7**, 91 (1973).
[12] O. Schmidt, K.-M. Knaak, R. Wynands, and D. Meschede, Phys. Rev. A **54**, R27 (1996); S. Brandt, A. Nagel, R. Wynands, and D. Meschede, Phys. Rev. A **56**, R1063 (1997).
[13] E. Arimondo, Phys. Rev. A **54**, 2216 (1996).
[14] D. J. Gauthier, Y. Zhu, and T. W. Mossberg, Phys. Rev. Lett. **66**, 2460 (1991).
[15] Y. Stalgies, I. Siemers, B. Appasamy, and P. E. Toschek, Europhys. Lett. **35**, 259 (1996).

Translated by D. H. McNeill

Stimulated scattering and wavefront conjugation in an inhomogeneous plasma

Yu. V. Rostovtsev and I. V. Khazanov

Institute of Applied Physics, Russian Academy of Sciences, 603600 Nizhniĭ Novgorod, Russia
(Submitted 31 July 1996)
Zh. Éksp. Teor. Fiz. **113**, 168–180 (January 1998)

In the context of the problem of stimulated scattering we discuss the mechanism of wavefront conjugation in an inhomogeneous plasma proposed relatively recently, associated with a difference in suppression of scattering (due to inhomogeneities) for the inverted and uninverted components. We analyze the solutions of the integro-differential equations describing this process both numerically and analytically for different sound attenuation lengths (ν^{-1}). It is shown that for this effect to exist it is necessary that ν not be too small. We also consider extinction of the inverted wave in terms of this mechanism. © *1998 American Institute of Physics.* [S1063-7761(98)01101-9]

Along with processes of stimulated scattering in plane waves, which have already been investigated to a significant extent, analogous processes in waves with complex spatial structure are of interest, especially those in which efficient wavefront conjugation is possible.[1,2] Different methods of solution of this problem have been discussed in many works, and the stimulated-scattering method of wavefront conjugation (SS-WFC) is one of the main ones and possesses the important advantage over the others that it does not require high-power reference waves of high (diffraction) quality for its realization. In SS-WFC, in fact, self-conjugation of the wavefront takes place. This mechanism of wavefront conjugation is sometimes also called statistical since before being directed into the active (nonlinear) medium, where in fact stimulated scattering takes place, the radiation that is to be inverted is first passed through a screen with small-scale random inhomogeneities, as a result of Fresnel diffraction on which the structure of the pump field in the active medium is formed as a set of strong inhomogeneities (the so-called speckle structure). In the classical scheme of wavefront conjugation in such a field local coupling is necessary between the amplitude of the medium density perturbation and the amplitude of the electromagnetic waves, i.e, the condition

$$\nu l \gg 1 \tag{1}$$

must be fulfilled, where l is the characteristic scale of the pump inhomogeneities (longitudinal correlation length of the field). Kurin[3] addressed the possibility of substantially broadening the region in which this mechanism acts and avoiding the limitations of condition (1). In fact, on the basis of comparatively simple, interesting, but not obvious estimates he concluded that this is possible for an inhomogeneous plasma. The purpose of the present paper is to carry out a more rigorous examination of this question.

Thus, consider a longitudinally (in the x direction) inhomogeneous plasma with characteristic inhomogeneity length scale L_N. We let a pump wave, whose field we assume to be given, propagate in this direction, with amplitude

$$a_1 = \Psi_1(x,\mathbf{r})\exp[-i\varphi_1(x)], \tag{2}$$

frequency ω_1 and wave number k_1. In expression (2) $\varphi_1(x)$ is the phase excursion (phase detuning) due to the variation in the plasma density [we define $\varphi_1(x=0)=0$; the point $x=0$ will be called below the synchronization point]. In contrast to the case of plane waves, here we allow the function Ψ_1 to have a complex transverse structure (dependence on $\mathbf{r}$) with characteristic scale ρ much less than the beam width, varying from cross section to cross section (dependence on x) and constant mean intensity over the cross section, i.e.,

$$\frac{1}{S}\int |\Psi_1|^2 d^2r = \langle |\Psi_1|^2 \rangle = \text{const}, \tag{3}$$

where S is the cross-sectional area, and the angular brackets denote averaging over the ensemble of realizations of the speckle inhomogeneity. Let us consider the process of stationary inverse stimulated Brillouin scattering (SBS) of such a wave, about which more will be said below, developing in the vicinity of the synchronization point at which the well-known conditions

$$\omega_1 = \omega_2 + \omega_3, \quad \mathbf{k}_1 = \mathbf{k}_2 + \mathbf{k}_3, \quad \omega_3 \ll \omega_{1,2},$$

$$\mathbf{k}_1 \approx -\mathbf{k}_2, \quad k_3 \approx 2k_1$$

are satisfied (the indices 2 and 3 pertain, respectively, to the Stokes wave and the ion acoustic wave, and the notation of the corresponding quantities here and below are analogous to the preceding). The equations describing it differ from the corresponding equations for the usual SBS in an inhomogeneous medium by the presence of additional terms with transverse Laplacian $\Delta_\perp$ of the amplitudes of the waves, i.e., this in fact is the parabolic equation for the amplitudes with a quadratic nonlinearity, taking into account the inhomogeneities[3] $\sigma_i(x)$ of the medium.[1)] Expressing from one of them the amplitude of the ion acoustic wave with the help of the Green's function (or, what is the same thing, writing down its solution implicitly), it is not difficult to proceed from the system of equations to one integro-differential equation for the amplitude of the Stokes wave, which can be significantly simplified in the present case of pump speckle structure. Specifically, we represent its solution as the sum of the inverted and uninverted wave:

$$a_2 = [A_2(x)\Psi_1^*(x,\mathbf{r}) + \tilde{A}_2(x)\text{const}_2]\exp[i\varphi_2(x)]. \tag{4}$$

