Journal of Experimental and Theoretical Physics, Vol. 99, No. 5, 2004, pp. 1005–1017.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 5, 2004, pp. 1153–1166.

FLUIDS

Diffusion and Clustering of Sedimenting Tracers in Random Hydrodynamic Flows

V. I. Klyatskin

Oboukhov Institute of Atmospheric Physics, Russian Academy of Sciences, Moscow, 109017 Russia
Il'ichev Pacific Institute of Oceanology, Far East Division, Russian Academy of Sciences, Vladivostok, 690041 Russia
e-mail: klyatskin@yandex.ru

Received May 5, 2004

Abstract—The analysis presented in [1, 2] is extended to sedimenting low-inertia tracers advected by random divergence-free hydrodynamic flows. The key feature of the process is the clustering of the tracers due to the divergence of tracer-velocity field. This phenomenon has probability one; i.e., it takes place in almost every realization of the process. Both spatial diffusivity and diffusivity in the density space (responsible for clustering) are calculated. The low inertia of the tracers does not affect the spatial diffusivity. The indispensable use of a finite velocity correlation time leads to an anisotropic spatial diffusivity. The calculations performed in the study are based on a diffusion approximation. © *2004 MAIK "Nauka/Interperiodica".*

1. CURRENT STATUS OF THE PROBLEM AND BASIC EQUATIONS

The inertial tracer density field advected by a random flow satisfies the continuity equation

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial \mathbf{r}}\mathbf{V}(\mathbf{r}, t)\right)\rho(\mathbf{r}, t) = 0, \quad \rho(\mathbf{r}, t) = \rho_0(\mathbf{r}), \quad (1)$$

which can be rewritten as

$$\left(\frac{\partial}{\partial t} + \mathbf{V}(\mathbf{r}, t)\frac{\partial}{\partial \mathbf{r}}\right)\rho(\mathbf{r}, t) + \frac{\partial \mathbf{V}(\mathbf{r}, t)}{\partial \mathbf{r}}\rho(\mathbf{r}, t) = 0. \quad (2)$$

In the general case, the Eulerian tracer velocity $\mathbf{V}(\mathbf{r}, t)$ is different from the Eulerian flow velocity $\mathbf{u}(\mathbf{r}, t)$.

Effects due to molecular diffusion can be neglected at an early stage of the process. At a later stage, these effects must be taken into account:

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial \mathbf{r}}\mathbf{V}(\mathbf{r}, t)\right)\rho(\mathbf{r}, t) = \mu\Delta\rho(\mathbf{r}, t), \quad \rho(\mathbf{r}, 0) = \rho_0(\mathbf{r}), \quad (3)$$

where μ is molecular diffusivity. The total mass of tracers is conserved in the course of evolution:

$$M = M(t) = \int d\mathbf{r}\rho(\mathbf{r}, t) = \int d\mathbf{r}\rho_0(\mathbf{r}) = \text{const}.$$

The Eulerian velocity $\mathbf{V}(\mathbf{r}, t)$ of low-inertia particles in a flow field $\mathbf{u}(\mathbf{r}, t)$ can be described by the quasilinear partial differential equation (e.g., see [3])

$$\left(\frac{\partial}{\partial t} + \mathbf{V}(\mathbf{r}, t)\frac{\partial}{\partial \mathbf{r}}\right)\mathbf{V}(\mathbf{r}, t) = -\lambda[\mathbf{V}(\mathbf{r}, t) - \mathbf{u}(\mathbf{r}, t)] + \mathbf{g}\left(1 - \frac{\rho_0}{\rho_p}\right) \quad (4)$$

interpreted as a phenomenological model, where λ is the drag coefficient, $\mathbf{g}$ is the gravitational acceleration, ρ_p is the tracer particle density, and ρ_0 is the fluid density. In the general case, Eq. (4) may have multiple or discontinuous solutions. However, its solution is unique over a reasonable time interval in the asymptotic case of $\lambda \longrightarrow \infty$ (low-inertia particles), which is considered in the present study.

The linear drag force $\mathbf{F}(\mathbf{r}, t) = \lambda\mathbf{V}(\mathbf{r}, t)$ on the right-hand side of (4) is given by the Stokes law for a slowly moving particle treated approximately as a sphere: $\lambda = 6\pi a\eta/m_p$, where a is the tracer radius, η is the coefficient of dynamic viscosity, and m_p is the particle mass (see [4, 5]).

The velocity $\mathbf{v}$ of tracer sedimentation or rise (usually in the vertical direction) is determined by the balance between buoyancy and viscous forces:

$$\frac{\mathbf{g}}{\lambda}\left(1 - \frac{\rho_0}{\rho_p}\right) = \mathbf{v}.$$

By introducing the representation

$$\mathbf{V}(\mathbf{r}, t) = \mathbf{v} + \mathbf{v}(\mathbf{r}, t), \quad (5)$$

where $\mathbf{v}(\mathbf{r}, t)$ is the tracer velocity fluctuation about $\mathbf{v}$,

Eqs. (2) and (4) are rewritten as

$$\left(\frac{\partial}{\partial t} + [\mathbf{v} + \mathbf{v}(\mathbf{r}, t)]\frac{\partial}{\partial \mathbf{r}}\right)\rho(\mathbf{r}, t) = -\frac{\partial \mathbf{v}(\mathbf{r}, t)}{\partial \mathbf{r}}\rho(\mathbf{r}, t),$$
$$\rho(\mathbf{r}, 0) = \rho_0(\mathbf{r}), \tag{6}$$
$$\left(\frac{\partial}{\partial t} + [\mathbf{v} + \mathbf{v}(\mathbf{r}, t)]\frac{\partial}{\partial \mathbf{r}}\right)\mathbf{v}(\mathbf{r}, t) = -\lambda[\mathbf{v}(\mathbf{r}, t) - \mathbf{u}(\mathbf{r}, t)].$$

In the general case, a divergent Gaussian stationary, homogeneous, and isotropic random field velocity field $\mathbf{u}(\mathbf{r}, t)$ with zero mean ($\langle \mathbf{u}(\mathbf{r}, t)\rangle = 0$) is characterized by the correlation and spectral tensors

$$B_{ij}^{(\mathbf{u})}(\mathbf{r} - \mathbf{r}', t - t') = \langle u_i(\mathbf{r}, t)u_j(\mathbf{r}', t')\rangle$$
$$= \int d\mathbf{k} E_{ij}(\mathbf{k}, t - t')e^{i\mathbf{k}(\mathbf{r} - \mathbf{r}')},$$
$$E_{ij}(\mathbf{k}, t) = \frac{1}{(2\pi)^d}\int d\mathbf{r} B_{ij}^{(\mathbf{u})}(\mathbf{r}, t)e^{-i\mathbf{k}\cdot\mathbf{r}}, \tag{7}$$
$$E_{ij}(\mathbf{k}, t) = E_{ij}^{\mathrm{sol}}(\mathbf{k}, t) + E_{ij}^{\mathrm{pot}}(\mathbf{k}, t),$$

where d is the space dimension and the spectral tensor components have the form

$$E_{ij}^{\mathrm{sol}}(\mathbf{k}, t) = E^{\mathrm{sol}}(\mathbf{k}, t)\left(\delta_{ij} - \frac{k_i k_j}{k^2}\right),$$
$$E_{ij}^{\mathrm{pot}}(\mathbf{k}, t) = E^{\mathrm{pot}}(k, t)\frac{k_i k_j}{k^2} \tag{8}$$

with solenoidal and potential components $E^{\mathrm{sol}}(k, t)$ and $E^{\mathrm{pot}}(k, t)$.

The correlation time of the field $\mathbf{u}(\mathbf{r}, t)$ is defined as

$$\tau_0 = \frac{1}{\sigma_{\mathbf{u}}^2}\int_0^\infty d\tau B_{ii}^{(\mathbf{u})}(0, \tau);$$

the velocity variance, as

$$\sigma_{\mathbf{u}}^2 = B_{ii}^{(\mathbf{u})}(0, 0) = \langle \mathbf{u}^2(\mathbf{r}, t)\rangle$$
$$= \int d\mathbf{k}[(d - 1)E^{\mathrm{s}}(k, \tau) + E^{\mathrm{p}}(k, \tau)].$$

In the analysis presented below, $\sigma_{\mathbf{u}}^2 = \langle \mathbf{u}^2(\mathbf{r}, t)\rangle$ is treated as a small parameter.

1.1. Tracer Diffusion: Lagrangian Approach

The first-order partial differential Eulerian equations (6) are equivalent to the following ordinary differential equations in a Lagrangian approach:

$$\frac{d}{dt}\mathbf{r}(t) = \mathbf{v} + \mathbf{v}(\mathbf{r}(t), t), \quad \mathbf{r}(0) = \mathbf{r}_0,$$
$$\frac{d}{dt}\mathbf{v}(t) = -\lambda[\mathbf{v}(t) - \mathbf{u}(\mathbf{r}(t), t)], \quad \mathbf{v}(0) = \mathbf{v}_0(\mathbf{r}_0). \tag{9}$$

Note that these equations describe the Newtonian dynamics of a particle under the action of the random force $\mathbf{f}(t) = \lambda\mathbf{u}(\mathbf{r}(t), t)$ exerted by the hydrodynamic flow and the Stokes drag force $\mathbf{F}(t) = -\lambda\mathbf{v}(\mathbf{r}(t), t)$.

The dependence of the solution to (9) on $\mathbf{r}_0$ is denoted here by a vertical bar:

$$\mathbf{r}(t) = \mathbf{r}(t|\mathbf{r}_0), \quad \mathbf{v}(t) = \mathbf{v}(t|\mathbf{r}_0).$$

Accordingly, the Eulerian tracers density $\rho(\mathbf{r}, t)$ is expressed as [6–8]

$$\rho(\mathbf{r}, t) = \int d\mathbf{r}_0 \rho_0(\mathbf{r}_0)\delta(\mathbf{r}(t|\mathbf{r}_0) - \mathbf{r}). \tag{10}$$

The delta function on the right-hand side of (10) is referred to as the indicator function

$$\varphi(\mathbf{r}, t) = \delta(\mathbf{r}(t|\mathbf{r}_0) - \mathbf{r}), \tag{11}$$

and its value averaged over the ensemble of realizations of the random field $\mathbf{u}(\mathbf{r}, t)$ is the probability density function (PDF) of particle location [6–8]

$$P(\mathbf{r}, t) = \langle \varphi(\mathbf{r}, t)\rangle_{\mathbf{u}} = \langle \delta(\mathbf{r}(t|\mathbf{r}_0) - \mathbf{r})\rangle_{\mathbf{u}}. \tag{12}$$

As $\lambda \longrightarrow \infty$ (in the limit of inertialess tracers), it holds that

$$\mathbf{v}(\mathbf{r}, t) \approx \mathbf{u}(\mathbf{r}, t), \tag{13}$$

and Eqs. (6) and (9) reduce to

$$\frac{d}{dt}\mathbf{r}(t) = \mathbf{v} + \mathbf{u}(\mathbf{r}(t), t), \quad \mathbf{r}(0) = \mathbf{r}_0,$$
$$\left(\frac{\partial}{\partial t} + \mathbf{v}\frac{\partial}{\partial \mathbf{r}} + \frac{\partial}{\partial \mathbf{r}}\mathbf{u}(\mathbf{r}, t)\right)\rho(\mathbf{r}, t) = 0, \tag{14}$$
$$\rho(\mathbf{r}, 0) = \rho_0(\mathbf{r}).$$

In other words, the trajectories of inertialess particles in a hydrodynamic flow can be found by solving a kinematics problem.

First of all, one has to elucidate the statistical meaning of (13) and its scope. The applicability of this approximation in a statistical model depends on the order of limit operations (see [1, 2]).

Spatial diffusion of inertialess sedimenting particles in the absence of clustering in a divergence-free velocity field $\mathbf{u}(\mathbf{r}, t)$ ($\mathrm{div}\,\mathbf{u}(\mathbf{r}, t) = 0$) was analyzed in [9, 10], where spatial diffusivity was shown to be anisotropic with respect to the sedimentation direction defined by the vector $\mathbf{v}$. The anisotropy is due to finite correlation time τ_0 of the flow velocity field. (Spatial diffusion would be isotropic if the field $\mathbf{u}(\mathbf{r}, t)$ were delta-corre-

lated in time.) In those studies, a finite τ_0 was used in the framework of diffusion approximation.

In [1], Eqs. (9) were used to analyze diffusion and clustering of low-inertia particles in a divergence-free velocity field $\mathbf{u}(\mathbf{r}, t)$ in the absence of sedimentation ($\mathbf{v} = 0$). It was shown that fields $\mathbf{u}(\mathbf{r}, t)$ that are delta-correlated in time cannot be used in statistical models of particle transport in the inertialess limit; i.e., the limit operations $\lambda \longrightarrow \infty$ and $\tau_0 \longrightarrow 0$ are not interchangeable. However, these limits commute when statistical modeling is restricted to spatial diffusion. It was shown in [11] that this is also true for sedimenting particles ($\mathbf{v} \neq 0$) in a divergence-free field $\mathbf{u}(\mathbf{r}, t)$, in which case anisotropy of spatial diffusivity is also due solely to a finite correlation time τ_0. Moreover, it is obvious that the inertialess limit is equivalent to the following conditions for the parameter λ:

$$\lambda\tau_0 \gg 1, \quad \lambda \gg v/l_0, \quad \lambda \gg \sigma_u^2\tau_0/l_0^2, \tag{15}$$

where l_0 is the correlation length of a random velocity field $\mathbf{u}(\mathbf{r}, t)$.

1.2. Tracer Density Field: Eulerian Statistical Approach

To develop a statistical model of the tracer density field, consider the indicator function

$$\Phi(t, \mathbf{r}; \rho) = \delta(\rho(\mathbf{r}, r) - \rho), \tag{16}$$

defined on a surface $\rho(\mathbf{r}, t) = \rho = \text{const}$ or a contour in three- or two-dimensional flow, respectively. In the general case, the evolution of this function is described by the Liouville equation

$$\begin{aligned}\left(\frac{\partial}{\partial t} + [\mathbf{v} + \mathbf{v}(\mathbf{r}, t)]\frac{\partial}{\partial \mathbf{r}}\right)\Phi(t, \mathbf{r}; \rho) \\ = \frac{\partial \mathbf{v}(\mathbf{r}, t)}{\partial \mathbf{r}}\frac{\partial}{\partial \rho}\rho\Phi(t, \mathbf{r}; \rho), \\ \Phi(0, \mathbf{r}; \rho) = \delta(\rho_0(\mathbf{r}) - \rho),\end{aligned} \tag{17}$$

which can be rewritten as

$$\begin{aligned}\left(\frac{\partial}{\partial t} + \mathbf{v}\frac{\partial}{\partial \mathbf{r}} + \frac{\partial}{\partial \mathbf{r}}\mathbf{u}(\mathbf{r}, t)\right)\Phi(t, \mathbf{r}; \rho) \\ = \frac{\partial \mathbf{v}(\mathbf{r}, t)}{\partial \mathbf{r}}\left[1 + \frac{\partial}{\partial \rho}\rho\right]\Phi(t, \mathbf{r}; \rho), \\ \Phi(0, \mathbf{r}; \rho) = \delta(\rho_0(\mathbf{r}) - \rho)\end{aligned} \tag{18}$$

in the case of a divergence-free velocity field $\mathbf{v}(\mathbf{r}, t)$, i.e., if $\partial \mathbf{v}(\mathbf{r}, t)/\partial \mathbf{r} \neq 0$. The single-point PDF corresponding to equation of motion (6) is the indicator function averaged over the ensemble of realizations of the random field $\mathbf{v}(\mathbf{r}, t)$:

$$P(t, \mathbf{r}; \rho) = \langle \Phi(t, \mathbf{r}; \rho) \rangle.$$

In [1], the problem was analyzed without taking into account sedimentation. If a zero-mean Gaussian stationary, homogeneous, and isotropic field is characterized by the correlation tensor

$$\langle v_i(\mathbf{r}, t) v_j(\mathbf{r}', t') \rangle = B_{ij}^{(\mathbf{v})}(\mathbf{r} - \mathbf{r}', t - t'),$$

then the single-point PDF $P(t, \mathbf{r}; \rho)$ corresponding to Eq. (18) obeys the following equation in both in the approximation of field $\mathbf{v}(\mathbf{r}, t)$ delta-correlated in time and in diffusion approximation:

$$\begin{aligned}\left(\frac{\partial}{\partial t} - D_0\frac{\partial^2}{\partial \mathbf{r}^2}\right)P(t, \mathbf{r}; \rho) = D_\rho\frac{\partial^2}{\partial \rho^2}\rho^2 P(t, \mathbf{r}; \rho), \\ P(0, \mathbf{r}; \rho) = \delta(\rho_0(\mathbf{r}) - \rho),\end{aligned} \tag{19}$$

where the spatial diffusivity D_0 and the diffusivity D_ρ in the ρ space are defined as

$$\begin{aligned}D_0 = \frac{1}{d}\int_0^\infty d\tau \langle \mathbf{v}(\mathbf{r}, t + \tau)\mathbf{v}(\mathbf{r}, t) \rangle = \frac{1}{d}\tau_{\mathbf{v}}\langle \mathbf{v}^2(\mathbf{r}, t) \rangle, \\ D_\rho = \int_0^\infty d\tau \left\langle \frac{\partial \mathbf{v}(\mathbf{r}, t + \tau)}{\partial \mathbf{r}}\frac{\partial \mathbf{v}(\mathbf{r}, t)}{\partial \mathbf{r}} \right\rangle \\ = \tau_{\mathrm{div}\mathbf{v}}\left\langle \left(\frac{\partial \mathbf{v}(\mathbf{r}, t)}{\partial \mathbf{r}}\right)^2 \right\rangle.\end{aligned} \tag{20}$$

They characterize, respectively, the spatial dispersion and clustering of a density field $\rho(\mathbf{r}, t)$. Here, $\tau_{\mathbf{v}}$ and $\tau_{\mathrm{div}\mathbf{v}}$ are the correlation times of the random fields $\mathbf{v}(\mathbf{r}, t)$ and $\partial \mathbf{v}(\mathbf{r}, t)/\partial \mathbf{r}$, and d is the space dimension.

It is clear that the random field $\mathbf{v}(\mathbf{r}, t)$ described by the closed nonlinear equation (6) is not Gaussian in the general case. However, it is also clear that $\mathbf{v}(\mathbf{r}, t)$ can be treated approximately as a Gaussian field in deriving Eq. (19), because higher order cumulants of the field $\mathrm{div}\,\mathbf{v}(\mathbf{r}, t)$ are smaller in order of magnitude as compared to the second-order cumulant function.

Indicator function (16) characterizes the geometry of the tracer density field $\rho(\mathbf{r}, t)$. In particular, (16) can be used to express the total area of the two-dimensional regions where $\rho(\mathbf{r}, t) > \rho$,

$$S(t; \rho) = \int_\rho^\infty d\rho' \int d\mathbf{r}\Phi(t, \mathbf{r}; \rho'), \tag{21}$$

or the tracer mass inside the regions,

$$M(t;\rho) = \int_{\rho}^{\infty} \rho' d\rho' \int d\mathbf{r} \Phi(t, \mathbf{r}; \rho'). \tag{22}$$

If $\mathbf{v}(\mathbf{r}, t)$ is a divergence-free random field, then clustering occurs with probability one, i.e., tracers group together into clusters of higher density surrounded by regions of relatively low tracer density; accordingly, $S(t;\rho) \longrightarrow 0$ and $M(t;\rho) \longrightarrow M_0$. Coherent behavior of this type is observed in almost every realization of the process (see [1, 2, 12, 13]), and its occurrence is independent of the model used to describe fluctuations. However, its characteristics (e.g., clustering time and length scales) may strongly depend on the model. In particular, it follows from Eq. (19) that the average area of clusters where the tracer density exceeds a certain ρ decreases in the course of time (when $Dt \gg 1$) as

$$\langle S(t;\rho)\rangle = \frac{1}{\sqrt{\pi\rho D_\rho t}} \exp\left\{-\frac{D_\rho t}{4}\right\} \int \sqrt{\rho_0(\mathbf{r})} d\mathbf{r}, \tag{23}$$

while the average mass inside such clusters,

$$\langle M(t;\rho)\rangle = M_0 - \sqrt{\frac{\rho}{\pi D_\rho t}} \exp\left\{-\frac{D_\rho t}{4}\right\} \int d\rho_0(\mathbf{r}) d\mathbf{r}, \tag{24}$$

approaches the total mass $M_0 = \int \rho_0(\mathbf{r}) d\mathbf{r}$.

Note that inertialess tracers advected by a divergence-free flow may be characterized by nonzero D_ρ. For example, the two-dimensional divergence does not vanish for buoyant tracers moving in a plane, whereas the three-dimensional divergence vanishes [14] (see also [15, 16]).

The diffusivities D_0 and D_ρ defined by (20) were calculated for low-inertia tracers advected by a random velocity field $\mathbf{u}(\mathbf{r}, t)$ were calculated in [1, 2].

Thus, the dynamics of tracer sedimentation is described by Eq. (17) or (18) for $\Phi(t, \mathbf{r}; \rho)$, where the tracer particle velocity $\mathbf{v}(\mathbf{r}, t)$ in the random flow field $\mathbf{u}(\mathbf{r}, t)$ is governed by Eq. (6). It is clear from the foregoing discussion that all calculations should be performed in diffusion approximation with a random field $\mathbf{u}(\mathbf{r}, t)$ having a finite correlation time τ_0. This problem is analyzed in the present study.