 1063-7761/98/010093-08$15.00

where for the latter (uncorrelated component) we choose a plane wave as its ''typical representative'' and apply the well-known procedure[1,2] of statistical averaging, taking $\Psi_1(x,\mathbf{r})$ to be a Gaussian, statistically uniform random field. We thus obtain equations for the functions $A_2(x)$ and $\widetilde{A}_2(x)$ describing the amplification of both components:[3]

$$\frac{d(\mathscr{A}_2,\widetilde{\mathscr{A}}_2)}{dy}=-\eta\beta\,\exp(-i\delta y^2)\int_{-\mathscr{L}}^{y}\frac{[\mathscr{A}_2(\widetilde{y}),\widetilde{\mathscr{A}}_2(\widetilde{y})]}{1-i(y-\widetilde{y})\theta}\times\exp[-h(y-\widetilde{y})]\exp(i\delta\widetilde{y}^2)d\widetilde{y}. \tag{5}$$

In Eq. (5) we have introduced the notation $\mathscr{A}_2(y)=A_2(x)$, $\widetilde{\mathscr{A}}_2(y)=\widetilde{A}_2(x)$, $y=\gamma x$,

$$\delta=\frac{1}{2\gamma^2}(\varphi''_{1xx}+\varphi''_{2xx})|_{x=0}\sim\frac{k_1}{\gamma^2 L_N}$$

[here we make use of the usual simplification of replacing the phase detuning $\varphi_i(x)$ by the first term of the Taylor series expansion about the synchronization point], $\beta=\Gamma_{NL}^2/v_2v_3\gamma^2$ [Γ_{NL} is the temporal growth rate of the decay instability in the field of a plane wave with intensity equal to the mean intensity of the speckle structure, in an unbounded, homogeneous plasma, $v_{2,3}$ are the group velocities of the Stokes and the ionic-acoustic wave, $\gamma=l^{-1}\equiv(k_1\rho^2)^{-1}$], $h=\nu/\gamma$, $\mathscr{L}=\gamma L$ [(L,L) is the region of nonlinear interaction (scattering) centered on the synchronization point $x=0$], $\eta=1$, 2, and $\theta=0$, 1, 2. The values $\eta=2$, $\theta=2$ in system (5) describe the inverted wave, the values $\eta=1$, $\theta=1$ describe the uninverted wave. For $\theta=0$ system of Eqs. (5) goes over to the ''classical'' equation for stationary SBS in a planar pump field.[7]

Let us analyze the solutions of this system of equations in different situations. To start with, consider a homogeneous medium. Then $\delta=0$ and Eqs. (5) significantly simplify:

$$\frac{d(\mathscr{A}_2,\widetilde{\mathscr{A}}_2)}{dy}=-\eta\beta\int_{-\mathscr{L}}^{y}\frac{\exp[-h(y-\widetilde{y})][\mathscr{A}_2(\widetilde{y}),\widetilde{\mathscr{A}}_2(\widetilde{y})]}{1-i(y-\widetilde{y})\theta}d\widetilde{y}. \tag{6}$$

Hence, in the well known strong-damping limit (1) ($h\gg1$) the integrals of the kernels in Eqs. (6) converge rapidly and the unknown functions $\mathscr{A}_2$ and $\widetilde{\mathscr{A}}_2$ can be taken outside the integral at the point $\widetilde{y}=y$, thereby obtaining the corresponding gains $g=2\beta/h$ ($\eta=2$) and $\widetilde{g}=\beta/h$ ($\eta=1$).[1–3] These expressions, as can be easily seen from Eqs. (6), are valid for $g\ll h$. On the other hand, this condition obtains automatically since the very procedure of statistical averaging, with the help of which we obtained Eqs. (5), requires that the gain be small on the scale of one speckle spot, which in the given case means $g\ll1$. If these conditions are fulfilled, then $g/\widetilde{g}=2$ holds at any point of the medium and thus we have the wavefront conjugation effect.

In contrast to the simple situation just described, the opposite situation ($h\ll1$) is markedly more complicated and requires a more detailed analysis. Let us start out with $h=0$ and, for definiteness, take the uncorrelated component. Then from Eqs. (6) we have the equation ($\eta=1$, $\theta=1$)

$$\frac{d\widetilde{\mathscr{A}}_2}{dy}=-\beta\int_{-\mathscr{L}}^{y}\frac{\widetilde{\mathscr{A}}_2(\widetilde{y})d\widetilde{y}}{1-i(y-\widetilde{y})}. \tag{7}$$

It is interesting to note that this equation is very similar to the equation for ordinary time-independent SBS in a prescribed pump field in a homogeneous nondissipative medium. Such SBS is described, as is well known, by the equation of an ordinary harmonic oscillator, which follows directly from Eq. (7) in the absence of the imaginary term in the denominator of the kernel ($\theta=0$). Thus, the solution of Eq. (7) turns out to be not an exponential function, but a sum of trigonometric functions. One can convince oneself of this by substituting the trial solution

$$\widetilde{\mathscr{A}}_2(y)\propto\cos(\widetilde{b}y+\widetilde{d})+i\,\sin(\widetilde{b}y+\widetilde{d}), \tag{8}$$

in Eq. (7), where $\widetilde{b}$ and $\widetilde{d}$ are real constants. Substituting expression (8) on the right side of Eq. (7), we obtain an expression involving the sine-integral (si) and cosine-integral (ci) functions. In this expression the terms with $\mathrm{si}(u)$ and $\mathrm{ci}(u)$ [$u=\widetilde{b}(i+\mathscr{L}+y)$] can be neglected under the conditions

$$|\widetilde{b}|\lesssim1,\quad |\widetilde{b}(\mathscr{L}+y)|\gg1 \tag{9}$$

by virtue of the asymptotic expansions of these functions

$$\mathrm{si}(u),\ \mathrm{ci}(u)\sim\frac{(\sin u,\ \cos u)}{u}$$

(see Ref. 9), while in the remaining terms with $\mathrm{si}(i\widetilde{b})$ and $\mathrm{ci}(i\widetilde{b})$ given the first of conditions (9) it is sufficient to keep the first few terms of their series expansions.[9] As a result, by comparing with the left side of Eq. (7) we obtain the following equation for the period of the oscillations:

$$-|\widetilde{b}|=\beta\left(0.6+\ln|\widetilde{b}|-|\widetilde{b}|+\frac{1}{4}|\widetilde{b}|^2\right)e^{|\widetilde{b}|},\quad \widetilde{b}<0. \tag{10}$$