2. DIFFUSION AND CLUSTERING OF INERTIALESS SEDIMENTING PARTICLES

In the case of intertialess sedimenting tracers, statistical description of the density field in diffusion approximation relies on simplified Eqs. (17) and (18):

$$\left(\frac{\partial}{\partial t} + [\mathbf{v} + \mathbf{u}(\mathbf{r}, t)]\frac{\partial}{\partial \mathbf{r}}\right)\Phi(t, \mathbf{r}; \rho) = \frac{\partial \mathbf{u}(\mathbf{r}, t)}{\partial \mathbf{r}} \frac{\partial}{\partial \rho} \rho \Phi(t, \mathbf{r}; \rho), \tag{25}$$

$$\left(\frac{\partial}{\partial t} + \mathbf{v}\frac{\partial}{\partial \mathbf{r}} + \frac{\partial}{\partial \mathbf{r}}\mathbf{u}(\mathbf{r}, t)\right)\Phi(t, \mathbf{r}; \rho) = \frac{\partial \mathbf{u}(\mathbf{r}, t)}{\partial \mathbf{r}}\left[1 + \frac{\partial}{\partial \rho}\rho\right]\Phi(t, \mathbf{r}; \rho), \tag{26}$$

$$\Phi(0, \mathbf{r}; \rho) = \delta(\rho_0(\mathbf{r}) - \rho).$$

Equation (25) is well suited for finding the dependence of $\Phi(t, \mathbf{r}; \rho)$ on $\mathbf{u}(\mathbf{r}, t)$; Eq. (26), for averaging the desired function directly over the ensemble of realizations of $\mathbf{u}(\mathbf{r}, t)$.

Since both (25) and (26) are first-order equations with respect to time, the following dynamic causality condition holds:

$$\frac{\partial \Phi(t, \mathbf{r}; \rho)}{\delta u_j(\mathbf{r}', t')} = 0 \quad \text{for} \quad t' < 0 \quad \text{and} \quad t' > t, \tag{27}$$

i.e., the functional $\Phi(t, \mathbf{r}; \rho)$ depends on $u_j(\mathbf{r}, t')$ only within the interval $t_0 \le t' \le t$. As $t' \longrightarrow t$, the following equation for the variational derivative is obtained:

$$\frac{\delta \Phi(t, \mathbf{r}; \rho)}{\delta u_j(\mathbf{r}', t-0)} = \left[-\delta(\mathbf{r} - \mathbf{r}')\frac{\partial}{\partial r_j} + \frac{\partial \delta(\mathbf{r} - \mathbf{r}')}{\partial r_j}\frac{\partial}{\partial \rho}\rho\right]\Phi(t, \mathbf{r}; \rho). \tag{28}$$

The equation for $P(t, \mathbf{r}; \rho)$ is derived by averaging Eq. (26) over the ensemble of realizations of $\mathbf{u}(\mathbf{r}, t)$:

$$\left(\frac{\partial}{\partial t} + \mathbf{v}\frac{\partial}{\partial \mathbf{r}}\right)P(t, \mathbf{r}; \rho) = -\frac{\partial}{\partial \mathbf{r}}\langle \mathbf{u}(\mathbf{r}, t)\Phi(t, \mathbf{r}; \rho)\rangle + \left[1 + \frac{\partial}{\partial \rho}\rho\right]\left\langle \frac{\partial \mathbf{u}(\mathbf{r}, t)}{\partial \mathbf{r}}\Phi(t, \mathbf{r}; \rho)\right\rangle. \tag{29}$$

Equation (29) can be rewritten as

$$\left(\frac{\partial}{\partial t} + \mathbf{v}\frac{\partial}{\partial \mathbf{r}}\right)P(t, \mathbf{r}; \rho) = -\frac{\partial}{\partial r_i}\int d\mathbf{r}' \times \int_0^t dt' B_{ij}^{(\mathbf{u})}(\mathbf{r} - \mathbf{r}', t - t')\left\langle \frac{\delta \Phi(t, \mathbf{r}; \rho)}{\delta u_j(\mathbf{r}', t')}\right\rangle + \left[1 + \frac{\partial}{\partial \rho}\rho\right]\int d\mathbf{r}' \int_0^t dt' \frac{\partial B_{ij}^{(\mathbf{u})}(\mathbf{r} - \mathbf{r}', t - t')}{\partial r_j}\left\langle \frac{\delta \Phi(t, \mathbf{r}; \rho)}{\delta u_j(\mathbf{r}', t')}\right\rangle \tag{30}$$

by applying the Furutsu–Novikov theorem [17, 18]

$$\langle u_i(\mathbf{r}, t)R[t; \mathbf{u}(\mathbf{y}, \tau)]\rangle = \int d\mathbf{r}' \int_0^t dt' B_{ij}^{(\mathbf{u})}(\mathbf{r}-\mathbf{r}', t-t')\left\langle \frac{\delta R[t; \mathbf{u}(\mathbf{y}, \tau)]}{\delta u_j(\mathbf{r}', t')}\right\rangle \quad (31)$$

to the correlation function of a Gaussian random field with an arbitrary functional $R[t; \mathbf{u}(\mathbf{y}, \tau)]$, which is valid under dynamic causality condition (27).

In the diffusion approximation for Eq. (30), fluctuations over time intervals on the order of τ_0 are neglected, and the variational derivative in (30) obeys the equation

$$\left(\frac{\partial}{\partial t} + \mathbf{v}\frac{\partial}{\partial \mathbf{r}}\right)\frac{\delta\Phi(t, \mathbf{r}; \rho)}{\delta u_j(\mathbf{r}', t')} = 0 \quad (32)$$

subject to initial condition (28); i.e.,

$$\left.\frac{\delta\Phi(t, \mathbf{r}; \rho)}{\delta u_j(\mathbf{r}', t')}\right|_{t=t'} = \left[-\delta(\mathbf{r}-\mathbf{r}')\frac{\partial}{\partial r_j} + \frac{\partial\delta(\mathbf{r}-\mathbf{r}')}{\partial r_j}\frac{\partial}{\partial\rho}\rho\right]\Phi(t', \mathbf{r}; \rho). \quad (33)$$

The solution to problem (32), (33) has the form

$$\frac{\delta\Phi(t, \mathbf{r}; \rho)}{\delta u_j(\mathbf{r}', t')} = \exp\left\{-\mathbf{v}(t-t')\frac{\partial}{\partial \mathbf{r}}\right\} \times\left[-\delta(\mathbf{r}-\mathbf{r}')\frac{\partial}{\partial r_j} + \frac{\partial\delta(\mathbf{r}-\mathbf{r}')}{\partial r_j}\frac{\partial}{\partial\rho}\rho\right]\Phi(t', \mathbf{r}; \rho). \quad (34)$$

The evolution of the function $\Phi(t, \mathbf{r}; \rho)$ over a time interval on the order of τ_0 is governed by the equation

$$\left(\frac{\partial}{\partial t} + \mathbf{v}\frac{\partial}{\partial \mathbf{r}}\right)\Phi(t, \mathbf{r}; \rho) = 0,$$

$$\Phi(t, \mathbf{r}; \rho)|_{t=t'} = \Phi(t', \mathbf{r}; \rho).$$

Hence,

$$\Phi(t', \mathbf{r}; \rho) = \exp\left\{\mathbf{v}(t-t')\frac{\partial}{\partial \mathbf{r}}\right\}\Phi(t, \mathbf{r}; \rho). \quad (35)$$

Substituting (35) into (34), one obtains a final expression for the variational derivative valid in the diffusion approximation ($\tau = t - t'$):

$$\frac{\delta\Phi(t, \mathbf{r}; \rho)}{\delta u_j(\mathbf{r}', t')} = \left[-\delta(\mathbf{r}-\mathbf{r}'-\mathbf{v}\tau)\frac{\partial}{\partial r_j} + \frac{\partial\delta(\mathbf{r}-\mathbf{r}'-\mathbf{v}\tau)}{\partial r_j}\frac{\partial}{\partial\rho}\rho\right]\Phi(t, \mathbf{r}; \rho). \quad (36)$$

A closed equation for the tracer-density PDF in the diffusion approximation is obtained by substituting (36) into (30) and integrating the result over $\mathbf{r}'$:

$$\left(\frac{\partial}{\partial t} + \mathbf{v}\frac{\partial}{\partial \mathbf{r}}\right)P(t, \mathbf{r}; \rho) = \int_0^t d\tau B_{ij}^{(\mathbf{u})}(\mathbf{v}\tau, \tau)\frac{\partial^2}{\partial r_i\partial r_j}P(t, \mathbf{r}; \rho) - \int_0^t d\tau\frac{\partial B_{ij}^{(\mathbf{u})}(\mathbf{v}\tau, \tau)}{\partial r_i}\frac{\partial}{\partial r_j}P(t, \mathbf{r}; \rho) - \int_0^t d\tau\frac{\partial B_{ij}^{(\mathbf{u})}(\mathbf{v}\tau, \tau)}{\partial r_i\partial r_j}\frac{\partial^2}{\partial\rho^2}\rho^2 P(t, \mathbf{r}; \rho). \quad (37)$$

In the long-time limit ($t \gg \tau_0$, $t \gg l_0/v$), the upper limits in the integrals on the right-hand side of (37) can be set to infinity. The resulting PDF equation is

$$\left(\frac{\partial}{\partial t} + \mathbf{v}\frac{\partial}{\partial \mathbf{r}}\right)P(t, \mathbf{r}; \rho) = D_{ij}(\mathbf{v})\frac{\partial^2}{\partial r_i\partial r_j}P(t, \mathbf{r}; \rho) + G_j(\mathbf{v})\frac{\partial}{\partial r_j}P(t, \mathbf{r}; \rho) + D_\rho(\mathbf{v})\frac{\partial^2}{\partial\rho^2}\rho^2 P(t, \mathbf{r}; \rho), \quad (38)$$

where the transport coefficients are defined as

$$D_{ij}(\mathbf{v}) = \int_0^\infty d\tau B_{ij}^{(\mathbf{u})}(\mathbf{v}\tau, \tau) = \int_0^\infty d\tau\langle u_i(\mathbf{r}+\mathbf{v}\tau, t+\tau)u_j(\mathbf{r}, t)\rangle,$$

$$G_j(\mathbf{v}) = -\int_0^\infty d\tau\frac{\partial B_{ij}^{(\mathbf{u})}(\mathbf{v}\tau, \tau)}{\partial r_i} = -\int_0^\infty d\tau\left\langle\frac{\partial\mathbf{u}(r+\mathbf{v}\tau, t+\tau)}{\partial\mathbf{r}}u_j(\mathbf{r}, t)\right\rangle, \quad (39)$$

$$D_\rho(\mathbf{v}) = -\int_0^\infty d\tau\frac{\partial B_{ij}^{(\mathbf{u})}(\mathbf{v}\tau, \tau)}{\partial r_i\partial r_j} = \int_0^\infty d\tau\left\langle\frac{\partial\mathbf{u}(r+\mathbf{v}\tau, t+\tau)}{\partial\mathbf{r}}\frac{\partial\mathbf{u}(\mathbf{r}, t)}{\partial\mathbf{r}}\right\rangle.$$

The coefficients $D_{ij}(\mathbf{v})$ and $G_j(\mathbf{v})$ characterize the spatial dispersion of the density field for inertialess tracers, whereas clustering is controlled by the diffusivity $D_\rho(\mathbf{v})$. Note that the asymptotic laws of evolution of

the tracer density field are described by functionals (23) and (24) with D_ρ replaced by $D_\rho(\mathbf{v})$.

If $\operatorname{div}\mathbf{u}(\mathbf{r}, t) = 0$, then both $G_j(\mathbf{v})$ and $D_\rho(\mathbf{v})$ vanish in (38), the inertialess-tracer density field evolves without clustering, and clustering is possible only for low-inertia tracers.

An equation for the mean tracer density is obtained by integrating the product of Eq. (38) with ρ over ρ:

$$\left(\frac{\partial}{\partial t} + \mathbf{v}\frac{\partial}{\partial \mathbf{r}}\right)\langle \rho(\mathbf{r}, t)\rangle = D_{ij}(\mathbf{v})\frac{\partial^2}{\partial r_i \partial r_j}\langle \rho(\mathbf{r}, t)\rangle + G_j(\mathbf{v})\frac{\partial}{\partial r_j}\langle \rho(\mathbf{r}, t)\rangle. \tag{40}$$

In view of (10) and (12), this equation is similar to the equation for the PDF of particle location. Effects due to molecular diffusion can easily be taken into account by averaging Eq. (3) in the inertialess limit over the ensemble of realizations of the field $\mathbf{u}(\mathbf{r}, t)$ and using (31). The resulting equation for the mean density is

$$\left(\frac{\partial}{\partial t} + \mathbf{v}\frac{\partial}{\partial \mathbf{r}}\right)\langle \rho(\mathbf{r}, t)\rangle = \mu\Delta\langle \rho(\mathbf{r}, t)\rangle - \frac{\partial}{\partial r_i}\int d\mathbf{r}' \int_0^t dt' B_{ij}^{(\mathbf{u})}(\mathbf{r} - \mathbf{r}', t - t')\left\langle \frac{\delta \rho(\mathbf{r}, t)}{\delta u_j(\mathbf{r}', t')}\right\rangle, \tag{41}$$
$$\langle \rho(\mathbf{r}, 0)\rangle = \rho_0(\mathbf{r}).$$

In the diffusion approximation, the variational derivative in Eq. (41) satisfies the equation

$$\left(\frac{\partial}{\partial t} + \mathbf{v}\frac{\partial}{\partial \mathbf{r}}\right)\frac{\delta \rho(\mathbf{r}, t)}{\delta u_j(\mathbf{r}', t')} = \mu\Delta\frac{\delta \rho(\mathbf{r}, t)}{\delta u_j(\mathbf{r}', t')} \tag{42}$$

subject to an initial condition that follows from (3),

$$\left.\frac{\delta \rho(\mathbf{r}, t)}{\delta u_j(\mathbf{r}', t')}\right|_{t = t'} = -\frac{\partial}{\partial r_j}\delta(\mathbf{r} - \mathbf{r}')\rho(\mathbf{r}, t). \tag{43}$$

The solution to problem (42), (43) has the form ($\tau = t - t'$)

$$\frac{\delta \rho(\mathbf{r}, t)}{\delta v_j(\mathbf{r}', t')} = -\exp\left\{\left(\mu\Delta - \mathbf{v}\frac{\partial}{\partial \mathbf{r}}\right)\tau\right\}\frac{\partial}{\partial r_j}\delta(\mathbf{r} - \mathbf{r}')\rho(\mathbf{r}, t'). \tag{44}$$

The evolution of $\rho(\mathbf{r}, t)$ over a time interval on the order of $\tau_\mathbf{v}$ is governed by the equation

$$\left(\frac{\partial}{\partial t} + \mathbf{v}\frac{\partial}{\partial \mathbf{r}}\right)\rho(\mathbf{r}, t) = \mu\Delta\rho(\mathbf{r}, t),$$
$$\rho(\mathbf{r}, t)|_{t = t'} = \rho(\mathbf{r}, t').$$

Hence,

$$\rho(\mathbf{r}, t') = \exp\left\{-\left(\mu\Delta - \mathbf{v}\frac{\partial}{\partial \mathbf{r}}\right)\tau\right\}\rho(\mathbf{r}, t). \tag{45}$$

Substituting (45) into the right-hand side of (44) yields a final expression for the variational derivative valid in the diffusion approximation ($\tau = t - t'$):

$$\frac{\delta \rho(\mathbf{r}, t)}{\delta u_j(\mathbf{r}', t')} = -e^{\mu\Delta\tau}\frac{\partial}{\partial r_j} \times \delta(\mathbf{r} - \mathbf{r}' - \mathbf{v}\tau)e^{-\mu\Delta\tau}\rho(\mathbf{r}, t). \tag{46}$$

A closed equation describing the evolution of the mean density in the diffusion approximation is obtained by substituting (46) into (41) and integrating the result over $\mathbf{r}'$:

$$\left(\frac{\partial}{\partial t} + \mathbf{v}\frac{\partial}{\partial \mathbf{r}}\right)\langle \rho(\mathbf{r}, t)\rangle = \mu\Delta\langle \rho(\mathbf{r}, t)\rangle + \frac{\partial}{\partial r_i}\int d\mathbf{r}' \int_0^t d\tau B_{ij}^{(\mathbf{u})}(\mathbf{r} - \mathbf{r}', \tau) \times e^{\mu\tau\Delta}\frac{\partial}{\partial r_j}\delta(\mathbf{r} - \mathbf{r}' - \mathbf{v}\tau)e^{-\mu\tau\Delta}\langle \rho(\mathbf{r}, \tau)\rangle, \tag{47}$$
$$\langle \rho(\mathbf{r}, 0)\rangle = \rho_0(\mathbf{r}).$$

In the long-time limit ($t \gg t_0$, $t \gg l_0/v$), the upper limits in the integrals on the right-hand side of (47) can be set to infinity. A solution to the resulting equation for the mean tracer density,

$$\left(\frac{\partial}{\partial t} + \mathbf{v}\frac{\partial}{\partial \mathbf{r}}\right)\langle \rho(\mathbf{r}, t)\rangle = \mu\Delta\langle \rho(\mathbf{r}, t)\rangle + \frac{\partial^2}{\partial r_i \partial r_j}\int d\mathbf{r}' \int_0^\infty d\tau B_{ij}^{(\mathbf{u})}(\mathbf{r} - \mathbf{r}', \tau) \times e^{\mu\tau\Delta}\delta(\mathbf{r} - \mathbf{r}' - \mathbf{v}\tau)e^{-\mu\tau\Delta}\langle \rho(\mathbf{r}, \tau)\rangle - \frac{\partial}{\partial r_i}\int d\mathbf{r}' \int_0^\infty d\tau \frac{\partial B_{ij}^{(\mathbf{u})}(\mathbf{r} - \mathbf{r}', \tau)}{\partial r_j} \times e^{\mu\tau\Delta}\delta(\mathbf{r} - \mathbf{r}' - \mathbf{v}\tau)e^{-\mu\tau\Delta}\langle \rho(\mathbf{r}, \tau)\rangle,$$

can be obtained in explicit form.

By introducing the Fourier transform of the density field,

$$\rho(\mathbf{r}, t) = \int d\mathbf{q}\rho_\mathbf{q}(t)e^{i\mathbf{q}\cdot\mathbf{r}},$$
$$\rho_\mathbf{q}(t) = \frac{1}{(2\pi)^d}\int d\mathbf{r}\rho(\mathbf{r}, t)e^{-i\mathbf{q}\cdot\mathbf{r}},$$

and using the spectral tensor defined in (8), the follow-

ing equation is obtained:

$$\left(\frac{\partial}{\partial t} + i\mathbf{v}\cdot\mathbf{q}\right)\langle\rho_{\mathbf{q}}(t)\rangle$$
$$= -\{\mu\mathbf{q}^2 + q_i q_j D_{ij}(\mathbf{q},\mathbf{v}) - q_i G_i(\mathbf{q},\mathbf{v})\}\langle\rho_{\mathbf{q}}(t)\rangle,$$

where

$$D_{ij}(\mathbf{q},\mathbf{v}) = \int d\mathbf{k}\int_0^\infty d\tau E_{ij}(\mathbf{k},\tau)$$
$$\times\exp\{-\mu(\mathbf{k}^2 - 2\mathbf{k}\cdot\mathbf{q})\tau + i\mathbf{k}\cdot\mathbf{v}\tau\},$$

$$G_i(\mathbf{q},\mathbf{v}) = \int d\mathbf{k}\int_0^\infty d\tau k_j E_{ij}(\mathbf{k},\tau)$$
$$\times\exp\{-\mu(\mathbf{k}^2 - 2\mathbf{k}\cdot\mathbf{q})\tau + i\mathbf{k}\cdot\mathbf{v}\tau\}.$$

Its solution is

$$\langle\rho_{\mathbf{q}}(t)\rangle = \rho_{\mathbf{q}}(0)$$
$$\times\exp\{-i\mathbf{v}\cdot\mathbf{q}t - \mu\mathbf{q}^2 t - q_i q_j D_{ij}(\mathbf{q},\mathbf{v})t + q_i G_i(\mathbf{q},\mathbf{v})t\}.$$

Therefore,

$$\langle\rho(\mathbf{r},t)\rangle = \int d\mathbf{r}'\rho_0(\mathbf{r}')P(\mathbf{r},t|\mathbf{r}'),$$

where

$$P(\mathbf{r},t|\mathbf{r}') = \frac{1}{(2\pi)^d}\int d\mathbf{q}\exp\{i\mathbf{q}(\mathbf{r}-\mathbf{r}'-\mathbf{v}t)$$
$$-\mu\mathbf{q}^2 t - q_i q_j D_{ij}(\mathbf{q},\mathbf{v})t + q_i G_i(\mathbf{q},\mathbf{v})t\} \quad (48)$$

can be interpreted, in view of (10) and (12), as the PDF of particle location affected by molecular diffusion if a Gaussian random forcing term $\mathbf{f}(t)$ is introduced into Eq. (9):

$$\frac{d}{dt}\mathbf{r}(t) = \mathbf{v} + \mathbf{u}(\mathbf{r}(t),t) + \mathbf{f}(t), \quad \mathbf{r}(0) = \mathbf{r}',$$

$$\langle\mathbf{f}(t)\rangle = 0, \quad \langle f_i(t) f_j(t')\rangle = 2\mu\delta_{ij}\delta(t-t').$$

In the general case, the PDF given by (48) is non-Gaussian, but it simplifies to a Gaussian one in the long-time limit ($t \gg \tau_0$, $t \gg l_0/v$):

$$P(\mathbf{r},t|\mathbf{r}') = \frac{1}{(2\pi)^d}\int d\mathbf{q}\exp\{i\mathbf{q}\cdot(\mathbf{r}-\mathbf{r}'-\mathbf{v}t)$$
$$-\mu\mathbf{q}^2 t - q_i q_j D_{ij}(\mathbf{v})t + q_i G_i(\mathbf{v})t\}, \quad (49)$$

where

$$D_{ij}(\mathbf{v}) = D_{ij}(0,\mathbf{v})$$
$$= \int d\mathbf{k}\int_0^\infty d\tau E_{ij}(\mathbf{k},\tau)\exp\{-\mu\mathbf{k}^2\tau + i\mathbf{k}\cdot\mathbf{v}\tau\},$$

$$G_i(\mathbf{v}) = G_i(0,\mathbf{v})$$
$$= \int d\mathbf{k}\int_0^\infty d\tau k_j E_{ij}(\mathbf{k},\tau)\exp\{-\mu\mathbf{k}^2\tau + i\mathbf{k}\cdot\mathbf{v}\tau\}.$$

The diffusivity tensor $D_{ij}(\mathbf{v})$ can be represented as

$$D_{ij}(\mathbf{v}) = A(v)\frac{v_i v_j}{\mathbf{v}^2} + B(v)\Delta_{ij}(\mathbf{v}),$$

where

$$A(v) = D_{ij}(\mathbf{v})\frac{v_i v_j}{\mathbf{v}^2}, \quad B(v) = \frac{1}{d-1}D_{ij}(\mathbf{v})\Delta_{ij}(\mathbf{v})$$

and $\Delta_{ij}(\mathbf{v}) = \delta_{ij} - v_i v_j/\mathbf{v}^2$. According to this representation, if the z axis is aligned with $\mathbf{v}$, then diffusion of the density field along the z axis and in the transverse plane $\mathbf{R}$ is controlled by $D_{zz}(v) = A(v)$ and $D_\perp(v) = B(v)$, respectively. Moreover, additional particle transport along the z axis takes place, due to the divergent nature of the field $\mathbf{u}(\mathbf{r},t)$, and

$$G_j(\mathbf{v}) = G(v)\frac{v_j}{\mathbf{v}^2} = \frac{v_j}{\mathbf{v}^2}\int d\mathbf{k}\int_0^\infty d\tau E^{\mathrm{p}}(k,\tau)(i\mathbf{k}\cdot\mathbf{v})$$
$$\times\exp\{-\mu\mathbf{k}^2 + i\mathbf{k}\cdot\mathbf{v}\tau\}. \quad (50)$$

To estimate the degree of diffusivity anisotropy, suppose that

$$E_{ij}(\mathbf{k},\tau) = E_{ij}(\mathbf{k})\exp\{-|\tau|/\tau_0\},$$

where τ_0 is the correlation time of the random velocity field. Then,

$$D_{ij}(\mathbf{v}) = \frac{1}{v}\int\frac{d\mathbf{k}}{k}E_{ij}(\mathbf{k})\frac{p(k,v)}{1+p^2(k,v)\cos^2\theta},$$

where $\cos^2\theta = (\mathbf{k}\cdot\mathbf{v})^2/k^2\mathbf{v}^2$ and

$$p(k,v) = \frac{kv\tau_0}{1+\mu\tau_0\mathbf{k}^2}.$$

In the three-dimensional case,

$$G_j(\mathbf{v}) = -\frac{v_j}{\mathbf{v}^2}\int d\mathbf{k}E^{\mathrm{pot}}(k)\frac{p^2(k,v)\cos^2\theta}{1+p^2(k,v)\cos^2\theta}$$
$$= -\frac{4\pi v_j}{\mathbf{v}^2}\int_0^\infty dk k^2 E^{\mathrm{pot}}(k)\left\{1 - \frac{1}{p(k,v)}\arctan p(k,v)\right\}.$$

If the velocity field is divergence-free (e.g., for $d = 3$), then $E_{ij}(\mathbf{k}) = E^s(k)\Delta_{ij}(\mathbf{k})$, and

$$D_{zz}(v) = \frac{4\pi}{v}\int_0^\infty dk k E^{sol}(k) f_\parallel(k, v),$$

$$D_\perp(v) = \frac{4\pi}{v}\int_0^\infty dk k E^{sol}(k) f_\perp(k, v),$$

where

$$f_\parallel(k, v) = \left\{\arctan p(k, v) + \frac{1}{p(k, v)}\left(\frac{1}{p(k, v)}\arctan p(k, v) - 1\right)\right\},$$

$$f_\perp(k, v) = \left\{\arctan p(k, v) - \frac{1}{p(k, v)}\left(\frac{1}{p(k, v)}\arctan p(k, v) - 1\right)\right\}.$$

If p is small ($v\tau_0 \ll l_0$), then both $f_\parallel(k, v)$ and $f_\perp(k, v)$ are close to $2p/3$, and diffusion is independent of the sedimentation velocity v. If p is large ($v\tau_0 \gg l_0$), then $f_\parallel(k, v) = 2f_\perp(k, v) \approx \pi/2$. The anisotropy is explained by the shorter times required for particles to pass through regions of correlated velocities when tracers diffuse relative to turbulent motion and by the fact that the transverse correlation length of a random velocity field is one half of its longitudinal correlation length [6, 19]. If $\mu\tau_0 \ll l_0^2$, then $D_{ij}(\mathbf{v})$ is independent of μ. Note that these estimates are valid if the integrals involved are convergent. Diffusion in the well-developed turbulent flow characterized by the Kolmogorov power spectrum was analyzed in [11].

In the general case of a coordinate system oriented in the direction of sedimentation ($\mathbf{r} = \{z, \mathbf{R}\}$), Eq. (38) has the form

$$\left(\frac{\partial}{\partial t} + \mathbf{v}\frac{\partial}{\partial \mathbf{r}}\right)P(t, \mathbf{r}; \rho) = A(v)\frac{\partial^2}{\partial z^2}P(t, \mathbf{r}; \rho) + B(v)\frac{\partial^2}{\partial \mathbf{R}^2}P(t, \mathbf{r}; \rho) + G(v)\frac{\partial}{\partial z}P(t, \mathbf{r}; \rho) + D_\rho(v)\frac{\partial^2}{\partial \rho^2}\rho^2 P(t, \mathbf{r}; \rho), \quad (51)$$

where

$$G(v) = -\int_0^\infty d\tau \frac{\partial B_{iz}^{(\mathbf{u})}(v\tau, 0, \tau)}{\partial r_i}.$$

3. LOW-INERTIA TRACERS

3.1. General Analysis

It was noted above that analysis of clustering of sedimenting tracers in a divergence-free flow field must take into account the inertia of particles. Accordingly, recall Eqs. (6),

$$\left(\frac{\partial}{\partial t} + [\mathbf{v} + \mathbf{v}(\mathbf{r}, t)]\frac{\partial}{\partial \mathbf{r}}\right)\rho(\mathbf{r}, t) = -\frac{\partial \mathbf{v}(\mathbf{r}, t)}{\partial \mathbf{r}}\rho(\mathbf{r}, t), \quad \rho(\mathbf{r}, 0) = \rho_0(\mathbf{r}), \quad (52)$$

$$\left(\frac{\partial}{\partial t} + [\mathbf{v} + \mathbf{v}(\mathbf{r}, t)]\frac{\partial}{\partial \mathbf{r}}\right)\mathbf{v}(\mathbf{r}, t) = -\lambda[\mathbf{v}(\mathbf{r}, t) - \mathbf{u}(\mathbf{r}, t)], \quad (53)$$

and assume that $\mathbf{v}(\mathbf{r}, t)$ is a zero-mean Gaussian stationary and homogeneous field characterized by the correlation tensor

$$\langle v_i(\mathbf{r}, t) v_j(\mathbf{r}', t')\rangle = B_{ij}^{(\mathbf{v})}(\mathbf{r} - \mathbf{r}', t - t').$$

The single-point PDF $P(t, \mathbf{r}; \rho)$ corresponding to equation of motion (52) is governed by an equation similar to (38):

$$\left(\frac{\partial}{\partial t} + \mathbf{v}\frac{\partial}{\partial \mathbf{r}}\right)P(t, \mathbf{r}; \rho) = D_{ij}(\mathbf{v}; t)\frac{\partial^2}{\partial r_i \partial r_j}P(t, \mathbf{r}; \rho) + G_j(\mathbf{v}; t)\frac{\partial}{\partial r_j}P(t, \mathbf{r}; \rho) + D_\rho(\mathbf{v}; t)\frac{\partial^2}{\partial \rho^2}\rho^2 P(t, \mathbf{r}; \rho), \quad (54)$$

with transport coefficients defined as

$$D_{ij}(\mathbf{v}; t) = \int_0^t d\tau B_{ij}^{(\mathbf{v})}(\mathbf{v}\tau, \tau) = \int_0^t dt' \langle v_i(\mathbf{r} + \mathbf{v}t, t) v_j(\mathbf{r} + \mathbf{v}t', t')\rangle,$$

$$G_j(\mathbf{v}; t) = -\int_0^t d\tau \frac{\partial B_{ij}^{(\mathbf{v})}(\mathbf{v}\tau, \tau)}{\partial r_i} = -\int_0^t dt' \left\langle \frac{\partial \mathbf{v}(\mathbf{r} + \mathbf{v}t, t)}{\partial \mathbf{r}} v_j(\mathbf{r} + \mathbf{v}t', t')\right\rangle, \quad (55)$$

$$D_\rho(\mathbf{v}; t) = -\int_0^t d\tau \frac{\partial B_{ij}^{(\mathbf{v})}(\mathbf{v}\tau, \tau)}{\partial r_i \partial r_j} = \int_0^t dt' \left\langle \frac{\partial \mathbf{v}(\mathbf{r} + \mathbf{v}t, t)}{\partial \mathbf{r}}\frac{\partial \mathbf{v}(\mathbf{r} + \mathbf{v}t', t')}{\partial \mathbf{r}}\right\rangle.$$

Here, $D_{ij}(\mathbf{v}; t)$ characterizes the spatial dispersion of the density field for low-inertia tracers, whereas clustering is controlled by the diffusivity $D_\rho(\mathbf{v}; t)$.

Thus, stochastic equation (53) must be used to evaluate transport coefficients (55), i.e., spatiotemporal correlation functions for the random fields $\mathbf{v}(\mathbf{r}, t)$ and $\partial\mathbf{v}(\mathbf{r}, t)/\partial\mathbf{r}$.

Equation (53) can be rewritten as

$$\exp\left(-\mathbf{v}t\frac{\partial}{\partial\mathbf{r}}\right)\frac{\partial}{\partial t}\left[\exp\left(\mathbf{v}t\frac{\partial}{\partial\mathbf{r}}\right)\mathbf{v}(\mathbf{r}, t)\right]$$

$$= -\left(\mathbf{v}(\mathbf{r}, t)\frac{\partial}{\partial\mathbf{r}}\right)\mathbf{v}(\mathbf{r}, t) - \lambda[\mathbf{v}(\mathbf{r}, t) - \mathbf{u}(\mathbf{r}, t)].$$

When represented in terms of

$$\tilde{\mathbf{v}}(\mathbf{r}, t) = \mathbf{v}(\mathbf{r} + \mathbf{v}t, t), \quad \tilde{\mathbf{u}}(\mathbf{r}, t) = \mathbf{u}(\mathbf{r} + \mathbf{v}t, t),$$

it does not contain $\mathbf{v}$:

$$\frac{\partial}{\partial t}\tilde{\mathbf{v}}(\mathbf{r}, t) = -\left(\tilde{\mathbf{v}}(\mathbf{r}, t)\frac{\partial}{\partial\mathbf{r}}\right)\tilde{\mathbf{v}}(\mathbf{r}, t) - \lambda[\tilde{\mathbf{v}}(\mathbf{r}, t) - \tilde{\mathbf{u}}(\mathbf{r}, t)]. \tag{56}$$

The new variables are the velocity fields in a coordinate system moving with the settling tracers. Transport coefficients (55) are expressed accordingly:

$$D_{ij}(\mathbf{v}; t) = \int_0^t d\tau B_{ij}^{(\mathbf{v})}(\mathbf{v}\tau, \tau)$$

$$= \int_0^t dt' \langle \tilde{V}_i(\mathbf{r}, t)\tilde{V}_j(\mathbf{r}, t')\rangle = \int_0^t d\tau B_{ij}^{(\tilde{\mathbf{v}})}(0, \tau).$$

$$G_j(\mathbf{v}; t) = -\int_0^t d\tau\frac{\partial B_{ij}^{(\mathbf{v})}(\mathbf{v}\tau, \tau)}{\partial r_i}$$

$$= -\int_0^t dt' \left\langle\frac{\partial\tilde{\mathbf{v}}(\mathbf{r}, t)}{\partial\mathbf{r}}\tilde{V}_j(\mathbf{r}, t')\right\rangle = -\int_0^t d\tau\frac{\partial B_{ij}^{(\tilde{\mathbf{v}})}(0, \tau)}{\partial r_i}, \tag{57}$$

$$D_\rho(\mathbf{v}; t) = -\int_0^t d\tau\frac{\partial B_{ij}^{(\mathbf{v})}(\mathbf{v}\tau, \tau)}{\partial r_i\partial r_j}$$

$$= \int_0^t dt' \left\langle\frac{\partial\tilde{\mathbf{v}}(\mathbf{r}, t)}{\partial\mathbf{r}}\frac{\partial\tilde{\mathbf{v}}(\mathbf{r}, t')}{\partial\mathbf{r}}\right\rangle = -\int_0^t d\tau\frac{\partial B_{ij}^{(\tilde{\mathbf{v}})}(0, \tau)}{\partial r_i\partial r_j}.$$

In the long-time limit ($t \gg \tau_0$, $t \gg l_0/v$), the upper limits in the integrals in (57) can be set to infinity. The resulting transport coefficients

$$D_{ij}(\mathbf{v}) = \int_0^\infty d\tau B_{ij}^{(\tilde{\mathbf{v}})}(0, \tau),$$

$$G_j(\mathbf{v}) = -\int_0^\infty d\tau\frac{\partial B_{ij}^{(\tilde{\mathbf{v}})}(0, \tau)}{\partial r_i}, \tag{58}$$

$$D_\rho(\mathbf{v}) = -\int_0^\infty d\tau\frac{\partial B_{ij}^{(\tilde{\mathbf{v}})}(0, \tau)}{\partial r_i\partial r_j}$$

are determined by spatiotemporal statistical characteristics of the solution to Eq. (56).

Consider a divergence-free zero-mean Gaussian stationary, homogeneous, and isotropic field $\mathbf{u}(\mathbf{r}, t)$ characterized by the correlation tensor

$$B_{ij}^{(\mathbf{u})}(\mathbf{r} - \mathbf{r}', t - t') = \langle u_i(\mathbf{r}, t)u_j(\mathbf{r}', t')\rangle.$$

In a coordinate system moving with the settling tracers,

$$B_{ij}^{(\tilde{\mathbf{u}})}(\mathbf{r} - \mathbf{r}', t - t') = \langle\tilde{u}_i(\mathbf{r}, t)\tilde{u}_j(\mathbf{r}', t')\rangle = B_{ij}^{(\mathbf{u})}(\mathbf{r} - \mathbf{r}' + \mathbf{v}(t - t'), t - t'), \tag{59}$$

and a spatial spectral density of the field $\mathbf{u}(\mathbf{r}, t)$ can be defined:

$$B_{ij}^{(\mathbf{u})}(\mathbf{r}, t) = \int d\mathbf{k}E_{ij}(\mathbf{k}, t)e^{i\mathbf{k}\cdot\mathbf{r}},$$

$$E_{ij}(\mathbf{k}, t) = E(k, t)\left(\delta_{ij} - \frac{k_ik_j}{\mathbf{k}^2}\right). \tag{60}$$

The following representation is valid for a fourth-rank tensor used below:

$$-\frac{\partial^2 B_{ij}^{(\mathbf{u})}(0, 0)}{\partial r_k\partial r_l} = \frac{D}{d(d + 2)} \times [(d + 1)\delta_{kl}\delta_{ij} - \delta_{ki}\delta_{lj} - \delta_{kj}\delta_{li}], \tag{61}$$

where

$$D = \int d\mathbf{k}\mathbf{k}^2E(k) = -\frac{1}{d - 1}\langle\mathbf{u}(\mathbf{r}, t)\Delta\mathbf{u}(\mathbf{r}, t)\rangle \tag{62}$$

is associated with the solenoidal component of the divergence-free random field $\mathbf{u}(\mathbf{r}, t)$.

In the diffusion approximation, the Eulerian statistical model of clustering for the density field $\rho(\mathbf{r}, t)$ governed by Eq. (56) is obtained by evaluating the transport coefficients in (58).

If $\sigma_{\mathbf{u}}^2$ is small, then Eq. (56) can be linearized with respect to $\tilde{\mathbf{u}}(\mathbf{r}, t)$. The simplified equation

$$\left(\frac{\partial}{\partial t} + \tilde{\mathbf{u}}(\mathbf{r}, t)\frac{\partial}{\partial \mathbf{r}}\right)\tilde{\mathbf{v}}(\mathbf{r}, t)$$

$$= -\left(\tilde{\mathbf{v}}(\mathbf{r}, t)\frac{\partial}{\partial \mathbf{r}}\right)\tilde{\mathbf{u}}(\mathbf{r}, t) - \lambda[\tilde{\mathbf{v}}(\mathbf{r}, t) - \tilde{\mathbf{u}}(\mathbf{r}, t)]$$

can be rewritten as

$$\left(\frac{\partial}{\partial t} + \lambda\right)\tilde{v}_i(\mathbf{r}, t) = -\tilde{u}_k(\mathbf{r}, t)\frac{\partial \tilde{v}_i(\mathbf{r}, t)}{\partial r_k} - \frac{\partial \tilde{u}_i(\mathbf{r}, t)}{\partial r_k}\tilde{v}_k(\mathbf{r}, t) + \lambda\tilde{u}_i(\mathbf{r}, t), \tag{63}$$

where summation over repeated indices is assumed.