Hence, as can be easily seen, it follows that $|\widetilde{b}|\sim\beta$. A similar equation for the correlated component b ($\eta=2$, $\theta=2$) obtains analogously and has the same form as (10) with the substitution $\widetilde{b}\to b/2$, $\beta\to\beta/2$. Thus, the spatial frequencies b and $\widetilde{b}$ of the oscillations of the relative amplitudes $\mathscr{A}_2$ and $\widetilde{\mathscr{A}}_2$ are roughly identical, although the first is somewhat larger. Solutions (8)–(10) are in good agreement with numerical calculation of Eq. (7) by computer (Fig. 1). They break down, as can be seen from the second condition (9) only near the left boundary of the interaction region. The second of conditions (9) implies the smallness of variation of $\mathscr{A}_2$ and $\widetilde{\mathscr{A}}_2$ over the scale of one speckle spot, the need for which was noted earlier. Note that, as is clear from the limits of integration in $\int_{-\mathscr{L}}^{y}$, numerical calculation, both of this equation [Eq. (7)] and the more complicated Eqs. (6) and (5), simplifies substantially if we solve not the direct problem, as is customary, but the inverse problem, i.e., assign the Stokes wave not at the entrance to the layer—$\mathscr{A}_2(\mathscr{L})$, $\widetilde{\mathscr{A}}_2(\mathscr{L})$, but at the exit from the interaction region: $\mathscr{A}_2(-\mathscr{L})$, $\widetilde{\mathscr{A}}_2(-\mathscr{L})$.

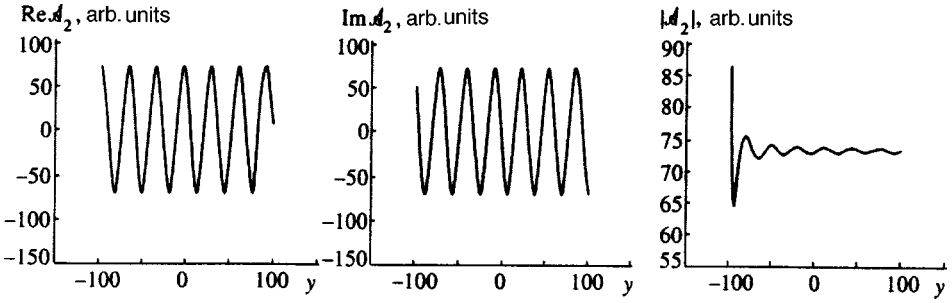


FIG. 1. Relative amplitude $\mathscr{A}_2$ correlated with the speckle-pump Stokes wave in inverse stimulated Brillouin scattering (the stationary problem) in a homogeneous layer without dissipation versus the dimensionless coordinate y (along the layer): $\beta=0.1$, $\eta=2$, $\theta=2$, $\delta=0$, $h=0$, $\mathscr{L}=100$, $\mathscr{A}_2(-\mathscr{L})=72+72i$.

Just this calculational scheme is used in this work. Thus, as follows from the solutions, for $h=0$ the intensities of the scattered waves in the medium do not undergo significant variations:

$$|\mathscr{A}_2(y)|\sim|\mathscr{A}_2(\mathscr{L})|, \quad |\tilde{\mathscr{A}}_2(y)|\sim|\tilde{\mathscr{A}}_2(\mathscr{L})|,$$

i.e., they remain at the level of the thermal fluctuations. In this regard, note that, continuing the above analogy with the problem of inverse SBS of a planar pump wave, an instability of the given stationary states may be expected, which in a speckle pump field is probably also absolute.[10,11]

The effect of damping of the longitudinal (sound) wave can substantially alter these solutions. To allow for it, it is necessary to return to the more general Eqs. (6). Thus, following, as before, the analogy of the processes considered here with ordinary SBS or with an ordinary harmonic oscillator, it may be expected that the oscillating solutions (8) break down, transforming into exponential solutions if the damping scale becomes less than the period of the oscillations (in the absence of damping), i.e., for $h>\tilde{b}$ or $h>b$ (for the correlated component). Under these conditions, as the calculations show,

$$-\mathrm{Re}\ g\simeq-\mathrm{Re}\ \tilde{g}\simeq\beta. \tag{11}$$

Note that this value corresponds to the growth rate for similar processes ($k_2\sim k_3$) defined in other models—of one-dimensional pump fluctuations or wide-frequency-band packets—in Refs. 12 and 13. The method for solving the general equations (5), proposed in Ref. 3, by simply taking the functions $\mathscr{A}_2(\tilde{y})$ and $\tilde{\mathscr{A}}_2(\tilde{y})$ outside the integral on the right side of the equation gives acceptable accuracy here. Indeed, taking this approach and setting

$$\mathscr{A}_2(y),\tilde{\mathscr{A}}_2(y)\propto\exp\left\{\int_{-\mathscr{L}}^{y}[g(\tilde{y}),\tilde{g}(\tilde{y})]d\tilde{y}\right\},$$

we obtain from Eqs. (6) approximately the same value for the real parts $\mathrm{Re}(g,\tilde{g})$, and likewise for the imaginary parts: $\mathrm{Im}(g,\tilde{g})\sim-\beta\ \ln(1/h)$, which agrees with the numerical solutions [the period of the oscillations in $\mathscr{A}_2(y)$ is not much less than in $\tilde{\mathscr{A}}_2(y)$]. For $h<b,\tilde{b}$ the gains of the Stokes waves are significantly decreased (Fig. 2).

Let us now analyze stimulated scattering of a pump with speckle structure in inhomogeneous media. Such scattering also turns out to be largely similar to ordinary stimulated scattering of a plane wave. For this purpose, we need to return to the original equations (5). As before, let us first consider the situation in which we neglect damping ($h=0$). Then for $\theta=0$ (a planar pump wave) Eqs. (5) are easily reduced to the equation of a parabolic cylinder, and its solutions are expressed in terms of the function of a parabolic cylinder.[7] The gradient of the amplitude levels of the Stokes wave corresponds to the well-known criterion[7,14]

$$\frac{|\mathscr{A}_2(-\mathscr{L})|}{|\mathscr{A}_2(\mathscr{L})|}=\exp(\pi p), \quad p=\frac{\eta\beta}{2\delta}\geqslant 1. \tag{12}$$