3.2. Diffusion Approximation

The random field $\tilde{\mathbf{u}}(\mathbf{r}, t)$ correlates with $\tilde{\mathbf{v}}(\mathbf{r}, t)$, which is a functional of $\tilde{\mathbf{u}}(\mathbf{r}, t)$. According to (59), correlations are decoupled for the Gaussian field $\tilde{\mathbf{u}}(\mathbf{r}, t)$ by applying the Furutsu–Novikov theorem rewritten as

$$\langle \tilde{u}_k(\mathbf{r}, t)R[t; \tilde{\mathbf{u}}(\mathbf{y}, \tau)]\rangle = \int d\mathbf{r}'\int_0^t dt' \times B_{kl}^{(\mathbf{u})}(\mathbf{r} - \mathbf{r}' + \mathbf{v}(t - t'), t - t')\left\langle \frac{\delta R[t; \tilde{\mathbf{u}}(\mathbf{y}, \tau)]}{\delta \tilde{u}_l(\mathbf{r}', t')}\right\rangle. \tag{64}$$

In the diffusion approximation, the equation

$$\left(\frac{\partial}{\partial t} + \lambda\right)\frac{\delta \tilde{v}_i(\mathbf{r}, t)}{\delta \tilde{u}_l(\mathbf{r}', t')} = 0$$

for the variational derivative is subject to the initial condition

$$\left.\frac{\delta \tilde{v}_i(\mathbf{r}, t)}{\delta \tilde{u}_l(\mathbf{r}', t')}\right|_{t = t' + 0} = -\left[\delta(\mathbf{r} - \mathbf{r}')\frac{\partial \tilde{v}_i(\mathbf{r}, t')}{\partial r_l} + \delta_{il}\frac{\partial \delta(\mathbf{r} - \mathbf{r}')}{\partial r_k}\tilde{v}_k(\mathbf{r}, t')\right] + \lambda\delta_{il}\delta(\mathbf{r} - \mathbf{r}'),$$

at $t = t'$, which follows from (63). Its solution has the form

$$\frac{\delta \tilde{v}_i(\mathbf{r}, t)}{\delta \tilde{u}_l(\mathbf{r}', t')} = e^{-\lambda(t - t')}\left\{-\left[\delta(\mathbf{r} - \mathbf{r}')\frac{\partial \tilde{v}_i(\mathbf{r}, t')}{\partial r_l} + \frac{\partial \delta(\mathbf{r} - \mathbf{r}')}{\partial r_k}\delta_{il}\tilde{v}_k(\mathbf{r}, t')\right] + \lambda\delta_{il}\delta(\mathbf{r} - \mathbf{r}')\right\}.$$

Since

$$\tilde{\mathbf{v}}(\mathbf{r}, t) = e^{-\lambda(t - t')}\tilde{\mathbf{v}}(\mathbf{r}, t')$$

and

$$\tilde{\mathbf{v}}(\mathbf{r}, t') = e^{\lambda(t - t')}\tilde{\mathbf{v}}(\mathbf{r}, t)$$

in the diffusion approximation, the variational derivative is expressed as

$$\frac{\delta \tilde{v}_i(\mathbf{r}, t)}{\delta \tilde{u}_l(\mathbf{r}', t')} = -\left[\delta(\mathbf{r} - \mathbf{r}')\frac{\partial \tilde{v}_i(\mathbf{r}, t)}{\partial r_l} + \delta_{il}\frac{\partial \delta(\mathbf{r} - \mathbf{r}')}{\partial r_\mu}\tilde{v}_\mu(\mathbf{r}, t)\right] + \lambda e^{-\lambda(t - t')}\delta_{il}\delta(\mathbf{r} - \mathbf{r}'). \tag{65}$$

3.3. Spatiotemporal Correlation Tensor for $\tilde{\mathbf{v}}(\mathbf{r}, t)$

The spatiotemporal correlation tensor of the field $\tilde{\mathbf{v}}(\mathbf{r}, t)$ at $t > t_1$ obeys the equation

$$\left(\frac{\partial}{\partial t} + \lambda\right)\langle \tilde{v}_i(\mathbf{r}, t)\tilde{v}_j(\mathbf{r}_1, t_1)\rangle = \lambda\langle \tilde{u}_i(\mathbf{r}, t)\tilde{v}_j(\mathbf{r}_1, t_1)\rangle - \frac{\partial}{\partial r_k}\langle \tilde{u}_k(\mathbf{r}, t)\tilde{v}_i(\mathbf{r}, t)\tilde{v}_j(\mathbf{r}_1, t_1)\rangle - \left\langle \frac{\partial \tilde{u}_i(\mathbf{r}, t)}{\partial r_k}\tilde{v}_k(\mathbf{r}, t)\tilde{v}_j(\mathbf{r}_1, t_1)\right\rangle.$$

By Furutsu–Novikov theorem (64) and expression (65) for the variational derivative, the problem for the function $\langle \tilde{v}_i(\mathbf{r}, t)\tilde{v}_j(\mathbf{r}_1, t_1)\rangle$ is formulated as

$$\left(\frac{\partial}{\partial \tau} + \lambda\right)\langle \tilde{v}_i(\mathbf{r}, t + \tau)\tilde{v}_j(\mathbf{r}_1, t)\rangle = \lambda^2 e^{\lambda\tau}\int_\tau^\infty d\tau_1 B_{ij}^{(\mathbf{u})}(\mathbf{r} - \mathbf{r}_1 + \mathbf{v}\tau_1, \tau_1)e^{-\lambda\tau_1}, \tag{66}$$

$$\langle \tilde{v}_i(\mathbf{r}, t + \tau)\tilde{v}_j(\mathbf{r}_1, t)\rangle_{\tau = 0} = \langle \tilde{v}_i(\mathbf{r}, t)\tilde{v}_j(\mathbf{r}_1, t)\rangle,$$

with $\tau = t - t_1$, where the initial condition $\langle \tilde{v}_i(\mathbf{r}, t)\tilde{v}_j(\mathbf{r}_1, t)\rangle$ is independent of t for a stationary field. The terms of order $\sigma_{\mathbf{u}}^4$ are neglected in Eq. (66), because λ is assumed to satisfy conditions (15).

Now, the diffusivity tensor $D_{ij}(\mathbf{v})$ can be calculated by using (58). The integral of Eq. (66) with respect to τ over the interval $(0, \infty)$ yields

$$\lambda\int_0^\infty d\tau\langle \tilde{v}_i(\mathbf{r}, t + \tau)\tilde{v}_j(\mathbf{r}_1, t)\rangle = \langle \tilde{v}_i(\mathbf{r}, t)\tilde{v}_j(\mathbf{r}_1, t)\rangle + \lambda\int_0^\infty d\tau B_{ij}^{(\mathbf{u})}(\mathbf{r} - \mathbf{r}_1 + \mathbf{v}\tau, \tau)[1 - e^{-\lambda\tau}]. \tag{67}$$

The expression for $D_{ij}(\mathbf{v})$ is obtained by setting $\mathbf{r} = \mathbf{r}_1$:

$$\lambda D_{ij}(\mathbf{v}) = \langle \tilde{v}_i(\mathbf{r}, t)\tilde{v}_j(\mathbf{r}, t)\rangle + \lambda\int_0^\infty d\tau B_{ij}^{(\mathbf{u})}(\mathbf{v}\tau, \tau)[1 - e^{-\lambda\tau}]. \quad (68)$$

Furthermore, Eq. (63) entails an equation for the equal-time spatial correlation tensor for the field $\tilde{\mathbf{v}}(\mathbf{r}, t)$:

$$\left(\frac{\partial}{\partial t} + 2\lambda\right)\langle \tilde{v}_i(\mathbf{r}, t)\tilde{v}_j(\mathbf{r}_1, t)\rangle$$
$$= -\frac{\partial}{\partial r_k}\langle \tilde{v}_k(\mathbf{r}, t)\tilde{v}_i(\mathbf{r}, t)\tilde{v}_j(\mathbf{r}_1, t)\rangle$$
$$-\frac{\partial}{\partial r_{1k}}\langle \tilde{u}_k(\mathbf{r}_1, t)\tilde{v}_i(\mathbf{r}, t)\tilde{v}_j(\mathbf{r}_1, t)\rangle$$
$$-\left\langle \frac{\partial \tilde{u}_i(\mathbf{r}, t)}{\partial r_k}\tilde{v}_k(\mathbf{r}, t)\tilde{v}_j(\mathbf{r}_1, t)\right\rangle$$
$$-\left\langle \frac{\partial \tilde{u}_j(\mathbf{r}_1, t)}{\partial r_{1k}}\tilde{v}_k(\mathbf{r}_1, t)\tilde{v}_i(\mathbf{r}, t)\right\rangle$$
$$+ \lambda[\langle \tilde{u}_i(\mathbf{r}, t)\tilde{v}_j(\mathbf{r}_1, t)\rangle + \langle \tilde{u}_j(\mathbf{r}_1, t)\tilde{v}_i(\mathbf{r}, t)\rangle].$$

By Furutsu–Novikov theorem (64) and expression (65) for the variational derivative, the time-independent correlation tensor $F_{ij}(\mathbf{r} - \mathbf{r}_1) = \langle \tilde{v}_i(\mathbf{r}, t)\tilde{v}_j(\mathbf{r}_1, t)\rangle$ satisfies the following equation (in the limit of $\mathbf{r} - \mathbf{r}_1 \longrightarrow \mathbf{r}$):

$$2\lambda_i F_{ij}(\mathbf{r}) = 2\int_0^\infty d\tau[B_{\beta\gamma}^{(\mathbf{u})}(\mathbf{v}\tau, \tau) - B_{\beta\gamma}^{(\mathbf{u})}(\mathbf{r} + \mathbf{v}\tau, \tau)]$$
$$\times\frac{\partial^2}{\partial r_\beta \partial r_\gamma}F_{ij}(\mathbf{r}) - \int_0^\infty d\tau\frac{\partial B_{\beta j}^{(\mathbf{u})}(\mathbf{r} + \mathbf{v}\tau, \tau)}{\partial r_\gamma}\frac{\partial}{\partial r_\beta}F_{i\gamma}(\mathbf{r})$$
$$-\int_0^\infty d\tau\frac{\partial B_{\beta i}^{(\mathbf{u})}(\mathbf{r} + \mathbf{v}\tau, \tau)}{\partial r_\gamma}\frac{\partial}{\partial r_\beta}F_{\gamma j}(\mathbf{r})$$
$$-\int_0^\infty d\tau\frac{\partial B_{i\gamma}^{(\mathbf{u})}(\mathbf{r} + \mathbf{v}\tau, \tau)}{\partial r_\beta}\frac{\partial}{\partial r_\gamma}F_{\beta j}(\mathbf{r}) \quad (69)$$
$$-\int_0^\infty d\tau\frac{\partial B_{j\gamma}^{(\mathbf{u})}(\mathbf{r} + \mathbf{v}\tau, \tau)}{\partial r_\beta}\frac{\partial}{\partial r_\gamma}F_{i\beta}(\mathbf{r})$$
$$-2\int_0^\infty d\tau\frac{\partial^2 B_{ij}^{(\mathbf{u})}(\mathbf{r} + \mathbf{v}\tau, \tau)}{\partial r_\beta \partial r_\gamma}F_{\beta\gamma}(\mathbf{r})$$
$$+ 2\lambda^2\int_0^\infty d\tau e^{-\lambda\tau}B_{ij}^{(\mathbf{u})}(\mathbf{r} + \mathbf{v}\tau, \tau).$$

The time-independent correlation $\langle \tilde{v}_i(\mathbf{r}, t)\tilde{v}_j(\mathbf{r}, t)\rangle$ is obtained by setting $\mathbf{r} = 0$ and neglecting the terms of order $\sigma_{\mathbf{u}}^4$ in Eq. (69):

$$\langle \tilde{v}_i(\mathbf{r}, t)\tilde{v}_j(\mathbf{r}, t)\rangle = \lambda\int_0^\infty d\tau e^{-\lambda\tau}B_{ij}^{(\mathbf{u})}(\mathbf{v}\tau, \tau). \quad (70)$$

The spatial diffusivity tensor $D_{ij}(\mathbf{v})$ is obtained by substituting (70) into (68):

$$D_{ij}(\mathbf{v}) = \int_0^\infty d\tau B_{ij}^{(\mathbf{u})}(\mathbf{v}\tau, \tau). \quad (71)$$

This expression is independent of λ and obviously identical to that corresponding to inertialess tracers.

The drift coefficient given by the second expression in (58) is calculated in a similar manner. It is obvious that the result will be on the order of $\sigma_{\mathbf{u}}^4$; i.e., the drift is slow as compared to sedimentation with velocity $\mathbf{v}$.

3.4. Spatiotemporal Correlation Tensor for $\mathrm{div}\tilde{\mathbf{v}}(\mathbf{r}, t)$

The ρ-space diffusivity $D_\rho(\mathbf{v})$ is obtained by differentiating (67) with respect to r_i and r_{1j} and setting $\mathbf{r} = \mathbf{r}_1$:

$$D_\rho(\mathbf{v}) = \int_0^\infty d\tau\left\langle \frac{\partial\tilde{\mathbf{v}}(\mathbf{r}, t + \tau)}{\partial\mathbf{r}}\frac{\partial\tilde{\mathbf{v}}(\mathbf{r}, t)}{\partial\mathbf{r}}\right\rangle = \frac{1}{\lambda}\left\langle\left(\frac{\partial\tilde{\mathbf{v}}(\mathbf{r}, t)}{\partial\mathbf{r}}\right)^2\right\rangle. \quad (72)$$

Hence, the correlation time for the field $\partial\tilde{\mathbf{v}}(\mathbf{r}, t)/\partial\mathbf{r}$ is

$$\tau_{\mathrm{div}\tilde{\mathbf{v}}} = \frac{1}{\lambda}.$$

This expression is valid for sufficiently large values of λ. In particular, it holds under condition (15), i.e., when $\mathbf{v}(\mathbf{r}, t) \approx \mathbf{u}(\mathbf{r}, t)$ and $\tau_{\tilde{\mathbf{v}}} = \tau_0$.

The equation for

$$\left\langle \frac{\partial\tilde{v}_i(\mathbf{r}, t)}{\partial r_k}\frac{\partial\tilde{v}_j(\mathbf{r}, t)}{\partial r_l}\right\rangle = -\frac{\partial^2 F_{ij}(\mathbf{r})}{\partial r_k \partial r_l}\bigg|_{\mathbf{r}=0}$$

follows from (69):

$$2\lambda\frac{\partial^2 F_{ij}(0)}{\partial r_k \partial r_l} = 2\lambda^2\int_0^\infty d\tau e^{-\lambda\tau}\frac{\partial^2 B_{ij}^{(\mathbf{u})}(\mathbf{v}\tau,\tau)}{\partial r_k \partial r_l}$$
$$-2\int_0^\infty d\tau\frac{\partial^2 B_{\beta\gamma}^{(\mathbf{u})}(\mathbf{v}\tau,\tau)}{\partial r_k \partial r_l}\frac{\partial^2 F_{ij}(0)}{\partial r_\beta \partial r_\gamma}$$
$$-\int_0^\infty d\tau\frac{\partial^2 B_{\beta j}^{(\mathbf{u})}(\mathbf{v}\tau,\tau)}{\partial r_\gamma \partial r_k}\frac{\partial^2 F_{i\gamma}(0)}{\partial r_\beta \partial r_l}$$
$$-\int_0^\infty d\tau\frac{\partial^2 B_{\beta j}^{(\mathbf{u})}(\mathbf{v}\tau,\tau)}{\partial r_\gamma \partial r_l}\frac{\partial^2 F_{i\gamma}(0)}{\partial r_\beta \partial r_k}$$
$$-\int_0^\infty d\tau\frac{\partial^2 B_{\beta i}^{(\mathbf{u})}(\mathbf{v}\tau,\tau)}{\partial r_\gamma \partial r_k}\frac{\partial^2 F_{\gamma j}(0)}{\partial r_\beta \partial r_l}$$
$$-\int_0^\infty d\tau\frac{\partial^2 B_{\beta i}^{(\mathbf{u})}(\mathbf{v}\tau,\tau)}{\partial r_\gamma \partial r_l}\frac{\partial^2 F_{\gamma j}(0)}{\partial r_\beta \partial r_k}$$
$$-\int_0^\infty d\tau\frac{\partial^2 B_{i\gamma}^{(\mathbf{u})}(\mathbf{v}\tau,\tau)}{\partial r_\beta \partial r_k}\frac{\partial^2 F_{\beta j}(0)}{\partial r_\gamma \partial r_l} \quad (73)$$
$$-\int_0^\infty d\tau\frac{\partial^2 B_{i\gamma}^{(\mathbf{u})}(\mathbf{v}\tau,\tau)}{\partial r_\beta \partial r_l}\frac{\partial^2 F_{\beta j}(0)}{\partial r_\gamma \partial r_k}$$
$$-\int_0^\infty d\tau\frac{\partial^2 B_{j\gamma}^{(\mathbf{u})}(\mathbf{v}\tau,\tau)}{\partial r_\beta \partial r_k}\frac{\partial^2 F_{i\beta}(0)}{\partial r_\gamma \partial r_l}$$
$$-\int_0^\infty d\tau\frac{\partial^2 B_{j\gamma}^{(\mathbf{u})}(\mathbf{v}\tau,\tau)}{\partial r_\beta \partial r_l}\frac{\partial^2 F_{i\beta}(0)}{\partial r_\gamma \partial r_k}$$
$$-2\int_0^\infty d\tau\frac{\partial^2 B_{ij}^{(\mathbf{u})}(\mathbf{v}\tau,\tau)}{\partial r_\beta \partial r_\gamma}\frac{\partial^2 F_{\beta\gamma}(0)}{\partial r_k \partial r_l}$$
$$-2\int_0^\infty d\tau\frac{\partial^4 B_{ij}(0,\tau)}{\partial r_\beta \partial r_\gamma \partial r_k \partial r_l}F_{\beta\gamma}(0).$$

Setting $i = k$ and $j = l$ in (73) yields

$$\lambda\left\langle\left(\frac{\partial\tilde{\mathbf{v}}(\mathbf{r},t)}{\partial\mathbf{r}}\right)^2\right\rangle = 4\int_0^\infty d\tau\frac{\partial^2 B_{\beta\gamma}^{(\mathbf{u})}(\mathbf{v}\tau,\tau)}{\partial r_i \partial r_j}\frac{\partial^2 F_{ij}(0)}{\partial r_\beta \partial r_\gamma} \quad (74)$$

with

$$\frac{\partial^2 F_{ij}(0)}{\partial r_\beta \partial r_\gamma} = \lambda\int_0^\infty d\tau e^{-\lambda\tau}\frac{\partial^2 B_{ij}^{(\mathbf{u})}(\mathbf{v}\tau,\tau)}{\partial r_\beta \partial r_\gamma},$$

which is given by (61) under conditions (15):

$$\frac{\partial^2 F_{ij}(0)}{\partial r_\beta \partial r_\gamma} = \frac{\partial^2 B_{ij}^{(\mathbf{u})}(0,0)}{\partial r_\beta \partial r_\gamma}$$
$$= \frac{D}{d(d+2)}[(d+1)\delta_{\beta\gamma}\delta_{ij} - \delta_{\beta i}\delta_{\gamma j} - \delta_{\beta j}\delta_{\gamma i}],$$

where the parameter D associated with the solenoidal component of the field $\mathbf{u}(\mathbf{r}, t)$ is given by (62).

The resulting expression for $D_\rho(\mathbf{v})$ is

$$D_\rho(\mathbf{v}) = \frac{4D(d+1)}{d(d+2)\lambda^2}\int_0^\infty d\tau\frac{\partial^2 B_{\beta\beta}^{(\mathbf{u})}(\mathbf{v}\tau,\tau)}{\partial\mathbf{r}^2}$$
$$= \frac{4(d+1)}{d(d+2)(d-1)\lambda^2}\frac{\partial^2 B_{\alpha\alpha}^{(\mathbf{u})}(0,0)}{\partial\mathbf{r}^2}\int_0^\infty d\tau\frac{\partial^2 B_{\beta\beta}^{(\mathbf{u})}(\mathbf{v}\tau,\tau)}{\partial\mathbf{r}^2}; \quad (75)$$

i.e., $D_\rho(\mathbf{v}) \sim \sigma_{\mathbf{u}}^4$. The solenoidal component of $\mathbf{u}(\mathbf{r}, t)$ generates the solenoidal component of $\mathbf{v}(\mathbf{r}, t)$ directly via a linear mechanism unrelated to advection, and the solenoidal component of $\mathbf{v}(\mathbf{r}, t)$ is coupled to its divergent component via advection. Thus, a smaller $D_\rho(\mathbf{v})$ is obtained if sedimentation is taken into account; i.e., the clustering time increases.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, project nos. 04-05-64044 and 02-05-64375.

REFERENCES

1. V. I. Klyatskin and T. Elperin, Zh. Éksp. Teor. Fiz. **122**, 327 (2002) [JETP **95**, 282 (2002)].
2. V. I. Klyatskin, Usp. Fiz. Nauk **173**, 689 (2003) [Phys. Usp. **46**, 667 (2003)].
3. M. R. Maxey, J. Fluid Mech. **174**, 441 (1987).
4. H. Lamb, *Hydrodynamics*, 6th ed. (Dover, New York, 1932; Gostekhizdat, Moscow, 1947).
5. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 6: *Fluid Mechanics* (Nauka, Moscow, 1986; Pergamon, New York, 1987).
6. A. S. Monin and A. M. Yaglom, *Statistical Fluid Mechanics* (Nauka, Moscow, 1965 and 1967; MIT Press, Cambridge, 1971 and 1975), Parts 1 and 2.
7. V. I. Klyatskin, *Stochastic Equations by the Eyes of a Physicist (Principal Ideas, Precise Results, and Asymp-*

totic Approximations) (Fizmatlit, Moscow, 2001) [in Russian].

8. V. I. Klyatskin, *Dynamics of Stochastic Systems: Course of Lectures* (Fizmatlit, Moscow, 2002) [in Russian].
9. V. P. Dokuchaev, Izv. Ross. Akad. Nauk, Fiz. Atmos. Okeana **31**, 275 (1995).
10. V. I. Klyatskin and O. G. Nalbandyan, Izv. Ross. Akad. Nauk, Fiz. Atmos. Okeana **33**, 291 (1997) [Izv., Atmos. Ocean. Phys. **33**, 263 (1997)].
11. E. Z. Gribova, I. S. Zhukova, S. A. Lapinova, *et al.*, Zh. Éksp. Teor. Fiz. **123**, 543 (2003) [JETP **96**, 480 (2003)].
12. V. I. Klyatskin and D. Gurariĭ, Usp. Fiz. Nauk **169**, 171 (1999) [Phys. Usp. **42**, 165 (1999)].
13. V. I. Klyatskin, Izv. Ross. Akad. Nauk, Fiz. Atmos. Okeana **36**, 177 (2000) [Izv., Atmos. Ocean. Phys. **36**, 162 (2000)].
14. V. I. Klyatskin and A. I. Saichev, Zh. Éksp. Teor. Fiz. **111**, 1297 (1997) [JETP **84**, 716 (1997)].
15. J. R. Cressman and W. I. Goldburg, J. Stat. Phys. **113**, 875 (2003).
16. J. R. Cressman, W. I. Goldburg, and J. Schumacher, Europhys. Lett. **66**, 219 (2004).
17. K. Furutsu, J. Res. Natl. Bur. Stand., Sect. D **67**, 303 (1963).
18. E. A. Novikov, Zh. Éksp. Teor. Fiz. **47**, 1919 (1964) [Sov. Phys. JETP **20**, 1290 (1964)].
19. G. K. Batchelor, *The Theory of Homogeneous Turbulence* (Cambridge Univ. Press, Cambridge, 1953; Inostrannaya Literatura, Moscow, 1955).

Translated by A. Betev

Journal of Experimental and Theoretical Physics, Vol. 99, No. 5, 2004, pp. 1018–1027.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 5, 2004, pp. 1167–1179.

FLUIDS

Chaotic Scattering, Transport, and Fractals in a Simple Hydrodynamic Flow

M. V. Budyansky, M. Yu. Uleysky, and S. V. Prants

Il'ichev Pacific Oceanological Institute, Far East Division, Russian Academy of Sciences, Vladivostok, 690041 Russia

e-mail: prants@poi.dvo.ru

Received May 18, 2004

Abstract—Advection of passive tracers in an unsteady hydrodynamic flow consisting of a background stream and a vortex is analyzed as an example of chaotic particle scattering and transport. A numerical analysis reveals a nonattracting chaotic invariant set Λ that determines the scattering and trapping of particles from the incoming flow. The set has a hyperbolic component consisting of unstable periodic and aperiodic orbits and a nonhyperbolic component represented by marginally unstable orbits in the particle-trapping regions in the neighborhoods of the boundaries of outer invariant tori. The geometry and topology of chaotic scattering are examined. It is shown that both the trapping time for particles in the mixing region and the number of times their trajectories wind around the vortex have hierarchical fractal structure as functions of the initial particle coordinates. The hierarchy is found to have certain properties due to an infinite number of intersections of the stable manifold in Λ with a material line consisting of particles from the incoming flow. Scattering functions are singular on a Cantor set of initial conditions, and this property must manifest itself by strong fluctuations of quantities measured in experiments. *© 2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Deterministic chaos in classical Hamiltonian dynamical systems having compact phase spaces is a well-developed area of nonlinear physics. A variety of analytical, numerical, and experimental methods are used to study nonlinear dynamics. Most scattering theories deal with open systems having unbounded phase spaces in which particles enter an interaction region along completely regular trajectories and escape from it along asymptotically regular particle trajectories. Thus, typical scattering trajectories are not chaotic in rigorous sense, because chaos is interpreted as complicated motion over infinite time intervals. However, motion in the interaction region may have all characteristics of true Hamiltonian chaos, including homoclinic structures, fractals, strange invariant sets, positive Lyapunov exponents, etc. This phenomenon, known as irregular (chaotic) scattering [1–4], is the subject of intensive studies in celestial mechanics [5, 6], molecular dynamics [7], atomic physics [8–10], fluid dynamics [11–18], theory of potential scattering [1, 3, 19], and other areas of physics. One can say that regular scattering is as scarce in the realm of scattering processes as are integrable systems among the totality of bounded Hamiltonian systems.

In numerous studies, including those mentioned above, it was found that chaotic scattering obeys certain fundamental laws. Transient chaos in the interaction region is due to the existence of at least one nonattracting chaotic invariant set consisting of an infinite number of localized unstable periodic and aperiodic orbits. The chaotic invariant set Λ has both stable and unstable manifolds, which extend into the regions of asymptotically regular motion. Even though Λ is not a global attractor, it plays an important role in scattering loosely analogous to that played by the scattering matrix. The particles that belong to the stable manifold Λ_s from the outset remain in the interaction region forever. Particles that are initially close to Λ_s remain in the interaction region for a long time, wandering in the neighborhoods of unstable orbits in Λ. As a result, the physically measured scattering functions that relate characteristics of outgoing trajectories to those of incoming ones are singular on a Cantor set of initial conditions. The existence of this set can be used as a basis for a definition of chaotic scattering. Even though these singularities constitute a set of measure zero, they manifest themselves in experiments by strong fluctuations of measured quantities.

A number of rigorous mathematical results have been obtained for hyperbolic scattering systems, i.e., in cases when there are no KAM tori in the phase space and all periodic orbits are unstable. For such systems, the probability of particle trapping in the interaction region is an exponentially decreasing function of trapping time, there exist simple relations between the fractal dimension of Λ and both the corresponding exponent and the average positive Lyapunov exponent (see [20, 4]), and the set of singularities of a scattering function is a simple fractal. In nonhyperbolic systems, KAM tori coexist with zones of chaotic motion, the phase-space topology strongly depends on the system's parameters, the long-time tail of the trapping probabil-

1063-7761/04/9905-1018$26.00

ity follows a power law, and the set of singularities is not a simple fractal. More or less realistic models belong to the class of nonhyperbolic systems, and analysis of chaotic scattering and transport in such models is of practical importance.

In this paper, we analyze chaotic advection in an unconfined unsteady hydrodynamic flow as a model of oceanic vortices located over topographic heights [21, 22]. The laboratory prototype of such a topographic vortex is the cylindrical vortex in a homogeneous hydrodynamic flow over an underwater obstacle known as the Taylor column [23, 24]. Passive tracers (e.g., pollutants) are advected by flow with a periodic component (tidal flow) into a mixing region, where their motion may be chaotic [18], and then are washed away into a region of regular outflow. Thus, tracer transport and mixing must be described in the framework of a typical problem of chaotic scattering. Since the phase space of a two-dimensional incompressible flow is identical to its configuration space, geophysical flows and laboratory dyeing experiments offer unique opportunities to observe directly spatial patterns that illustrate fundamental structures and properties of dynamical systems, such as invariant sets, fractal boundaries, Lévy flights, dynamical traps, etc. [16, 17, 25, 26]. In the present study, we determine the flow topology and describe tracer transport by geometrical methods.

2. INVARIANT SETS OF THE FLOW

2.1. Two-Dimensional Unsteady Flow with a Topographic Vortex

The two-dimensional flow model introduced in [27] is defined by the dimensionless streamfunction

$$\Psi = \ln\sqrt{x^2 + y^2} + \varepsilon x + \xi x \sin\tau, \tag{1}$$

where $\tau = \omega t$ is a normalized time variable, and x and y denote Cartesian coordinates. The first term in (1) represents a steady point vortex with a singular point at $x = y = 0$. The second and third terms describe steady and unsteady two-dimensional flows characterized by dimensionless parameters ε and ξ, respectively. This is a simplified model of an oceanic flow with a topographic vortex embedded in a background steady flow having a periodic tidal component. We use the Lagrangian approach to analyze the kinematics of tracers (i.e., inertialess nondiffusive particles). It is well known that the Hamiltonian equations of motion for tracers in an incompressible two-dimensional flow (with $\mathrm{div}\,\mathbf{v} = 0$, where $\mathbf{v} = (v_x, v_y)$) are written as

$$\begin{aligned} \dot{x} &= v_x(x, y, \tau) = -\frac{\partial\Psi}{\partial y}, \\ \dot{y} &= v_y(x, y, \tau) = \frac{\partial\Psi}{\partial x}, \end{aligned} \tag{2}$$

where the streamfunction $\Psi(x, y, \tau)$ plays the role of a Hamiltonian. Thus, the configuration space of an advected particle is the phase space of dynamical system (2). The equations of motion for streamfunction (1) are

$$\begin{aligned} \dot{x} &= -\frac{y}{x^2 + y^2}, \\ \dot{y} &= \frac{x}{x^2 + y^2} + \varepsilon + \xi\sin\tau, \end{aligned} \tag{3}$$

where the dot denotes derivative with respect to τ. In the absence of perturbation ($\xi = 0$), the phase portrait of the system consists of finite and infinite orbits separated by a separatrix encompassing the vortex and passing through the saddle point with the coordinates $x = -1/\varepsilon$ and $y = 0$. In the polar coordinates defined by the relations $x = \rho\cos\varphi$ and $y = \rho\sin\varphi$, the unperturbed equations are solved in quadratures:

$$\dot{\rho} = \varepsilon\sqrt{1 - \left(\frac{E - \ln\rho}{\varepsilon\rho}\right)^2}, \tag{4}$$

where $E = \varepsilon\rho\cos\varphi + \ln\rho$ is an integral of motion. Depending on initial conditions, particles move either along closed streamlines encompassed by the separatrix loop or around the loop along infinite streamlines. In [27], it was shown numerically and analytically that an arbitrarily small perturbation splits the separatrix and gives rise to transversal intersections of stable and unstable saddle-point manifolds and to an infinite variety of periodic and aperiodic orbits. The trajectories of tracers deviate from the steady-flow streamlines. Some information about them is provided by Poincaré sections, which depict the locations of particles in the (x, y) plane at $\tau = 2\pi m$ ($m = 0, 1, 2, \ldots$). Borrowing terminology from [14], we define the free-stream region (with incoming and outgoing components), mixing region, and vortex core as the sets of trajectories for which the number of times they wind around the vortex is zero, finite, and infinite, respectively.

The phase-space topology strongly depends on the values of ε and ξ, because they are defined in terms of steady-flow velocity, unsteady-flow period, and vortex intensity (which determines the particle rotation frequency). Their relative values determine the orders of nonlinear resonances in the system. As the value of ε/ξ increases, the vortex core (occupied by regular trajectories) grows and the orders of surviving resonances increase, while the mixing region shrinks correspondingly. When $\varepsilon/\xi \gg 1$, the system exhibits almost regular dynamics. In that study, numerical computations were performed for $\varepsilon = 0.5$ and $\xi = 0.1$, in which case the mixing region is abundant with various topological structures.

2.2. KAM Tori and Cantori

Now, we describe the invariant sets of dynamical system (3), which make up the building blocks of its structure. Particles belonging to different sets exhibit

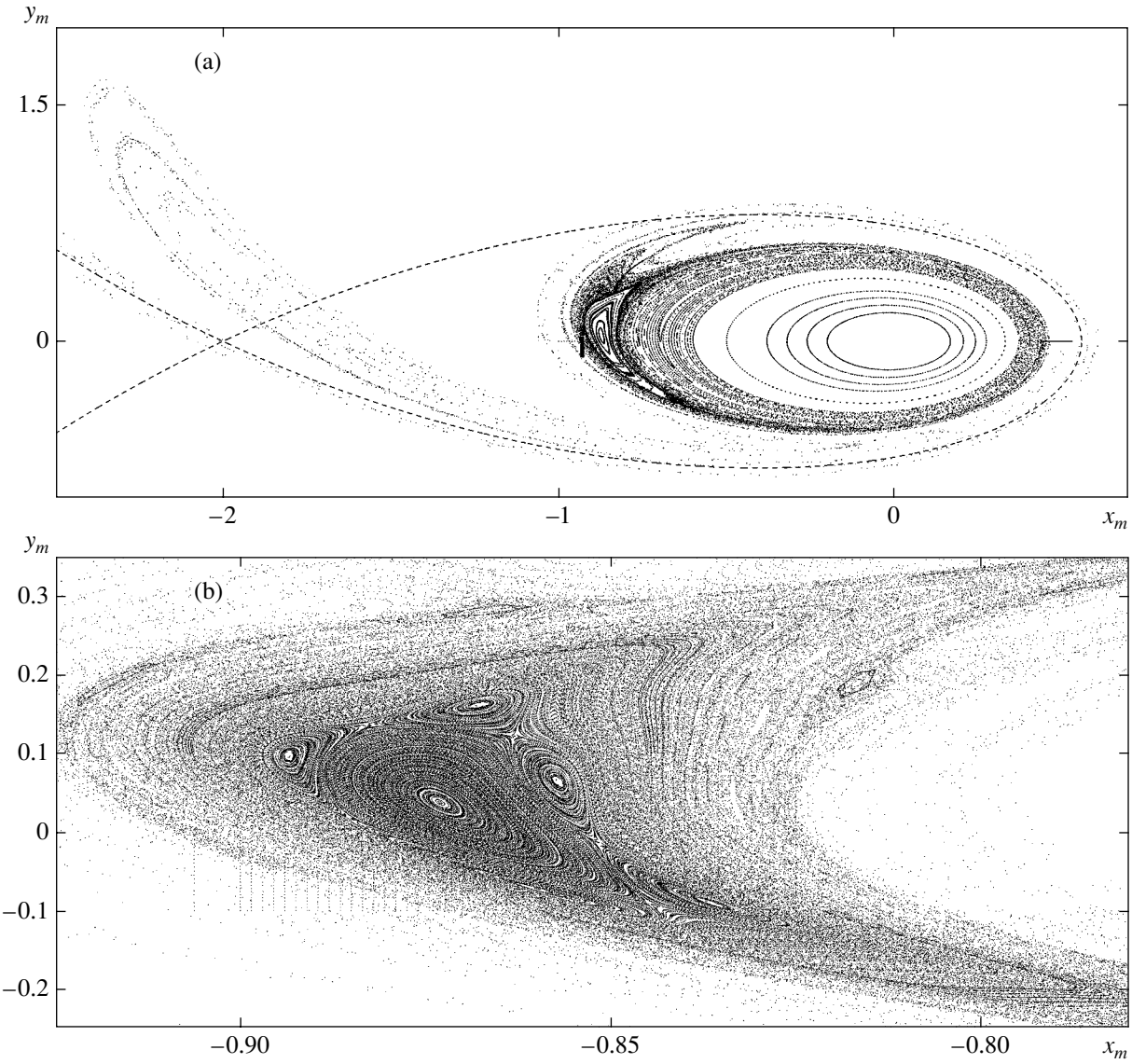


Fig. 1. (a) Plane of the Poincaré section of the flow. Dashed curves represent the unperturbed separatrix. (b) Secondary resonance with period 10π in the neighborhood of the half-integer primary resonance with period π.

qualitatively different behavior. The simplest examples of invariant sets are the entire phase space, a stationary point, a periodic orbit, and any orbit defined on the time interval $[-\infty, \infty]$. The set of trajectories that do not wind around the vortex is excluded from the present analysis. The set of invariant curves shown in Fig. 1 represents sections of KAM tori. It is the set of periodic and quasiperiodic tracer motions around the vortex center. In the Poincaré section, they make up families of nested closed smooth curves. Most of them lie inside the vortex core. The Poincaré section is the set of points with coordinates $x_m = x(\tau = 2\pi m)$ and $y_m = y(\tau = 2\pi m)$, where $m = 0, 1, 2, \ldots$. Figure 1 depicts the Poincaré section of a set of approximately 10^3 orbits with initial coordinates x and y lying in the intervals $[-0.9, -0.85]$ and $[-0.1, 0.1]$, respectively. Other families of invariant curves make up stable islands centered at elliptic points, which are located both in the vortex core and in the mixing region. The islands arise from nonlinear resonances of various orders between particle motions in the vortex and the 2π-periodic perturbation. The main island in the chaotic sea is arises from the half-integer (π-periodic) primary resonance and is surrounded by higher order resonances (see Fig. 1b). Fig. 1b also clearly depicts the secondary resonance with period 10π. Figure 2 shows a periodic orbit at the center of the primary resonance; the inset thereto, a fragment of the periodic orbit corresponding to the 10π-periodic secondary resonance depicted in Fig. 1b. Asterisks represent the elliptic points of the secondary resonance.

The vortex core also contains islands and chains of islands. High-resolution images demonstrate that these

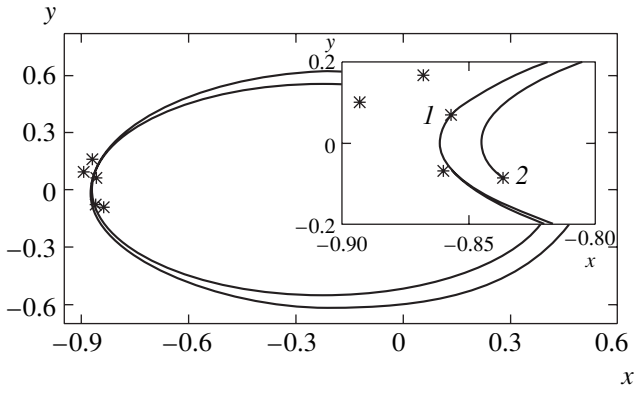


Fig. 2. Periodic trajectory at the center of the primary resonance. Inset shows a fragment of the periodic trajectory of the 10π-periodic secondary resonance originating at point *1* and terminating at point *2* over a time interval of 2π.

chains are surrounded by narrow stochastic layers. The vortex core is preserved for any combination of ε and ξ; i.e., it is a robust structure. Since the particle rotation frequency in the vortex core is much higher than the (unit) perturbation frequency, the perturbation can be treated as adiabatic with respect to most orbits inside the core and the orbits as regular, except for those lying in the neighborhoods of overlapping high-order resonances, which make up very narrow stochastic layers. The KAM tori make up impermeable barriers that limit tracer transport and mixing.

It is well known (see [28]) that perturbation gives rise to cantori replacing some KAM tori (primarily those with rotation numbers that do not satisfy the Diophantine condition in the KAM theorem). The cantori are invariant sets having Cantor structure with gaps and are characterized by a topological dimension at least lower than the measure of a curve. Motion on them is quasiperiodic. However, cantori are unstable and, therefore, have stable and unstable manifolds. Unlike KAM tori, cantori are permeable for tracers. The intersection of a material line with the stable manifold Λ_s in the incoming-flow region contains particles that reach the mixing-region–vortex-core boundary in the course of time and rotate in a region loosely encompassed by a bounded invariant KAM curve. Then, they rapidly cross the curve and dwell on its opposite side. The process repeats many times until the particles cross the boundary and escape. Such particle-trapping regions point to the existence of cantori with narrow gaps at vortex-core–stable-island boundaries. Figure 3 illustrates the trapping of a passive tracer with initial coordinates in the incoming-flow region. The graph of the coordinate x as a function of time (Fig. 3a) demonstrates that most of the time it executes a periodic motion in the neighborhood of the half-integer primary resonance (depicted in Fig. 3b). Figure 3c shows a frag-

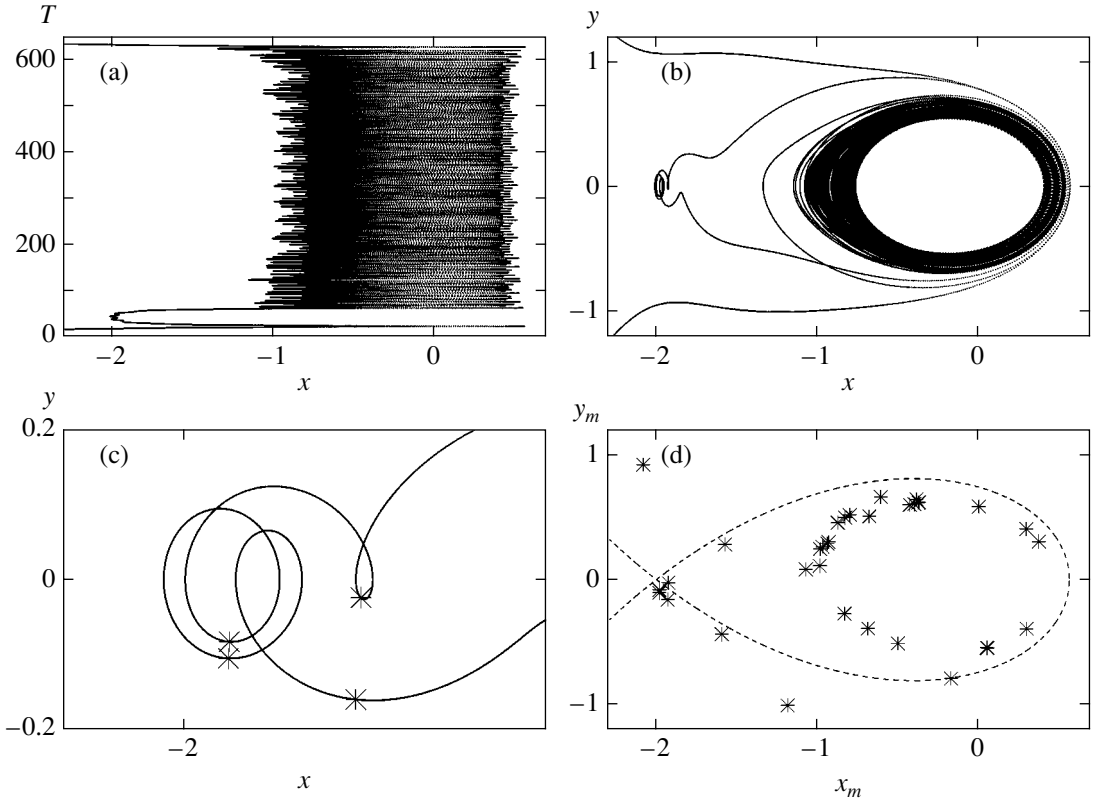


Fig. 3. Trapping of a passive tracer at the edge of the chaotic sea: (a) variation of coordinate x with time; (b) tracer trajectory; (c) fragment of trajectory near the saddle point; (d) Poincaré section.