At the same time, the absolute value of the amplitude in the time-independent problem falls quite slowly: $|\mathscr{A}_2(y)|\propto 1/y$ (here $y\gg\delta^{-1/2}$ holds, and almost the entire variation of this function right up to its boundary value $|\mathscr{A}_2(\mathscr{L})|$ takes place in this region). Consequently, since the amplitude of the Stokes field must grow beyond the thermal noise level by a factor of $\exp(15)$ as a benchmark figure in order to observe stimulated scattering, a very large system is required. Therefore, in reality inverse stimulated scattering (from the thermal noise level) in nondissipative media is possible only as a time-dependent process. It has been considered in many studies,[7,14,15] in just such a formulation, where it was shown that the process of stimulated scattering begins to develop near the synchronization point as an absolute instability,

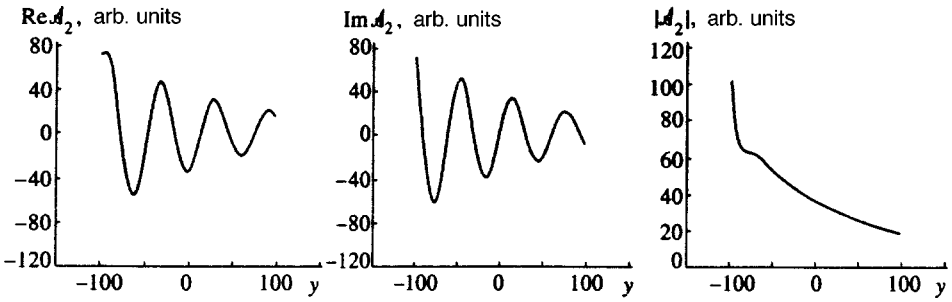


FIG. 2. The same as in Fig. 1, but for the case with dissipation: $\beta=0.04$, $\eta=2$, $\theta=2$, $\delta=0$, $h=0.03$, $\mathscr{L}=100$, $\mathscr{A}_2(-\mathscr{L})=72+72i$.

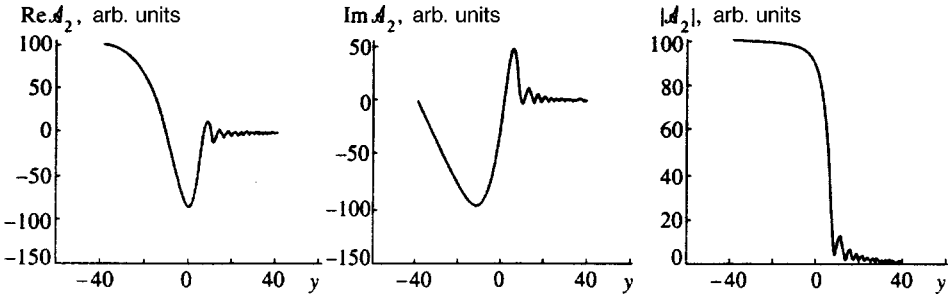


FIG. 3. The same as in Fig. 1, but for an inhomogeneous layer: $\beta=0.1$, $\eta=2$, $\theta=2$, $\delta=0.05$, $h=0$, $\mathscr{L}=40$, $\mathscr{A}_2(-\mathscr{L})=100$.

reaching the level (12) in a narrow region, after which the resulting Stokes wave and sound wave pulses broaden on both sides, encompassing ever larger regions of plasma. Here, if the length of the inhomogeneous layer is finite, the wave amplitudes cannot reach the described stationary state and when the pulse fronts reach the boundaries of the layer the process can again take on the character of an absolute instability, but with a different growth rate.

On this plane, the situation is essentially the same for a pump with speckle structure ($\theta\neq 0$). In this case it is not possible to find exact solutions of Eqs. (5); however, as numerical calculations show, they have the same form (Fig. 3). The total gain of the Stokes components is described by the same formula (12) and thus the total discrimination of the uncorrelated wave relative to the correlated wave after traversing the layer is the same as in the classical case,[1,2] i.e., it is equal to two. For the parameters of the cases shown in Fig. 3, the function $\mathscr{A}_2(y)$ converges to its limiting values (12) comparatively rapidly. At the same time, the length of the layer necessary for this grows abruptly as p increases. The values of $\mathscr{A}_2(y)$ and $\tilde{\mathscr{A}}_2(y)$ on the left side of the layer ($y<0$) not close to the synchronization point ($|y|\gg\delta^{-1/2}$) are given roughly by the formula in this region for the Stokes component for a planar pump wave[7]

$$\left.\begin{matrix}\mathscr{A}_2(y)\\ \tilde{\mathscr{A}}_2(y)\end{matrix}\right\}\propto\left\{\begin{matrix}\mathscr{A}_2(-\mathscr{L})\\ \tilde{\mathscr{A}}_2(-\mathscr{L})\end{matrix}\right\}\exp\{ip\ \ln[(2\delta)^{1/2}|y|]\},\qquad(13)$$

although the variable period of their oscillations, while maintaining the same constant amplitude, is somewhat larger than in Eqs. (13).[2)] On the right side of the layer the frequency of the oscillations grows rapidly, as in the case with planar pumping, and their amplitude decreases. The functional dependence of the falloff of $|\mathscr{A}_2(y)|$ and $|\tilde{\mathscr{A}}_2(y)|$ can be determined here as follows. Taking the correlated component ($\eta=2$, $\theta=2$) for definiteness, we write Eqs. (5) for it in the form (recall that $h=0$)

$$-\frac{1}{2\beta}\exp(i\delta y^2)\frac{d\mathscr{A}_2}{dy}=\int_{-\mathscr{L}}^{y_0}\frac{\mathscr{A}_2(\tilde{y})\exp(i\delta\tilde{y}^2)}{1-2i(y-\tilde{y})}d\tilde{y}+\int_{y_0}^{y}\frac{\mathscr{A}_2(\tilde{y})\exp(i\delta\tilde{y}^2)}{1-2i(y-\tilde{y})}d\tilde{y}\qquad(14)$$

where $y_0>0$ is such that $\delta y_0^2\gg 1$. It is probable that the main contribution to the first integral in Eq. (14) comes from the region near the synchronization point ($\delta\tilde{y}^2\sim 1$) since rapid oscillations of the exponential are happening on the left side of the layer, superposed on the slow variation of the amplitude (13). Therefore, setting $y>y_0$ it is possible to take the denominator of the integrand outside the integral, denoting the remaining part by c_1. On the other hand, for the indicated values of y, as can be seen from Eq. (14), it is possible to distinguish a slow and a fast part of the solution itself $\mathscr{A}_2(y)$ by representing it as

$$\mathscr{A}_2(y)=J(y)\exp(-i\delta y^2),\quad y>y_0.\qquad(15)$$

Substituting (15) in (14) and dropping small terms, we arrive at a Volterra integral equation of the second kind for the function $J(y)$. The solutions of such equations, as is well known,[16] can be found by the method of successive iterations. Applying it here, already in the zeroth iteration we obtain

$$J_0(y)\simeq\frac{pc_1}{2y^2},\quad c_1\equiv\int_{-\mathscr{L}}^{y_0}\mathscr{A}_2(\tilde{y})\exp(i\delta\tilde{y}^2)d\tilde{y}.\qquad(16)$$

Subsequent iterations have little effect, in contrast to the plane wave case, where the solution can also be simplified in a similar way by virtue of the more rapid falloff in this case of the function $J_0(y)$ with increasing y (Ref. 16). (For a planar pump wave, an additional phase factor arises in successive iterations that depends on y.) Following this discussion and in analogy with the plane wave case, where the exact solution is known, we define the quantity c_1 as

$$|c_1|\sim\delta^{-1/2}|\mathscr{A}_2(-\mathscr{L})|.$$

As a result we obtain

$$\mathscr{A}_2(y)\propto|\mathscr{A}_2(y)|\exp(-i\delta y^2),$$

$$|\mathscr{A}_2(y)|\sim\frac{p}{\eta\delta^{1/2}y^2}|\mathscr{A}_2(-\mathscr{L})|.\qquad(17)$$

As can easily be seen, the same dependences also obtain for the uncorrelated wave ($\eta=1$, $\theta=1$); however, its intensity, in accordance with the above remarks, falls to its limit much sooner. Computer calculations agree to within an order of magnitude with the values of $|\mathscr{A}_2(y)|$ and $|\tilde{\mathscr{A}}_2(y)|$ found from formulas (17), exceeding them severalfold. We also note that in the numerical solutions for a finite layer ($-\mathscr{L},\mathscr{L}$) the profiles of variation of these quantities clearly reveal oscillations about mean values corresponding to the values obtained from Eqs. (17). However, as the length of the layer is increased, these oscillations slowly smooth out and become less pronounced.

Although the falloff of $\mathscr{A}_2(y)$ according to formula (17) is more rapid than in the case of a plane wave (see above), if the Stokes wave grows above the thermal noise level, then the length of the layer needed to observe such (stationary) stimulated scattering is still large. Time-dependent stimulated scattering in a speckle pump field probably also develops in a way similar to the above picture for a planar pump field. Thus, an absolute instability should develop in a small neighborhood of the synchronization point also in this case. However, for an absolute instability an inverted wave is not formed,[3)] and therefore the previously indicated growth of the intensity levels of the Stokes waves with discrimination 2 (in amplitude) is not realized at this stage and apparently can be achieved only by passing to the stationary regime, which again requires a large system.

We also note here that in the case of a planar pump wave, as in formula (17), $\mathscr{A}_2(y) \propto \exp(-i\delta y^2)$. Therefore the fact, noted in Ref. 3, that the result (12) can be obtained in a very simple way is quite interesting. Specifically, if in a purely formal way we take the function $\mathscr{A}_2(y)$ outside the integral in Eqs. (5) ($\theta=0$) (although, generally speaking, it varies, as can be plainly seen, no more weakly than the kernel of this equation), we thereby introduce for $\mathscr{A}_2(y)$ in Eqs. (5) the growth rate $g(y)$ and integrate it in y from $-\infty$ to $+\infty$ ($\pm\mathscr{L}\to\pm\infty$), thereby determining the total gain for the Stokes wave in the layer. The resulting double integral is calculated immediately by transforming to different integration variables: $y,\tilde{y}\to y,\xi$ ($\xi\equiv y-\tilde{y}$). The main conclusions in Ref. 3 were in fact based on the described correspondence with (12). No less interesting is the fact that all of the numerical solutions for an inhomogeneous medium presented here confirm the final result for the total gain of the Stokes waves in the layer obtained by this approach.[3]

Below we describe the effect of damping of the longitudinal wave ($h\neq 0$) on the processes of forced scattering in inhomogeneous media discussed above. For a plane wave ($\theta=0$) Eqs. (5), like for $h=0$, are easily reduced to the equation of a parabolic cylinder, which when we transform from y to independent variable (complex) $\zeta\equiv y-ih/2\delta$ takes the same form as in the case $h=0$. Thus, here for $\mathscr{A}_2$ the same solutions[7] with the corresponding substitution are valid.

For a pump field with speckle structure in this situation, as in the above case of inverse scattering in the absence of dissipation ($h=0$), there are solutions that are noticeably similar to the corresponding solutions for a plane wave. First it must be said that the amplitudes $\mathscr{A}_2(y)$ and $\tilde{\mathscr{A}}_2(y)$ on the left side of the layer not far from the synchronization point have roughly the dependence of Eqs. (13) in which we must make the indicated substitution for y, i.e., they take the form

$$\left.\begin{matrix}\mathscr{A}_2(y)\\ \tilde{\mathscr{A}}_2(y)\end{matrix}\right\} \propto \left\{\begin{matrix}\mathscr{A}_2(-\mathscr{L})\\ \tilde{\mathscr{A}}_2(-\mathscr{L})\end{matrix}\right\} \exp\left\{p\left[i\ \ln[(2\delta)^{1/2}|y|]+\frac{h}{2\delta y}\right]\right\} \tag{18}$$

($h/2\delta|y|\ll 1$), but now not only are the periods of their infrequent oscillations somewhat larger than in formulas (18), but the amplitudes of the latter also fall off from the left boundary of the layer somewhat more slowly. On the right side of the layer the behavior of the solutions of Eq. (5) is much more varied and depends on the ratio of the parameters β and h. For $h<(\pi/2)\beta$, as will be shown below, to find the solutions in the region $y>0$ it is possible as before to represent them in the form given by the first of formulas (17) ($\delta y^2\gg 1$). Then, from an equation similar to Eq. (14) but allowing for the exponential factor of the kernel in Eqs. (5) associated with damping, we similarly find

$$\mathscr{A}_2(y)\propto|\mathscr{A}_2(y)|\exp(-i\delta y^2),$$

$$|\mathscr{A}_2(y)|\sim\frac{p}{\eta\delta^{1/2}y^2}\exp(-hy)|\mathscr{A}_2(-\mathscr{L})|,\quad y\gg\frac{h}{2\delta}. \tag{19}$$

In contrast to the nondissipative case, here the intensity of the Stokes wave decreases much more rapidly through the layer. The same behavior is obtained for the amplitude of the uncorrelated wave [as was the case for formulas (17)]. The results of numerical calculations in this case roughly also agree with formulas (19), just as the corresponding results agree with formulas (17).