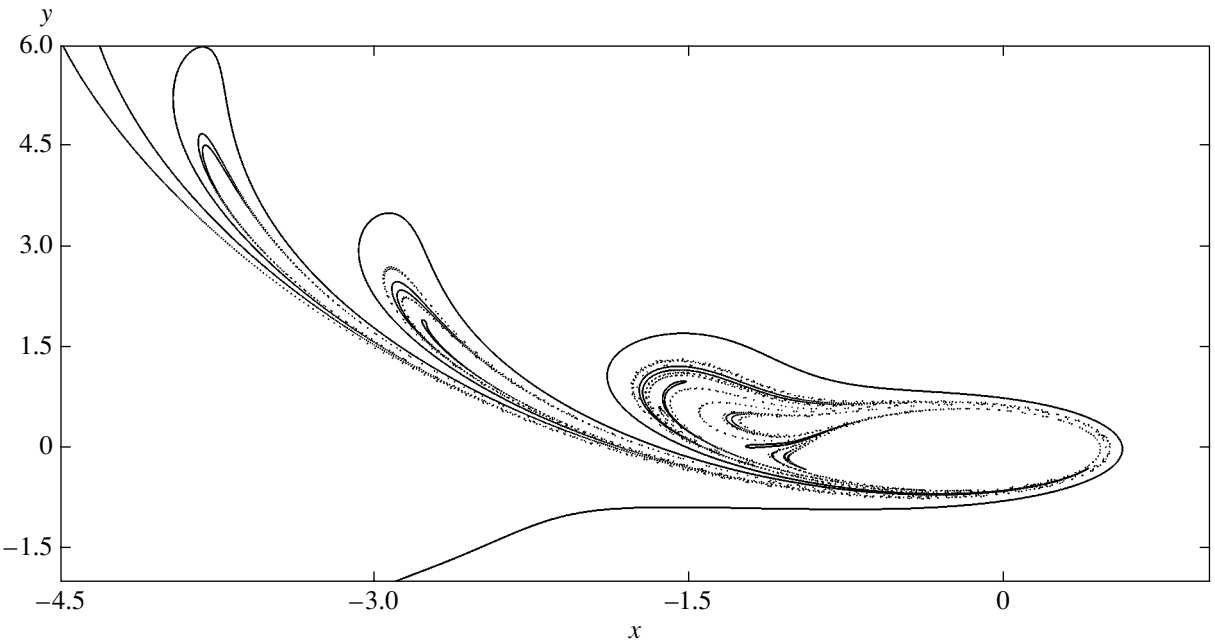


Fig. 4. Image of the unstable manifold Λ_u obtained as a "snapshot" of a dye streak in the flow at $\tau = 15\pi$.

ment of this periodic orbit in the neighborhood of the saddle point corresponding to an unstable periodic orbit. Figure 3d shows the corresponding Poincaré section. It is important that the existence of an invariant set of cantori and/or trapping regions implies that mixing is inhomogeneous, as manifested (e.g., in topographic maps of trapping) by power-law long-time tails of trapping-time distributions and by singular behavior of scattering functions [18].

2.3. Chaotic Invariant Set and Unstable Periodic Orbits

The chaotic invariant set Λ is defined as the set of all orbits (except for the KAM tori and cantori) that never leave the mixing region. The set consists of an infinite number of periodic and aperiodic (chaotic) orbits. All orbits in this set are unstable. If a tracer belongs to Λ at the initial moment, then it remains in the mixing region as $\tau \longrightarrow \infty$ or $\tau \longrightarrow -\infty$. The Poincaré section of Λ is a set of points of Lebesgue measure zero. Most trajectories of the tracers advected into the mixing region from the incoming flow sooner or later leave the mixing region with the outgoing flow. However, their behavior is largely determined by the presence of Λ. They can "trail" after trajectories of the saddle set, wandering in their neighborhoods.

Each orbit in the chaotic set and, therefore, the entire set Λ have both stable and unstable manifolds. The stable manifold Λ_s of the chaotic set is defined as the invariant set of orbits approaching those in Λ as $\tau \longrightarrow \infty$. The unstable manifold Λ_u is defined as the stable manifold corresponding to time-reversed dynamics. Following trajectories in Λ_s, tracers advected by the incoming flow enter the mixing region and remain there forever. It was mentioned above that the corresponding initial conditions make up a set of measure zero. The tracer trajectories that are initially close to those in the chaotic set follow the chaotic-set trajectories for a long time and eventually deviate from them, and leave the mixing region along the unstable manifold. This behavior offers a unique opportunity to extract important properties of Λ by measuring the characteristics of scattered particles and to observe unstable manifolds directly in laboratory experiments [11, 12, 16, 17, 29] or even in geophysical flows [30, 31].

An unstable manifold can be visualized by various methods. A blob consisting of many tracer particles initially belonging to the intersection of the incoming flow with the stable manifold spreads out and transforms into an intricate fractal curve approaching Λ_u in the course of time. A similar pattern develops in dyeing experiments. The stable manifold lies in the coordinate-plane region bounded by the separatrix locations at the times corresponding to the two extrema reached during the perturbation period. This region extends to $-\infty$ along the y axis, and its width is determined by the values of ε and ξ. Only particles located in this region reach the mixing region. Figure 4 shows an image of the unstable manifold of (3) at the time 15π obtained numerically by integrating the equations of motions for particles continuously injected into the incoming flow at the point with $x_0 = -4.357759744$ and $y_0 = -6$. This pattern oscillates with the flow. Tracer particles are advected along the fractal curve of the unstable manifold, which plays the role of an "attractor" in a Hamil-

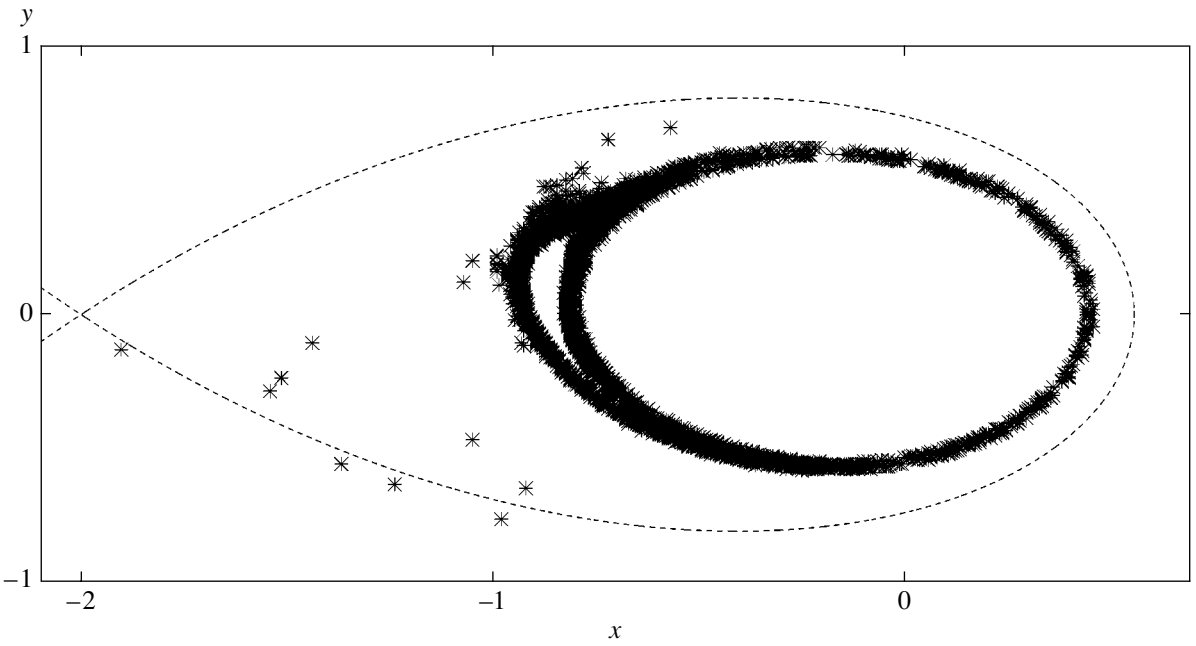


Fig. 5. Image of chaotic invariant set Λ.

tonian system (there are no "classical" attractors in incompressible flows). Direct computation of the so-called trapping map [18] provides an image of Λ_s. The intersection of Λ_s and Λ_u is the chaotic manifold Λ depicted in Fig. 5. This fractal set of points oscillates with the flow. The tracer particles starting from points in this set remain in the mixing region as $\tau \longrightarrow \pm\infty$. The vortex-core–free-stream boundary has a finite length equal to that of the separatrix loop in the unperturbed flow ($\xi = 0$), but it is infinitely long in a periodically perturbed flow.

The set shown in Fig. 5 was constructed by using the numerical procedure described above as a sketch of the actual chaotic invariant set Λ. Figure 5 provides an adequate representation of the nonhyperbolic component of Λ, which consists of marginally unstable quasiperiodic orbits in the neighborhoods of the boundaries of outer KAM tori. As mentioned above and demonstrated in Fig. 3, it is the region where the cantori and trapping regions are located and particles may remain for a long time, executing almost periodic motions with the Lyapunov exponent close to zero. Their dynamics can be interpreted as motions in the neighborhood of the half-integer primary resonance. The fractal characteristics of these motions are analyzed in the next section.

Periodic orbits in the hyperbolic component of Λ are essentially unstable, and the probability of finding them in the superposition of two fractal "dust clouds" is low. We propose the following numerical method to demonstrate their existence. Change (numerically) to new variables: angle Θ and action I. Define a set of initial angles Θ_0 and actions I_0 for particles in the mixing region. Find their values such that the angle changes by 2π in a time interval T. For the same initial conditions, calculate the change ΔE in particle energy over a certain time interval, say, 2π. The particles for which both angle and energy remain unchanged over this interval ($\Theta(T = 2\pi) = \Theta_0$ and $\Delta E = 0$) obviously execute a 2π-periodic motion. Using these results, draw a contour map of particle-energy change as a function of the initial angle and time T. Shaded areas in Fig. 6 represent certain values of ΔE. The intersections of the contours of $\Delta E = 0$ with the line $T = 2\pi = 6.2831852\ldots$ give the initial angles corresponding to particles that execute 2π-periodic motions in the initial action interval. Similar procedures can be used to find 4π-, 6π-, and $2m\pi$-periodic orbits. Their stability can be analyzed by standard methods.

3. GEOMETRY OF CHAOTIC SCATTERING

3.1. Fractal Structure of Scattering

In this section, we analyze the geometry of tracer transport in the mixing region. As an illustration, we consider the initial conditions on the segment of the line $y = -6$ in the free-stream region whose left and right endpoints are its intersections with the lower whisker of the perturbed separatrix loop at the times $3\pi/2 + 2\pi m(\sin\tau = -1)$ and $\pi/2 + 2\pi m$ $(\sin\tau = 1)$, respectively. All nontrivial scattering processes involve particles whose initial coordinates lie in the region bounded by these amplitude values of instantaneous separatrix streamlines. Figure 7 shows the snapshots of the evolution of this material line taken at $\tau = 8\pi$, 9π, 10π, and 11π. At $\tau = 0$, point A was at the intersection of the unperturbed separatrix ($\sin\tau = 0$) and the line $y_0 = -6$,

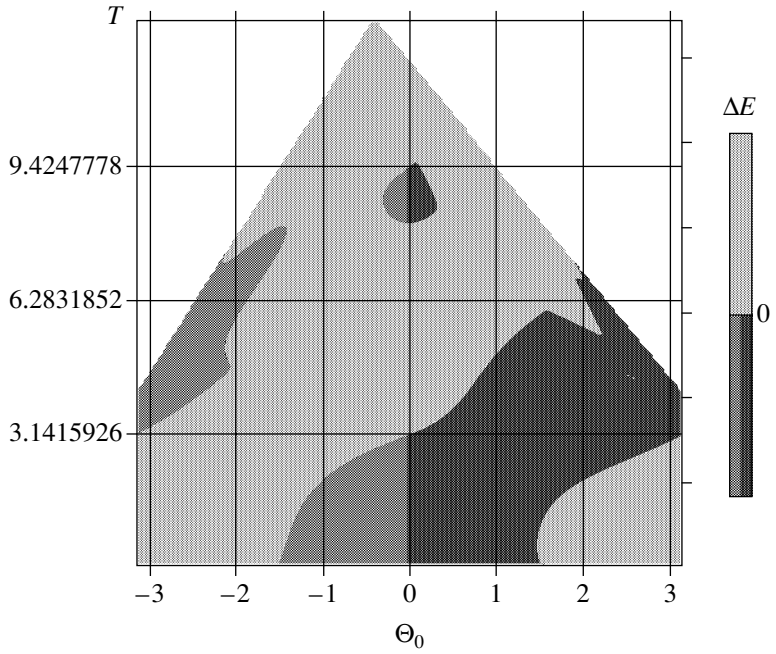


Fig. 6. Map of particle-energy change ΔE (as a function of initial angle Θ_0 and movement time T) for determining the initial coordinates of tracers executing 2π-periodic motions.

i.e., at $x_0(A) = -4.6447002$; point G, at the intersection of the separatrix at the times $\pi/2 + 2\pi m$ and the line, i.e., at $x_0(G) = -4.3577522$. The particles with $x_0 < -4.6447002$ and $x_0 > -4.3577522$ are not trapped in the vortex and are immediately washed away into the free-stream region (see dotted segments in Fig. 7). Points A and G move along the stable manifold into the neighborhood of the saddle periodic orbit and remain in the mixing region for a long (theoretically, infinite) time.

We calculated the total number n of turns executed by most particles before they escaped into the outgoing-flow region (the half-plane above the line $y = 6$). The graph of $n(x_0)$ (see Fig. 8) is an intricate hierarchy of sequences of fragments of material line AG. Their fractal properties are generated by the infinite sequence of intersections of the stable and unstable manifolds with the material line segment as it rotates about the vortex.

Following [32], we refer to the sequences of segments corresponding to each $n \geq 0$ as epistrophes. The epistrophes make up a hierarchy. The endpoints of each segment in an nth-level epistrophe are the limit points of an $(n + 1)$th-level epistrophe. For example, there is a single epistrophe on the zeroth level ($n = 0$), and the endpoints of the corresponding segments generate epistrophes b, c, d, e, g, etc. on the level $n = 1$, which converge to the corresponding limit points (see Fig. 8). Numerical experiments on epistrophes lying on different levels revealed the following trends: (i) each epistrophe converges to a limit point in the material line segment under consideration; (ii) the endpoints of each segment in an nth-level epistrophe are the limit points of an $(n + 1)$th-level epistrophe; (iii) the lengths of segments in an epistrophe decrease in geometric progression; (iv) the common ratio q of all progressions is related to the largest Lyapunov exponent for the saddle point as follows: $\lambda = -(1/2\pi)\ln q$.

Figure 9 shows the length l_j of an epistrophe segment as a functions of its index j for the zeroth-level epistrophe and the first-level epistrophes c, d, e, and g. The slopes of all graphs are $\ln q = -1.59$; i.e., the segment lengths in every epistrophe decrease in geometric progression: $l_j = l_0 q^j$ with $q \approx 0.2$. The largest Lyapunov exponent for the saddle point can be evaluated as the upper bound for the numerical error accumulated over a double integration step. Numerical experiments using 2×10^4 tracer particles uniformly distributed along AG showed that the upper bound for the numerical error accumulated before the particles escape from the mixing region lies on a line with slope $\lambda \approx 0.25$. A similar result is obtained by analyzing the linearized system (3) in the neighborhood of the saddle point.

The fractal depicted in Fig. 8 is not strictly self-similar, because it contains segments (called strophes in [32]) that do not belong to the epistrophes. Some of them are labeled by Greek letters in the graph. Thus, the fractal is characterized by partial self-similarity: each level contains both self-similar epistrophe sequences and additional elements (strophes), which are preserved in the asymptotic limit and do not fit into the regular structure. These results are in complete agreement with the epistrophe theorem proved in [32] for area- and orientation-preserving open maps of a plane with a

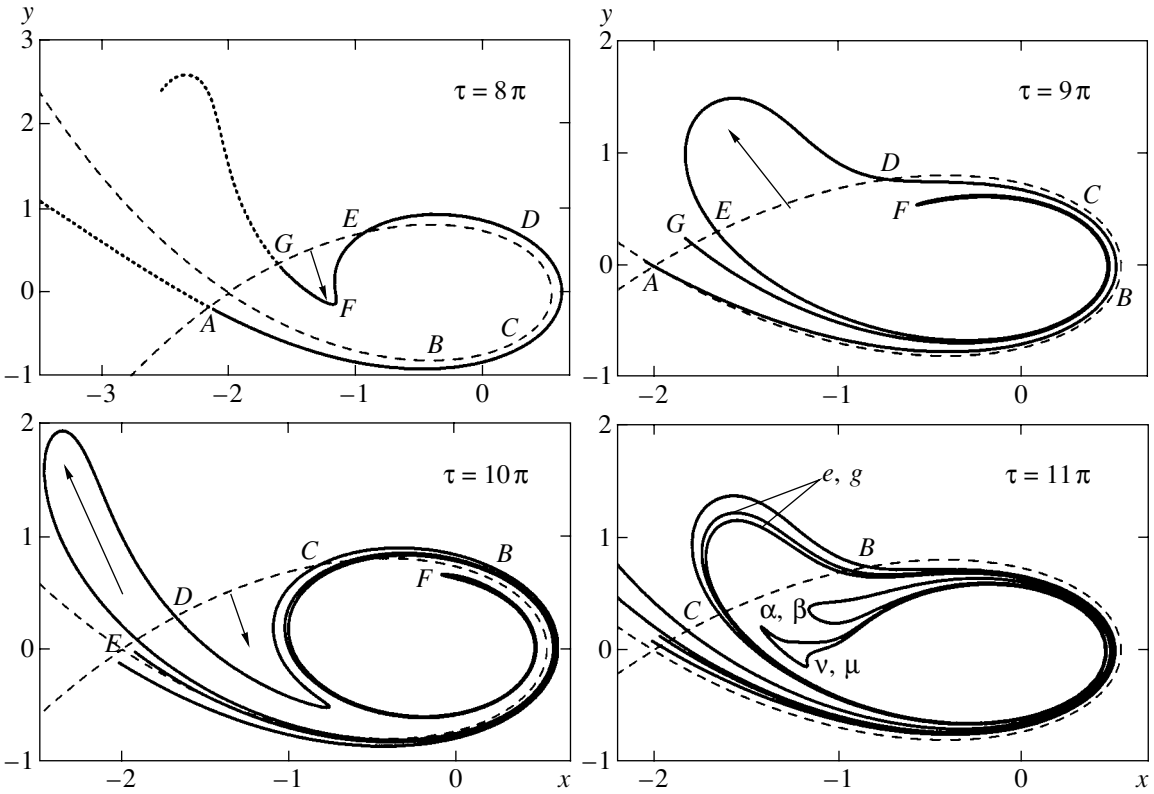


Fig. 7. Snapshots of evolution of a material line taken at successive instants to illustrate the development of "lobes" from elements of the epistrophes and strophes in the fractal shown in Fig. 8.

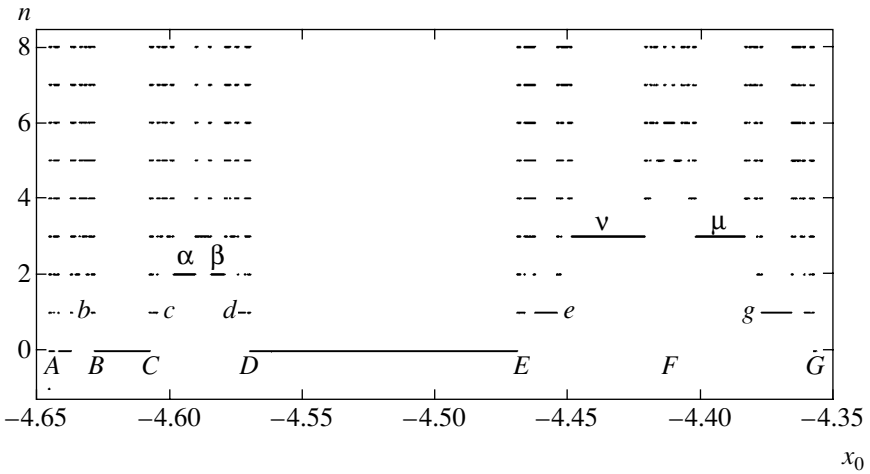


Fig. 8. Fractal set of the initial coordinates x_0 of incoming-flow tracers that escape from the mixing region after n turns around the vortex ($n \longrightarrow \infty$).

center and an unstable singular point whose stable and unstable manifolds intersect transversely.

3.2. Tracer Transport

The fractal depicted in Fig. 8 provides a comprehensive illustration of tracer transport. The line segment *AG* is stretches and bends as it winds around the point vortex, and then its part begins to fold as particles rotating around the vortex accelerate while other particles decelerate in the neighborhood of the saddle point. Figure 7 illustrates the formation of the first fold at $\tau = 8\pi$. The snapshots in Fig. 7 corresponding to later instants show that the segment *DE* (the "back" of the fold) escapes into the free-stream region, taking the shape of a "lobe." The segment *EFG* (the "front" of the fold)

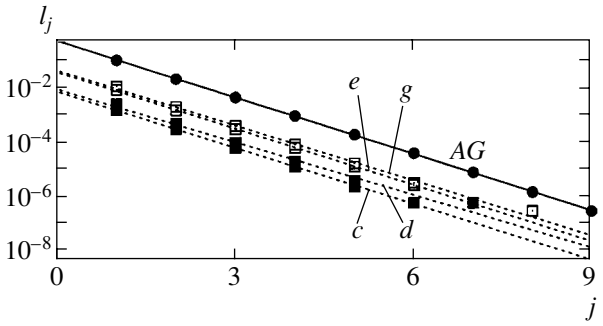


Fig. 9. Semilogarithmic plot of the decrease in the epistrophe-segment length l_j for $n = 0$ (*AG*) and $n = 1$ (*c*, *d*, *e*, *g*) with the segment index. The logarithm of the slope is minus the largest Lyapunov exponent.