For stronger damping of the ion acoustic wave

$$h>\frac{\pi}{2}\beta \tag{20}$$

the amplitudes of the Stokes components oscillate on the right side of the layer much more slowly than indicated by Eqs. (19). As the numerical calculations show, in the given case the magnitudes of the amplitudes $|\mathscr{A}_2|$ and $|\tilde{\mathscr{A}}_2|$ fall off on the right side of the layer with growth (damping)rates gradually decreasing in magnitude with distance from the synchronization point ($y=0$) due to the influence of the oscillation factors $\exp(\pm i\delta y^2)$ in Eqs. (5). For $|\tilde{\mathscr{A}}_2|$ this influence up to the point

$$y_*\approx\frac{1}{2\delta}, \tag{21}$$

and for $|\mathscr{A}_2|$, up to the point $2y_*$, turns out to be insubstantial and the corresponding rates on these intervals are equal to approximately the same values as in a homogeneous plasma:

$$\mathrm{Re}\ \tilde{g}\approx-\frac{\pi}{2}\beta,\quad \mathrm{Re}\ g\approx-\frac{\pi}{2}\beta. \tag{22}$$

The value of expression (21) arises from the condition that the scale of the oscillations of the given exponential factors in Eqs. (5) ($\theta=1$) at the given point, equal to $\pi/2\delta y_*$, should significantly exceed the longitudinal correlation length of the field ($\Delta y=1$), over which, as is not hard to see, the growth rate (11) is in fact formed. At the same time, since it is easy to show that $\theta=2$ for the correlated component in Eqs. (5), the value of y up to which it may be assumed that the growth rate for $|\mathscr{A}_2|$ is given by formula (22) is two times larger than the value given by formula (21). Similar simple arguments, however, do not apply to the left side of the layer, where, as was pointed out, the amplification of the Stokes waves is insignificant.[4)] The behavior of the functions $\mathscr{A}_2$ and $\tilde{\mathscr{A}}_2$ themselves on the right side of the layer is somewhat more complicated. Transforming to the

functions $\exp[i\alpha_2(y)] \equiv \mathscr{A}_2/|\mathscr{A}_2|$ and $\exp[i\tilde{\alpha}_2(y)] \equiv \tilde{\mathscr{A}}_2/|\tilde{\mathscr{A}}_2|$, we show that their oscillations, as was already noted, occur with comparatively large (variable) periods. The frequencies of the oscillations are subtended between the value $[\mathrm{Im}(g,\tilde{g})]_{\mathrm{hom}} \sim \beta \ln(1/h)$ (see above for the homogeneous medium) on the one side and values following from formulas (18) on the other. Starting from lengths y_0 where the magnitudes $|\mathscr{A}_2|$ and $|\tilde{\mathscr{A}}_2|$ become comparable with those satisfying criterion (12), the periods of the oscillations are in quite good agreement with the formulas (18). This is because at such lengths, as for a planar pump field, the amplitudes of the Stokes waves are described by the same dependence as on the left side of the layer divided by the exponential in Eq. (12). At lengths less than y_0 the periods vary weakly and for not too small β they are equal to several of the periods corresponding to $[\mathrm{Im}(g,\tilde{g})]_{\mathrm{hom}}$. For such lengths a phase lapse occurs in the oscillations in the region of y_* (21) (the inhomogeneity of the medium tends to increase their period in comparison with the period in the homogeneous medium, bringing it closer to ''its own'' (18). As can be easily seen by comparing $\mathrm{Re}(g,\tilde{g})$ (22) and $[\mathrm{Im}(g,\tilde{g})]_{hom}$, several oscillations in the wave amplitudes can occur up to the point y_0 for real values of the parameter $p \gg 1$ (12).

It is clear from the above remarks that these solutions, generally speaking, differ substantially from the preceding (19) and cannot be obtained by the same technique as the latter. Indeed, separating the two integrals in the applicable equation, similar to Eq. (14), with the dominant contribution coming from the first given conditions (20) and (22), is invalid since in this case the falloff of the exponential $\exp[-h(y-\tilde{y})]$ (5) outpaces the growth of $|\mathscr{A}_2(y)|$ in the y direction, thereby suppressing the main contribution to the integral from the region near the synchronization point. On the other hand, the scales of the oscillations of the functions $\exp[i\alpha_2(y)]$ and $\exp[i\tilde{\alpha}_2(y)]$ (see above) are formed in the given case precisely as a result of the influence of the attenuation (h). Therefore its spatial scale should be less than the scales of these oscillations, which in fact is what condition (20) says. The total gain due to the whole layer of both Stokes components for both this condition and the reverse inequality corresponds to criterion (12).[5)]

Thus, we have analyzed inverse stimulated scattering of a pump field with speckle structure for different attenuations of the sound (ion acoustic) wave on which the process develops. As can be seen, a substantial difference between the discrimination mechanism (noted in Ref. 3 and discussed here) connected with the difference in the gain, due to the inhomogeneities of the medium, of the correlated and uncorrelated (with the pump) Stokes waves, and the classical mechanism[1,2] arises because the latter gives the direct discrimination of the growth rate $\tilde{g}$ ($g/\tilde{g}=2$ at all points in the medium), while in the case discussed here ($h \ll 1$) it is manifested only on isolated segments and is collected together in the total discrimination integrated over the entire layer. However, for inverse stimulated scattering these segments are distributed less favorably so that at the final stage of amplification of the correlated component discrimination of the uncorrelated component is essentially absent. This makes the conditions necessary for wavefront conjugation more severe than in the case $h \gg 1$.

The point is that for wavefront conjugation, generally speaking, the discrimination of the gain of the uncorrelated waves revealed above is simply inadequate. It is also necessary that the specklon distortions (extinction of the correlated component) invariably arising on the scale of each spot of the pump speckle structure as a consequence of amplification at the center of the spot and at its edges (the so-called serpentine distortions, not taken into account in the first-order treatment of the specklon) and accumulating at the moment the inverted wave leaves the layer remain nevertheless substantially smaller than the intensity level of the inverted wave. In the well known situation treated in Refs. 1 and 2 this is achieved for $g \ll 1$ (small gain of the specklon on the scale of a spot) thanks to the increasing rate of amplification of the inverted component in comparison with the uncorrelated distortions ($g/\tilde{g}=2$). In the given case, as was described above, at a significant distance on either side of the synchronization point ($y=0$) no such discrimination exists. Even under more favorable conditions of greater damping of the sound wave (20) the quantities $\mathrm{Re}\, g$ and $\mathrm{Re}\, \tilde{g}$ up to lengths $y \sim y_*$ (21) differ only weakly [see Eqs. (22)] so that for them it may be assumed that specklon distortions from different speckle spots (inhomogeneities) simply add in intensity (by virtue of the randomness of the process). Thus, invoking the well known estimate[1] for the extinction coefficient[6)]

$$\mathscr{R} \approx |\Delta\mathscr{A}_2(y)/\mathscr{A}_2(y)|^2,$$

where $\Delta\mathscr{A}_2(y)$ is the variation of the complex amplitude $\mathscr{A}_2$ over the correlation length (speckle spot), we obtain the following estimate for the relative level of distortions on the interval $(0,y_*)$:

$$Q = \mathscr{R} y_* = \left(\frac{\pi}{2}\beta\right)^2 y_* = \pi p \frac{\pi\beta}{8}, \quad \eta = 2. \tag{23}$$

For simplicity, we have taken into account here in $\mathscr{R}$ only variations of $|\mathscr{A}_2|$. From the numerical calculations it is possible to obtain a more exact value for the total level of distortions Q_Σ with allowance for the contribution of the region to the left of the synchronization point $y=0$, and also phase variations in $\mathscr{A}_2$ (including the above-mentioned phase lapse in the oscillations at $y \sim y_*$). Specifically, calculating $\mathscr{R}(y)$ directly for each correlation length and summing these values up to y_*, we find that Q_Σ exceeds the estimate (23) by at least a factor of two. This means that since the factor πp on the rightmost side of Eq. (23) is the argument of the exponential in criterion (12) (which in order to be able to observe the inverted component must be equal approximately to 15, and the distortions themselves should be small, $Q_\Sigma \ll 1$), another necessary condition here is

$$\beta \leqslant 0.02. \tag{24}$$

Thus it follows from conditions (12) that $\delta \leqslant 0.005$. Let us now determine the length of the layer that is required for

amplification of a specklon. Even if we assume that the growth rate (22) is conserved for all $y>0$, it is easy to obtain while satisfying criterion (12) ($\eta=2$) that

$$y_{\min}=2/\delta.$$

The decrease in $|\mathrm{Re}\ g|$ as a function of y increases this quantity by roughly a factor of two, so that for $\delta\leqslant 0.005$ we have

$$\mathscr{L}_{\min}\geqslant 10^3. \tag{25}$$

Thus, at such lengths we have $Q_\Sigma\leqslant 0.2$. In the case of strong damping of the sound wave, it is easy to see that such a relative level of the distortions (in intensity) of the inverted wave as it leaves the layer, which in this case corresponds directly to the relative noise level (distortion) of the wave amplitude on the scale of a speckle spot,[1,2] is reached even for a length of the layer $2\mathscr{L}_{\min}\geqslant 70$.[7)] Note also that if we have in mind stimulated scattering in speckle beams (e.g., focused in the medium), where the pump intensity varies smoothly over its cross section, then the discrimination of the growth rate of the uncorrelated components in this situation is decreased since the specklon must also keep its structure on the periphery of the beam cross section, where the gain is weak, while the uncorrelated waves are concentrated into its central part, where the gain is maximum. As a result, the growth rates Re g and Re $\tilde{g}$ are similar in value at still greater distances than in the case under consideration of a pump with constant mean intensity over the cross section (3), and therefore the length $\mathscr{L}_{\min}$ is also increased in comparison with estimate (25). In this case, the arguments of the exponent in criterion (12) for the correlated and uncorrelated components probably differs not by a factor of two, as previously, but only by a factor of 1.4, in analogy with the decrease in the degree of discrimination of the uninverted wave in a focused beam in homogeneous media.[1,2]

Thus, as could be expected, the situation with strong damping of the sound wave ($h\gg 1$) (1) is more favorable for wavefront conjugation than its opposite. However, since under ordinary gas-dynamic conditions $\nu\propto k_3^2$, it becomes increasingly difficult to satisfy this condition at longer wavelengths (e.g., even for a CO_2-laser). From this point of view, a weakly collisional plasma is more promising since an ion acoustic wave in it attenuates mainly as described by the Landau theory (collisionlessly) and $\nu\propto k_3$, i.e., it falls off more slowly with decreasing k_3. Larger attenuation of the sound wave in a plasma is important for wavefront conjugation because[18,19] it gives shorter setup times for the stationary regime in which stimulated scattering is strongest. At the same time, in a weakly collisional plasma there is no collisional heating which can interfere with SBS and wavefront conjugation.[20,21] On the other hand, nonequilibrium conditions are also possible in such a plasma, with weaker damping of waves. Note that for a Maxwellian distribution function electron Landau damping of ion sound corresponds to the condition

$$h=\nu l=2\sqrt{\frac{\pi}{8}}(2\pi)^2\sqrt{\frac{m}{M}}\left(\frac{\rho}{\lambda_1}\right)^2>1$$

(m and M are respectively the mass of the electron and the ion, $\lambda_1=2\lambda_3$, and the scale of the speckle inhomogeneities $\rho\gg\lambda_1$).

[1)] For simplicity we do not take account in these equations of secondary nonlinearities such as generation of harmonics of ion sound,[4–6] assuming that the pump wave is not too strong, the more so since in the given case this process is noticeably damped as it proceeds incoherently.