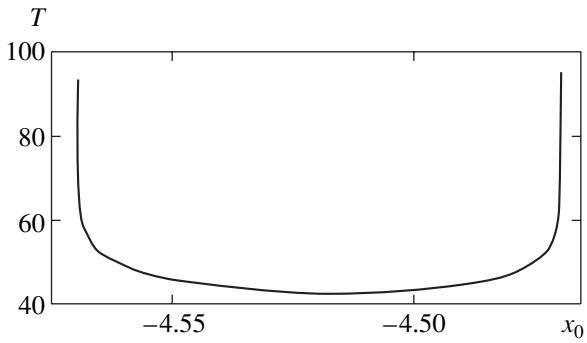


Fig. 10. Scattering function for particles belonging to the segment *DE* of the zeroth-level epistrophe (see Fig. 8). *T* is the particle-trapping time.

continues to wrap around the vortex. The segment *DE* in Fig. 8 is associated with the segment *DE* of the zeroth-level epistrophe (tracer particles that have not made a complete turn). The segment *EFG* is represented by an empty segment generating an infinite sequence of strophes and epistrophes in Fig. 8. At $\tau = 10\pi$, the second fold develops in the material line. The segment *BC* (its "back") escapes, giving rise to another "lobe" as its "front" (represented by segment *CD* in Fig. 8) gives rise to another turn about the vortex. After a time interval of $\tau = 11\pi$, two new lobes (represented by the segments *e* and *g* of the first-level epistrophe) begin to develop in the stretched portion of the first fold. The particles in these lobes escape together with the lobe *BC* before they complete their second turn about the vortex, giving rise to the second "finger" in Fig. 4. Furthermore, the snapshot taken at $\tau = 11\pi$ shows the folds that subsequently develop into the second-level strophes α and β and into the third-level strophes ν and μ. These strophes give rise to the four lobes that combine with zeroth- and first-level epistrophe segments to form the third "finger" in Fig. 4.

This process repeats iteratively; i.e., the portion corresponding to an epistrophe segment and an empty segment in Fig. 8 unwinds off the material segment's "tail" that lingers in the neighborhood of the saddle point with each turn about the vortex. The result is the infinite sequence of the zeroth-level epistrophe segments illustrated by Fig. 8. Each subsequent zeroth-level epistrophe segment escapes in a period after the preceding escape. This scenario describes the formation of epistrophes and strophes at all nonzero levels, except that each *n*th-level epistrophe segment generates two $(n + 1)$th-level epistrophes. In other words, these epistrophe segments also escape successively, a pair per period. In dyeing experiments, these events are visualized by periodic formation of lobe pairs. In the course of time, dye streaklines develop into a self-similar pattern (see Fig. 4) in the sense that new fingers with increasing number of lobes appear in each subsequent period.

Figure 10 shows a typical scattering function for particles belonging to the segment *DE* of the zeroth-level epistrophe. The endpoints *D* and *E* divide particles that fall into the stable and unstable manifolds. A similar role is played by the endpoints of all elements of strophes and epistrophes. These points make up a set of points that remain in the mixing region forever. The scattering function corresponding to strophes have a similar U shape with a more pronounced asymmetry.

4. CONCLUSIONS

A partition of the phase space into invariant sets is described for an unconfined periodic incompressible flow model with a stationary point vortex. Behavior of typical trajectories and tracer transport are analyzed. A numerical method is proposed for finding unstable periodic orbits with the use of contour maps of particle-energy change. A chaotic invariant set Λ is revealed. Its unstable manifold Λ_u is visualized in numerical experiments on evolution of material lines and multiple tracer trajectories. The scaling on the material line segment evolving from the intersection of a line in the incoming flow with the stable manifold Λ_s is shown to determine the chaotic behavior of tracers. It is demonstrated that the singularities of the trapping time for a particle in the mixing region as a function of its initial coordinate in the incoming flow are associated with particles that enter the mixing region along trajectories in Λ_s and escape along Λ_u. An analysis of the geometry and topology of chaotic scattering shows that both the particle-trapping time and the number of their turns about the vortex have intricate hierarchical fractal structure as functions of their initial coordinates. The hierarchy is shown to exhibit certain properties due to infinite number of intersections of the stable manifold in Λ with a material line consisting of particles from the incoming flow. The self-similar structure of the function is shown to consist of sequences of epistrophes, which determine tracer transport. Relationship between the topological and dynamical characteristics of chaotic advection is established.

Finally, we outline some open problems that deserve further scrutiny. The approach developed here can be applied to a broad class of problems in chaotic scattering in an inhomogeneous phase space where the dominant role is played by a nonattracting chaotic set with hyperbolic and nonhyperbolic components. For hyperbolic scattering, the fractal dimension d can be expressed in terms of the largest Lyapunov exponent λ and the average trapping time $\langle T \rangle$ for particles in the mixing region as $d \approx 1 - (\lambda/\langle T \rangle)^{-1}$ [4, 20]. When there exists a nonhyperbolic component, this expression is not valid and a deeper insight into relationship between topology, dynamics, and statistics of chaotic scattering is required. Of practical importance is analysis of the effect of noise on basic scattering characteristics. In geophysical advection problems, the effects due to viscosity (dissipative dynamics of finite-size particles) must be taken into account. Dynamics of chemically or biologically active diffusive tracers is also of practical interest [33]. Note that chaotic invariant sets play an important role in geophysical advection of active tracers, such as oceanic phytoplankton or chemical reactants in the atmosphere. These sets can be viewed as dynamical catalysts of biological productivity and chemical reactions.

NOTE ADDED IN PROOF (September 2004)

Using the method described at the end of Section 2, we found nonresonant periodic orbits (separated from elliptic and hyperbolic resonance points) and unstable periodic orbits related to broken resonances in the chaotic sea, which are not manifested in the Poincaré sections.

ACKNOWLEDGMENTS

This work was supported by Presidium of the Russian Academy of Sciences under the program "Mathematical Methods in Nonlinear Dynamics"; Russian Foundation for Basic Research, project no. 02-02-17796; and Presidium of the Far East Division of the Russian Academy of Sciences under the Program for Support of Basic Research.

REFERENCES

1. B. Eckhardt, Physica D (Amsterdam) **33**, 89 (1988).
2. R. Blumel and U. Smilansky, Phys. Rev. Lett. **60**, 477 (1988).
3. P. Gaspard and S. A. Rice, J. Chem. Phys. **90**, 2225 (1989).
4. E. Ott and T. Tel, Chaos **3**, 417 (1993).
5. J. M. Petit and M. Henon, Icarus **60**, 536 (1986).
6. P. T. Boyd and S. L. W. McMillan, Chaos **3**, 507 (1993).
7. D. W. Noid, S. Gray, and S. A. Rice, J. Chem. Phys. **84**, 2649 (1986).
8. S. V. Prants, Pis'ma Zh. Éksp. Teor. Fiz. **75**, 777 (2002) [JETP Lett. **75**, 651 (2002)].
9. V. Yu. Argonov and S. V. Prants, Zh. Éksp. Teor. Fiz. **123**, 946 (2003) [JETP **96**, 832 (2003)].
10. S. V. Prants and M. Yu. Uleysky, Phys. Lett. A **309**, 357 (2003).
11. J. Chaiken, R. Chevray, M. Tabor, and Q. M. Tan, Proc. R. Soc. London, Ser. A **408**, 165 (1986).
12. W. L. Chien, H. Rising, and J. M. Ottino, J. Fluid Mech. **170**, 355 (1986).
13. B. Eckhardt and H. Aref, Philos. Trans. R. Soc. London, Ser. A **326**, 655 (1988).
14. V. Rom-Kedar, A. Leonard, and S. Wiggins, J. Fluid Mech. **214**, 347 (1990).
15. C. Jang, T. Tel, and E. Ziemniak, Chaos **3**, 555 (1993).
16. T. N. Solomon, E. R. Weeks, and H. L. Swinney, Phys. Rev. Lett. **71**, 3975 (1993).
17. J. C. Sommerer, H.-C. Ku, and H. E. Gilreath, Phys. Rev. Lett. **77**, 5055 (1996).
18. M. V. Budyansky, M. Yu. Uleysky, and S. V. Prants, Dokl. Akad. Nauk **386**, 686 (2002).
19. B. Eckhardt and C. Jung, J. Phys. A **19**, 829 (1986).
20. H. Kautz and P. Grassberger, Physica D (Amsterdam) **17**, 75 (1985).
21. V. F. Kozlov, *Models of the Topographic Vortices in the Ocean* (Nauka, Moscow, 1983) [in Russian].
22. V. N. Zyryanov, *Topographic Vortices in Dynamics of Marine Flows* (Inst. Vodnykh Problem, Ross. Akad. Nauk, Moscow, 1995) [in Russian].
23. G. Proudman, Proc. R. Soc. London, Ser. A **92**, 408 (1916).
24. G. I. Taylor, Proc. R. Soc. London, Ser. A **104**, 213 (1923).
25. S. D. Danilov, V. A. Dovzhenko, and I. G. Yakushkin, Zh. Éksp. Teor. Fiz. **118**, 483 (2000) [JETP **91**, 423 (2000)].
26. T. Yuster and W. W. Hackborn, Chaos **7**, 769 (1997).
27. M. V. Budyansky and S. V. Prants, Pis'ma Zh. Tekh. Fiz. **27** (12), 51 (2001) [Tech. Phys. Lett. **27**, 508 (2001)].
28. V. I. Arnol'd, V. V. Kozlov, and A. I. Neĭshtadt, *Mathematical Aspects of Classical and Celestial Mechanics* (VINITI, Moscow, 1985; Springer, Berlin, 1993).
29. M. Van Dyke, *An Album of Fluid Motion* (Parabolic Press, Stanford, CA, 1982; Mir, Moscow, 1986).
30. R. T. Sutton, H. Maclean, R. Swinbank, *et al.*, J. Atmos. Sci. **51**, 2995 (1994).
31. K. L. Denman and A. E. Gargett, Annu. Rev. Fluid Mech. **27**, 225 (1995).
32. K. A. Mitchell, J. P. Handley, B. Tighe, *et al.*, Chaos **13**, 880 (2003).
33. Chaos **12** (2) (2002), Special Issue, Ed. by Z. Toroczkai and T. Tel.

Translated by A. Betev

Journal of Experimental and Theoretical Physics, Vol. 99, No. 5, 2004, pp. 1028–1038.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 5, 2004, pp. 1180–1191.

FLUIDS

Velocity, Temperature, and Reynolds-Stress Scaling in the Wall Region of Turbulent Boundary Layer on a Permeable Surface

I. I. Vigdorovich
Central Institute of Aviation Motors, Moscow, 111116 Russia
e-mail: ivigdorovich@ciam.ru
Received June 1, 2004

Abstract—A finite total number of flow parameters in the wall region of a turbulent boundary layer points to universal behavior of turbulent shear stress as a function of mean-velocity gradient and turbulent heat flux as a function of both mean-velocity and mean-temperature gradients. Combined with dimensional arguments, this fact is used to reduce the momentum and heat equations to first-order ordinary differential equations for temperature and velocity profiles amenable to general analysis. Scaling laws for velocity and temperature in boundary layer flows with transpiration are obtained as generalizations of well-known logarithmic laws. Scaling relations are also established for shear stress and rms transverse velocity fluctuation. The proposed method has substantial advantages as compared to the classical approach (which does not rely on fluid-dynamics equations [1–3]). It can be applied to establish scaling laws for a broader class of near-wall turbulence problems without invoking closure hypotheses. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The well-known derivation of the logarithmic velocity profile for the wall regions of turbulent near-wall flows proposed in [1–3] neither makes use of any hypothesis about the mechanism of turbulent transfer nor relies on fluid-dynamics equations. It is based only on dimensional reasoning combined with assumptions of negligible molecular viscosity outside the viscous sublayer and near-wall flow independence of external parameters. An analogous approach was subsequently applied in numerous studies. In particular, a logarithmic temperature distribution was obtained by Landau in 1944 (e.g., see [4]).

The scope of the classical method [1–3] is obviously restricted to problems with a limited number of parameters. In particular, it cannot be used to obtain any meaningful results in the case of nonzero transverse velocity at the wall.

The alternative approach proposed in this paper also makes use of dimensional analysis, but it essentially relies on equations of motion. As a result, velocity and temperature scaling laws for flows with transpiration are obtained as generalizations of the classical logarithmic laws. Scaling laws have also been established for shear stress and rms transverse velocity fluctuation. The present analysis is as physically reasonable and mathematical rigorous as the classical one.

The scaling law for the velocity profile was originally obtained in [5, 6] by invoking the Prandtl mixing-length theory, whereas the remaining relations are newly derived here.

2. STATEMENT OF THE PROBLEM

Consider the heat-conducting incompressible flow in the wall region of turbulent boundary layer on a smooth permeable surface with transpiration along the normal vector. Since the transverse gradients of averaged flow variables are much steeper than their longitudinal gradients in a thin wall region, the momentum and heat equations can be written in the first approximation as

$$\begin{aligned} -\langle u', v'\rangle + \nu\frac{du}{dy} &= \frac{\tau_w}{\rho} + v_w u, \\ -\langle \theta' v'\rangle + \chi\frac{d\theta}{dy} &= -j_w + v_w(\theta - \theta_w). \end{aligned} \tag{1}$$

Here, u is the longitudinal mean flow velocity; θ is the mean temperature; y is the distance from the wall; ν is molecular viscosity; χ is molecular thermal diffusivity; and v_w, θ_w, τ_w, and j_w are the wall values of transverse velocity, temperature, shear stress, and temperature flux, respectively.

Temperature is treated here as a passive scalar; i.e., the second equation in (1) applies to passive-contaminant transport as well.

Equations (1) describe turbulent flow over an infinite plane where the pressure and transverse velocity are constant, while the remaining flow variables depend

only on the distance from the plane. For a flow of this kind,

$$\begin{aligned}\frac{du}{dy} &= F_1\left(y, \nu, v_w, \frac{\tau_w}{\rho}\right),\\ \frac{d\theta}{dy} &= F_2\left(y, \nu, \chi, v_w, \frac{\tau_w}{\rho}, j_w\right),\\ \langle u'v'\rangle &= F_3\left(y, \nu, v_w, \frac{\tau_w}{\rho}\right),\\ \langle \theta'v'\rangle &= F_4\left(y, \nu, \chi, v_w, \frac{\tau_w}{\rho}, j_w\right),\end{aligned} \tag{2}$$

where $F_1, \ldots, F_4$ are universal functions. Thus, these quantities are assumed to be independent of external boundary-layer parameters and are determined only by wall boundary conditions and physical properties of the fluid.

The first and second equations in (2) are solved for τ_w/ρ and j_w, and the results are substituted into the third and fourth equations to obtain

$$\begin{aligned}\langle u'v'\rangle &= G_1\left(y, \nu, v_w, \frac{du}{dy}\right),\\ \langle \theta'v'\rangle &= G_2\left(y, \nu, \chi, v_w, \frac{du}{dy}, \frac{d\theta}{dy}\right).\end{aligned} \tag{3}$$

Applying Buckingham's Π-theorem to (3) and treating temperature as a passive scalar having a specific dimension, one obtains

$$\begin{aligned}\langle u'v'\rangle &= -\left(y\frac{du}{dy}\right)^2 S(\mathrm{Re}, \beta),\\ \langle \theta'v'\rangle &= -y^2\frac{d\theta}{dy}\frac{du}{dy}T(\mathrm{Re}, \mathrm{Pe}, \beta),\end{aligned} \tag{4}$$

$$\mathrm{Re} = \frac{y^2}{\nu}\frac{du}{dy}, \quad \mathrm{Pe} = \frac{y^2}{\chi}\frac{du}{dy}, \quad \beta = \frac{v_w}{\mathrm{Re}y}\frac{dy}{du}.$$

The local Reynolds number Re is defined here as the ratio of characteristic turbulent and molecular viscosity values. The functions S and T are assumed to be continuous at

$$0 \le \mathrm{Re} \le \infty, \quad 0 \le \mathrm{Pe} \le \infty, \quad -\infty \le \beta \le \infty,$$

and have partial derivatives with respect to their arguments within these intervals. In physical terms, these conditions are equivalent to the standard assumption that viscosity is essential only within a narrow viscous sublayer. It is also assumed that

$$S(\infty, 0) \ne 0, \quad T(\infty, \infty, 0) \ne 0.$$

Expressions (4) relate shear stress and temperature flux to the mean-velocity gradient. Since the dependence of these relations on transpiration must weaken with increasing distance from the wall, the parameter β is defined so that the denominator contains the local Reynolds number Re.

The behavior of the Reynolds-stress component $\langle v'^2\rangle$ (due to transverse velocity fluctuations) is similar to that of shear stress:

$$\langle v'^2\rangle = \left(y\frac{du}{dy}\right)^2 S_2(\mathrm{Re}, \beta), \tag{5}$$

where S_2 satisfies the same conditions as does the function S in (4).

The components $\langle u'^2\rangle$ and $\langle w'^2\rangle$ (due to velocity fluctuations parallel to the wall) are left outside the scope of the present analysis. Their behavior near the surface is more complicated (e.g., see [7–9]) and poorly known even in the absence of mass flux across the surface.

By changing to the wall variables

$$y_+ = \frac{y}{\nu}\sqrt{\frac{\tau_w}{\rho}}, \quad u_+ = u\sqrt{\frac{\rho}{\tau_w}}, \quad v_+ = v_w\sqrt{\frac{\rho}{\tau_w}},$$

$$\theta_+ = \frac{\theta_w - \theta}{j_w}\sqrt{\frac{\tau_w}{\rho}}$$

and using (4), Eqs. (1) are rewritten as

$$\left(y_+\frac{du_+}{dy_+}\right)^2 S(\mathrm{Re}, \beta) + \frac{du_+}{dy_+} = 1 + v_+u_+, \tag{6}$$

$$u_+(0) = 0,$$

$$y_+^2\frac{d\theta_+}{dy_+}\frac{du_+}{dy_+}T(\mathrm{Re}, \mathrm{PrRe}, \beta)$$

$$+\frac{1}{\mathrm{Pr}}\frac{d\theta_+}{dy_+} = 1 + v_+\theta_+, \quad \theta_+(0) = 0, \tag{7}$$

$$\mathrm{Re} = y_+^2\frac{du_+}{dy_+}, \quad \beta = \frac{v_+}{\mathrm{Re}y_+}\frac{dy_+}{du_+},$$

where $\mathrm{Pr} = \nu/\chi$ is the molecular Prandtl number.

Thus, the problem is reduced to an ordinary differential equation (6) for the velocity profile. The temperature distribution governed by (7) is determined by the integral

$$\ln(1 + v_+\theta_+) = \int_0^{y_+}\frac{\mathrm{Pr}\,v_+\,dy_+}{1 + \mathrm{PrRe}T(\mathrm{Re}, \mathrm{PrRe}, \beta)}. \tag{8}$$

According to (4), the turbulent Prandtl number is

$$\mathrm{Pr}_t(\mathrm{Re}, \mathrm{Pe}, \beta) = \frac{S(\mathrm{Re}, \beta)}{T(\mathrm{Re}, \mathrm{Pe}, \beta)}. \tag{9}$$

If $\mathrm{Pr}_t \equiv 1$ and $\mathrm{Pr} = 1$, then Eq. (7) has the solution

$$\theta_+ \equiv u_+.$$

3. IMPERMEABLE WALL

In the case of impermeable wall ($v_+ = 0$), Eq. (6) has the closed-form solution

$$u_+ = \int_0^R \frac{d\mathrm{Re}}{\sqrt{\mathrm{Re}^2 S(\mathrm{Re}, 0) + \mathrm{Re}}} - \frac{\sqrt{\mathrm{Re}}}{\sqrt{\mathrm{Re} S(\mathrm{Re}, 0) + 1}},$$

$$y_+ = \sqrt{\mathrm{Re}^2 S(\mathrm{Re}, 0) + \mathrm{Re}}, \quad 0 \le \mathrm{Re} < \infty, \tag{10}$$

and integral (8) yields

$$\theta_+ = \int_0^R \frac{\mathrm{Pr}\, d\sqrt{\mathrm{Re}^2 S(\mathrm{Re}, 0) + \mathrm{Re}}}{1 + \mathrm{PrRe} T(\mathrm{Re}, \mathrm{PrRe}, 0)}. \tag{11}$$

It follows from (10), (11), and the conditions imposed on the functions S and T that both velocity and temperature exhibit logarithmic asymptotic behavior in the outer part of the wall region:

$$u_+ = \frac{1}{\kappa}(\ln y_+ + C_0) + O(y_+^{-\alpha}), \tag{12}$$

$$\theta_+ = \frac{\mathrm{Pr}_t^0}{\kappa}[\ln y_+ + B(\mathrm{Pr})] + O(y_+^{-\alpha}),$$

$$y_+ \longrightarrow \infty, \quad \alpha > 0, \tag{13}$$

$$\kappa = \sqrt{S(\infty, 0)}, \quad \mathrm{Pr}_t^0 = \mathrm{Pr}_t(\infty, \infty, 0).$$

Based on experimental data, the values of the constants in asymptotic representation (12) are set as follows:

$$\kappa = 0.41, \quad C_0 = 2.05.$$

The most frequently recommended values of the turbulent Prandtl number Pr_t^0 in the logarithmic region vary between 0.85 and 0.95 [10]. In what follows, $\mathrm{Pr}_t^0 = 0.89$ in accordance with experimental data in [11].