[2)] Note that the distances between the maxima and minima in the dependence $\mathscr{A}_2(y)$ in the numerical solutions for a planar pump wave are well described by expressions (13); however, their positions may not necessarily coincide with those determined in Ref. 7, where solutions were found for an unbounded layer ($\mathscr{L}\to\infty$) and the boundary condition $\mathscr{A}_2(y\to\infty)=\exp\{ip\ln[(2\delta)^{1/2}y]\}$ [since Eqs. (5) are linear, any solution multiplied by an arbitrary constant factor is also a solution corresponding to the boundary condition of the first solution multiplied by the same coefficient (including complex coefficients); this fact is taken into account in the form of expressions (13)]. The number of trains of oscillations grows continuously as the length of the layer is increased.

[3)] We are grateful to G. A. Pasmanik for bringing this to our attention.

[4)] On the other hand, this picture can be explained here by employing the same procedure[3] of taking the functions $\mathscr{A}_2(\tilde{y})$ and $\tilde{\mathscr{A}}_2(\tilde{y})$ outside the integral in Eqs. (5) and introducing a growth rate calculated numerically for $h=0$, i.e., (for example, for $\tilde{\mathscr{A}}_2$)

$$\tilde{g}(y)=-\beta\int_0^{y+\mathscr{L}}\frac{\exp(i\delta\xi^2-i2\delta\xi y)}{1-i\xi}d\xi,\quad \xi=y-\tilde{y}.$$

The graph of the function Re $\tilde{g}(y)$ ($\mathscr{L}\to\infty$) provided in Ref. 3 shows that the interval where this quantity is not small is concentrated mainly in the region $y\gtrsim 0$. This is because the given integrand expression oscillates less in the integration region for $y>0$ [this region contains, in contrast to the region of values $y<0$, a point ξ for which $d(\delta\xi^2-2\delta\xi y)/d\xi=0$]. It may be expected that taking account in $\tilde{g}(y)$ of the sufficient damping assured by satisfaction of condition (20) leads to significantly better quantitative agreement between the results obtains by this method and those presented here.

[5)] Note that this criterion is also satisfied for the accompanying stimulated scattering, including, as we have verified, pump fields with speckle structure. Some generalizations of it for two-dimensional systems were obtained in Ref. 17.

[6)] Note also that an exact calculation of this coefficient in the given case is hardly possible: here it is a function of y and to find it, it is necessary to know the solution $\mathscr{A}_2(y)$ over the entire layer. Even in the simplest situation $h\gg 1$ its calculation is quite involved,[1] but the result turns out to be in good agreement with the estimate given here.

[7)] Note that at a pump frequency moderately exceeding the electron plasma frequency ($\omega_1\sim\omega_{pe}$) this length turns out to be much smaller than the inhomogeneity scale of the plasma $\mathscr{L}_N$ (on this segment, at least) that is necessary for the required level of scattering.[18] However, the same also applies to the case (25).

[1] B. Ya. Zel'dovich, N. F. Pilipetskiĭ, and V. V. Shkunov, *Principles of Phase Conjugation* (Springer-Verlag, Berlin, 1985).

[2] V. I. Bespalov and G. A. Pasmanik, *Nonlinear Optics and Adaptive Laser Systems* Nova Science, Commack, NY (1994).

[3] V. V. Kurin, Zh. Éksp. Teor. Fiz. **97**, 1467 (1990) [Sov. Phys. JETP **70**, 828 (1990)].

[4] V. P. Silin and V. T. Tikhonchuk, Zh. Éksp. Teor. Fiz. **83**, 1332 (1982) [Sov. Phys. JETP **56**, 765 (1982)].

[5] V. V. Kurin, Fiz. Plazmy **10**, 860 (1984) [Sov. J. Plasma Phys. **10**, 498 (1984)].

[6] J. A. Heikkinen, S. J. Karttunen, and R. R. E. Salomaa, Phys. Fluids **27**, 707 (1984).

[7] A. D. Piliya, Zh. Éksp. Teor. Fiz. **64**, 1237 (1973) [Sov. Phys. JETP **37**, 629 (1973)].

[8] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products* (Academic Press, New York, 1980).

[9] *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (Dover, New York, 1965).

[10] N. M. Kroll, J. Appl. Phys. **36**, 34 (1965).
[11] L. M. Gorbunov, Zh. Éksp. Teor. Fiz. **67**, 1386 (1974) [Sov. Phys. JETP **40**, 689 (1974)].
[12] E. J. Valeo and C. R. Oberman, Phys. Rev. Lett. **30**, 1035 (1973).
[13] V. N. Tsytovich, *Nonlinear Effects in Plasma* [in Russian] (Nauka, Moscow, 1967).
[14] M. N. Rosenbluth, Phys. Rev. Lett. **29**, 565 (1972).
[15] A. Bers, in *Handbook of Plasma Physics* edited by A. A. Galeev and R. Sudan, North-Holland, Amsterdam, (1984).
[16] M. L. Krasnov, A. I. Kiselev, G. I. Makarenko, *Integral Equations* [in Russian] (Nauka, Moscow, 1976).
[17] I. V. Khazanov, A. G. Litvak, A. M. Sergeev *et al.*, Phys. Fluids B **5**, 4347 (1993).
[18] A. A. Andreev and A. N. Shatsev, Pis'ma Zh. Tekh. Fiz. **10**, 883 (1984) [Sov. Phys. Tech. Phys. **10**, 371 (1984)].
[19] I. M. Bel'dyugin, M. G. Galushkin, and V. N. Soshnikov, Kvantovaya Elektron. **12**, 1387 (1985) [Sov. J. Quantum Electron. **15**, 919 (1985)].
[20] S. F.. Grigor'ev, O. P. Zaskal'ko, and V. V. Kuz'min, Zh. Éksp. Teor. Fiz. **92**, 1246 (1987) [Sov. Phys. JETP **65**, 697 (1987)].
[21] A. A. Andreev, A. A. Betin, O. V. Mitropol'skiĭ, and A. N. Shatsev, Zh. Éksp. Teor. Fiz. **92**, 1636 (1987) [Sov. Phys. JETP **65**, 918 (1987)].

Translated by Paul F. Schippnick