Asymptotic expressions for the function $B(\mathrm{Pr})$ in (13) at small and large molecular Prandtl numbers can be derived from the integral representation

$$B(\mathrm{Pr}) = \frac{\kappa \mathrm{Pr}}{\mathrm{Pr}_t^0}\int_0^1 \frac{d\sqrt{\mathrm{Re}^2 S(\mathrm{Re}, 0) + \mathrm{Re}}}{1 + \mathrm{PrRe} T(\mathrm{Re}, \mathrm{PrRe}, 0)}$$

$$-\int_1^\infty \frac{d\mathrm{Re}}{\mathrm{Re}(1 + \mathrm{PrRe} T(\mathrm{Re}, \mathrm{PrRe}, 0))} + \frac{\mathrm{Pr}}{\mathrm{Pr}_t^0}$$

$$\times \int_1^\infty \frac{\kappa d\sqrt{\mathrm{Re}^2 S(\mathrm{Re}, 0) + \mathrm{Re}} - \mathrm{Pr}_t^0 T(\mathrm{Re}, \mathrm{PrRe}, 0) d\mathrm{Re}}{1 + \mathrm{PrRe} T(\mathrm{Re}, \mathrm{PrRe}, 0)} \tag{14}$$

$$- \ln \kappa,$$

which follows from (11), (13), and (10). The asymptotic form corresponding to small Prandtl numbers is mainly determined by the behavior of the integrand in the second integral in (14) at high local Reynolds numbers:

$$B(\mathrm{Pr}) = \ln \mathrm{Pr} + b_1 + \ldots, \quad \mathrm{Pr} \longrightarrow 0,$$

$$b_1 = \int_0^1 \frac{T(\infty, \infty, 0) d\mathrm{Pe}}{1 + \mathrm{Pe} T(\infty, \mathrm{Pe}, 0)}$$

$$-\int_1^\infty \frac{[1 + \mathrm{Pe}(T(\infty, \mathrm{Pe}, 0) - T(\infty, \infty, 0))] d\mathrm{Pe}}{\mathrm{Pe}[1 + \mathrm{Pe} T(\infty, \mathrm{Pe}, 0)]} - \ln \kappa. \tag{15}$$

If turbulent thermal diffusivity is independent of χ, i.e., T is independent of the Peclet number (see [10]), then (15) yields

$$b_1 = \ln(\kappa / \mathrm{Pr}_t^0),$$

which is smaller by 1 as compared to the value suggested in [10].

At large Prandtl numbers, the dominant contribution to the asymptotic form of $B(\mathrm{Pr})$ is due to the first integral in (14), and the behavior of its integrand at the lower limit of integration is essential. Using the well-known estimate

$$\langle \theta' v' \rangle = O(y^3), \quad y \longrightarrow 0,$$

suppose that

$$T(\mathrm{Re}, \mathrm{PrRe}, 0) = k(\mathrm{Pr})\sqrt{\mathrm{Re}} + \ldots, \quad \mathrm{Re} \longrightarrow 0,$$

where $k(\mathrm{Pr})$ is some function. Then, the leading-order term in the asymptotic expression for $B(\mathrm{Pr})$ is

$$B(\mathrm{Pr}) = b_2 \mathrm{Pr}^{2/3} + \ldots, \quad \mathrm{Pr} \longrightarrow \infty,$$

$$b_2 = \frac{\kappa}{\mathrm{Pr}_t^0}\int_0^\infty \frac{dx}{1 + k(\infty) x^3} = \frac{2\pi\sqrt{3}\kappa}{9 \mathrm{Pr}_t^0 \sqrt[3]{k(\infty)}}.$$

This result is close to the values given by an approximate formula for b_2 proposed in [10].

By substituting (12) into (5), a constant value of the rms transverse velocity fluctuation in the outer subregion is obtained:

$$\sqrt{\langle v'^2 \rangle_+} = \frac{1}{\sigma_2} + O(y_+^{-\alpha}), \quad \sigma_2 = \frac{\kappa}{\sqrt{S_2(\infty, 0)}}, \tag{16}$$

$$y_+ \longrightarrow \infty.$$

The finite limit value $1/\sigma_2$ is a universal constant. Most experimental values of σ_2 are slightly less than unity. Henceforth, it is assumed that $\sigma_2 = 0.95$ (see experimental data in [12]).

4. TRANSPIRATION

In the general case of nonzero transverse velocity at the wall, the change to the variables

$$Y = \frac{y v_w}{\nu} = v_+ y_+, \quad w = \frac{2}{v_+}\sqrt{1 + v_+ u_+} \tag{17}$$

is performed in (6) to obtain the equation

$$\left(Y\frac{dw}{dY}\right)^2 S(\mathrm{Re}, \beta) + \frac{2}{w}\frac{dw}{dY} = 1, \quad w(0) = \frac{2}{v_+},$$
$$\mathrm{Re} = \frac{Y^2 w}{2}\frac{dw}{dY}, \quad \beta = \frac{2}{\mathrm{Re} Y w}\frac{dY}{dw}. \tag{18}$$

The variables in (17) are defined so that Eq. (18) is independent of v_+.

Since the integral curves of Eq. (18) are symmetric about the abscissa axis, it is sufficient to examine their behavior in the upper half-plane. The pattern of integral curves is qualitatively illustrated by Fig. 1, where solutions to Eq. (18) with $S = \kappa^2$ are plotted. Negative and positive values of Y correspond to injection and suction, respectively. The value of w increases from 0 to $+\infty$ along each curve.

In the case of suction, v_+ is small, because u_+ has a relatively large value at the outer edge of the wall region while the right-hand side of (6) remains positive. Therefore, the initial condition in (18) dictates that physical integral curves must intersect the ordinate axis at relatively large values of w (unphysical segments of integral curves are shown as dashed graphs in Fig. 1). In the general case, the velocity profile in the wall region of turbulent boundary layer with suction is described by integral-curve segments corresponding to large values of w.

All integral curves lying in the first quadrant are physical. Small values of $w(0)$ correspond to strong injection, i.e., large values of v_+.

4.1. Scaling Laws

Equation (18) entails

$$\frac{dw}{d\ln Y} = \frac{1}{\sqrt{S(\mathrm{Re}, \beta) + 1/\mathrm{Re}}}. \tag{19}$$

In the outer part of the wall region, the local Reynolds number is high and quantities on the order of 1/Re can be neglected in Eq. (19). Hence,

$$w = \frac{1}{\kappa}\ln Y + C_1(v_+) + O(Y^{-\alpha}), \quad \alpha > 0, \tag{20}$$

where $C_1(v_+)$ is some function. In the case of injection, expression (20) describes the asymptotic behavior of the solution to Eq. (6) as $Y \longrightarrow \infty$. In the case of suction, it corresponds to the intermediate asymptotic solution for the outer part of the wall region. After changing back to the original variables, (20) yields an asymptotic representation of velocity profile:

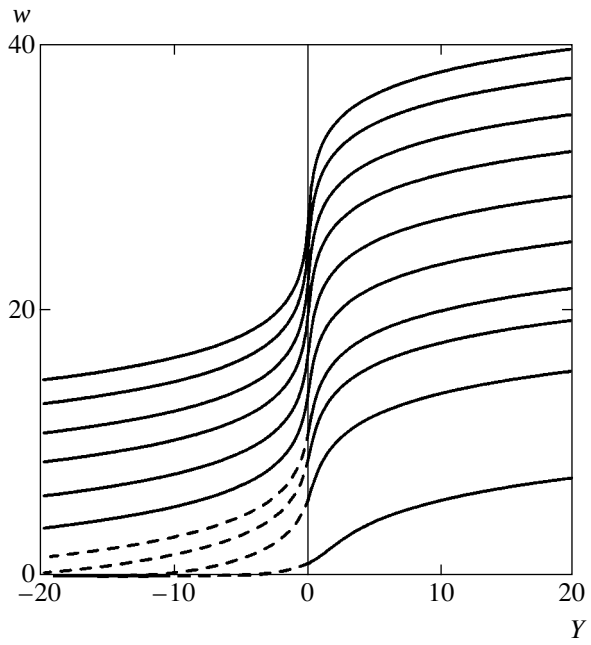


Fig. 1. Integral curves of Eq. (18) in the upper half-plane.

$$\frac{2}{v_+}(\sqrt{1 + v_+ u_+} - 1) = \frac{1}{\kappa}[\ln y_+ + C(v_+)] + O(y_+^{-\alpha}),$$
$$y_+ \longrightarrow \infty, \quad \alpha > 0, \tag{21}$$

where $C(v_+)$ is a universal function. Relation (21) extends the logarithmic law for velocity to flows with transpiration. Since it must reduce to (12) when $v_+ = 0$, it holds that

$$C(0) = C_0.$$

By virtue of (6), expression (21) also determines the shear-stress distribution in the outer part of the wall region:

$$\frac{2}{v_+}(\sqrt{-\langle u'v'\rangle_+} - 1) = \frac{1}{\kappa}[\ln y_+ + C(v_+)] + O(y_+^{-\alpha}),$$
$$y_+ \longrightarrow \infty. \tag{22}$$

An analogous expression for the rms transverse velocity fluctuation can be obtained by substituting (21) into (5):

$$\frac{2}{v_+}(\sigma_2\sqrt{\langle v'^2\rangle_+} - 1)$$
$$= \frac{1}{\kappa}[\ln y_+ + C(v_+)] + O(y_+^{-\alpha}), \quad y_+ \longrightarrow \infty, \tag{23}$$

where the constant σ_2 is defined in (16) and calculated by using results obtained for impermeable wall.

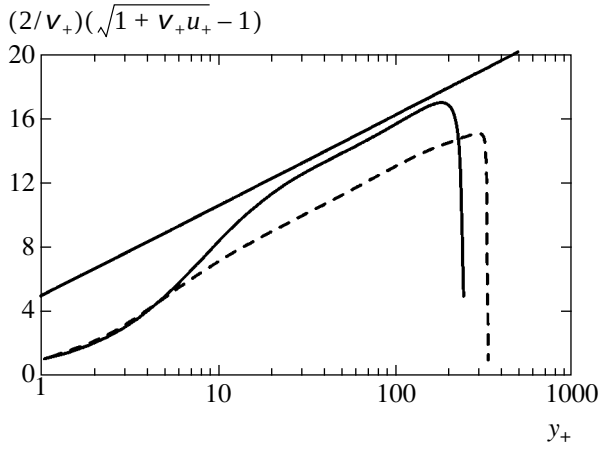


Fig. 2. Velocity profiles for Poiseuille flows with transverse mass flux in scaling variables. Solid and dashed curves represent data from [13, 14]. The solid line is $2.44\ln y_+ + 5.0$.

In view of (17) and (19), integral (8) is rewritten as

$$\ln(1+\nu_+\theta_+) = \int_{2/\nu_+}^{w} \frac{2\mathrm{Pr}[1+\mathrm{Re}S(\mathrm{Re},\beta)]dw}{w[1+\mathrm{PrRe}T(\mathrm{Re},\mathrm{PrRe},\beta)]}.$$

Then,

$$\frac{\ln(1+\nu_+\theta_+)}{2\mathrm{Pr}_t^0} = \ln|w| + \ln\left|\frac{\nu_+}{2}\right| + \int_{2/\nu_+}^{w}\left(\frac{\mathrm{Pr}[1+\mathrm{Re}S(\mathrm{Re},\beta)]}{\mathrm{Pr}_t^0[1+\mathrm{PrRe}T(\mathrm{Re},\mathrm{PrRe},\beta)]}-1\right)\frac{dw}{w}. \quad (24)$$

The integral in (24) is bounded for the outer subregion. After changing back to the original variables in (24) and using (21), it yields

$$\frac{2}{\nu_+}[(1+\nu_+\theta_+)^{1/2\mathrm{Pr}_t^0}-1] + \frac{1}{\kappa}D(\nu_+,\mathrm{Pr})(1+\nu_+\theta_+)^{1/2\mathrm{Pr}_t^0} = \frac{1}{\kappa}[\ln y_+ + C(\nu_+)] + O(y_+^{-\alpha}), \quad y_+ \longrightarrow \infty, \quad (25)$$

where $D(\nu_+, \mathrm{Pr})$ is some function. Since (25) must reduce to (13) when $\nu_+ = 0$, it holds that

$$D(0,\mathrm{Pr}) = C_0 - B(\mathrm{Pr}).$$

For unit turbulent and molecular Prandtl numbers, $D(\nu_+, 1) \equiv 0$.

Asymptotic relations (21)–(23) and (25) are scaling laws for longitudinal velocity, shear stress, rms transverse velocity fluctuation, and temperature, respectively, valid in the intermediate wall region outside the viscous sublayer.

Relation (21) was originally obtained in [5, 6] by invoking the Prandtl mixing-length theory, whereas scaling laws (22), (23), and (25) are newly derived here.

4.2. Small Values of ν_+

When ν_+ is small (e.g., in the case of flow with suction), the velocity, shear-stress, and temperature distributions can be represented as

$$\frac{2u_+}{1+\sqrt{1+\nu_+ u_+}} = u_+^0(y_+) + O(\nu_+),$$
$$\frac{2}{\nu_+}\left(\sqrt{-\langle u'v'\rangle_+ + \frac{du_+}{dy_+}}-1\right) = u_+^0(y_+) + O(\nu_+),$$
$$\left[\frac{2\mathrm{Pr}_t^0}{\nu_+} + \frac{\mathrm{Pr}_t^0}{\kappa}[C_0 - B(\mathrm{Pr})]\right][(1+\nu_+\theta_+)^{1/2\mathrm{Pr}_t^0}-1] = \theta_+^0(y_+) + O(\nu_+), \quad y_+ \geq 0, \quad (26)$$

where $u_+^0(y_+)$ and $\theta_+^0(y_+)$ are the velocity and temperature profiles for an impermeable wall. Substituting the first relation in (26) into (5) yields

$$\sqrt{\langle v'^2\rangle_+} = \sqrt{\langle v'^2\rangle_+^0} + \frac{\nu_+}{2\sigma_2}u_+^0(y_+), \quad y_+ \geq 0.$$

The first term on the right-hand side of this expansion is the rms velocity fluctuation in the boundary layer on impermeable surface.

Thus, in the case of arbitrary suction or weak injection (when ν_+ is small), the profiles in question can be expressed in terms of the characteristics of the flow over impermeable wall.

5. COMPARISON WITH EXPERIMENTAL AND NUMERICAL RESULTS

Figure 2 shows the velocity profiles for pressure-gradient-driven channel flow with transverse mass flux obtained by direct numerical simulation in [13, 14] and plotted in terms of scaling variables. The solid curve in Fig. 2 corresponds to $\nu_+ = 0.061$ and the value $\mathrm{Re}_m = 4357$ of the Reynolds number based on the mean velocity and the channel width; the dashed curve, to $\nu_+ = 0.241$ and $\mathrm{Re}_m = 8000$. Even though the Reynolds-number values used in the computations are relatively low, these distributions agree with (21): the graphs have intervals of almost linear growth with a slope close to $1/\kappa$ ($\kappa = 0.41$).

The graphs in Fig. 3 represent experimental velocity and shear-stress distributions in turbulent boundary layers with uniform transpiration reported in [15] (see also [16]). The data points plotted in Figs. 3a–3d corre-

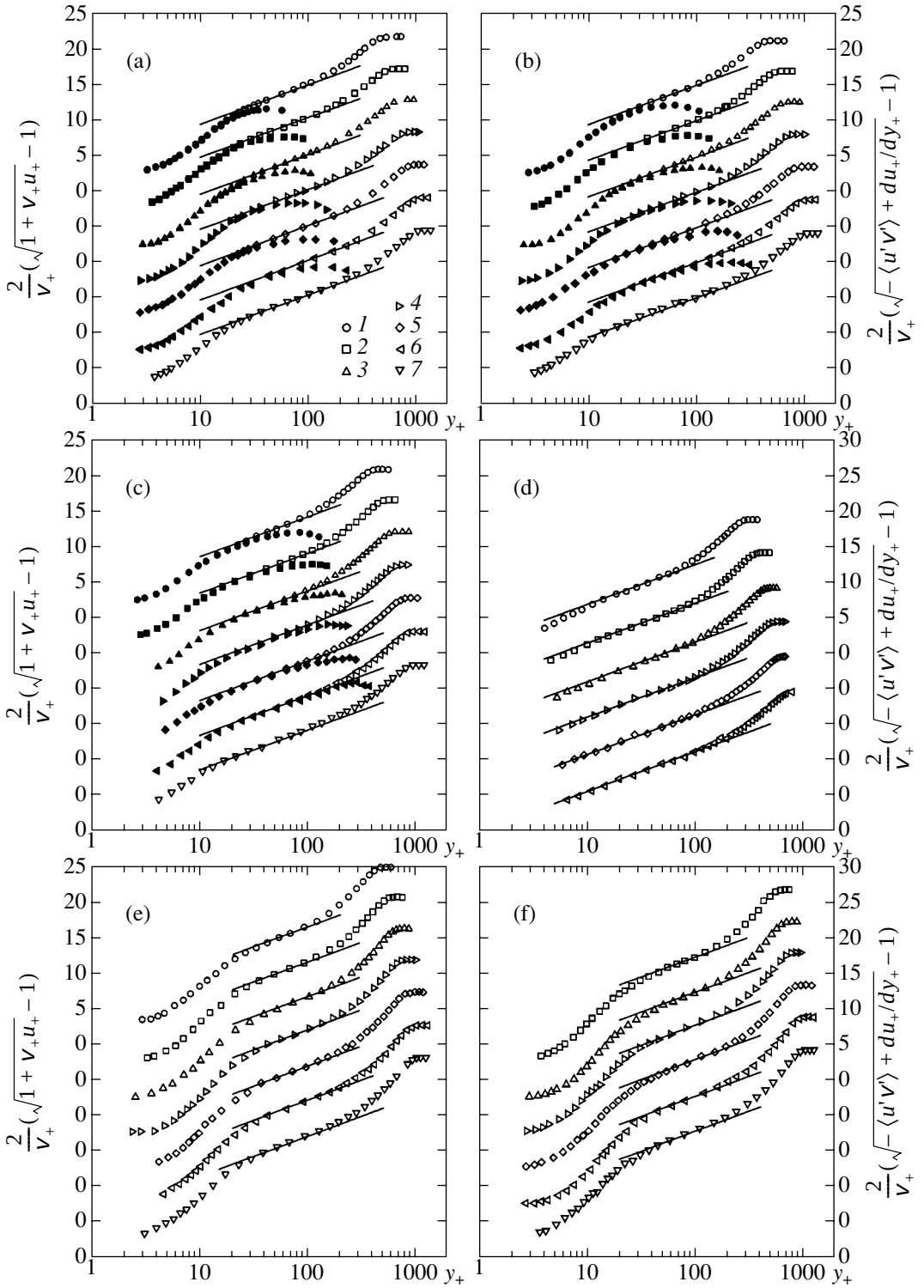


Fig. 3. Profiles of velocity (open symbols) and shear-stress (closed symbols) in boundary layer on a plate with transpiration at Reynolds numbers $R_x = 3.5\times10^5$ (*1*), 5.4×10^5 (*2*), 7.3×10^5 (*3*), 9.2×10^5 (*4*), 1.1×10^6 (*5*), 1.3×10^6 (*6*), and 1.4×10^6 (*7*) in scaling variables, based on data from [15]: $F = 0.001$ (a), 0.002 (b), 0.00375 (c), 0.008 (d), –0.001 (e), –0.002 (f).

spond to several values of $F = \nu_w/U_e$ for flows with injection and zero pressure gradient (U_e is free-stream velocity); the data points in Figs. 3e and 3f, to flows with suction and moderate adverse pressure gradient. The straight lines in Fig. 3 have the slope $1/\kappa$ and correspond to the right-hand side in (21) with $C(\nu_+)$ chosen to ensure the best approximation of experimental data (see Fig. 4). According to [15], experimental data

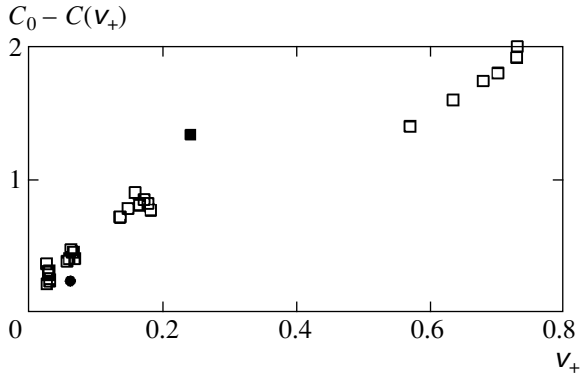


Fig. 4. Values of $C(v_+)$ based on experimental results from [15] (□) and numerical results from [13] (■) and [14] (●).

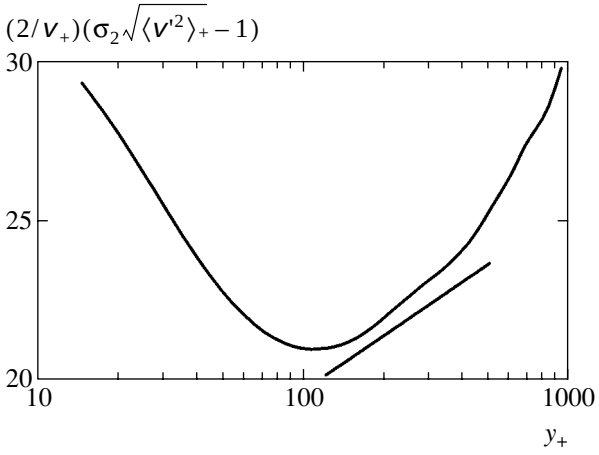


Fig. 5. Rms transverse velocity fluctuation in terms of scaling variables for the wall region based on DNS results from [17]. The line segment has the slope 1/κ.

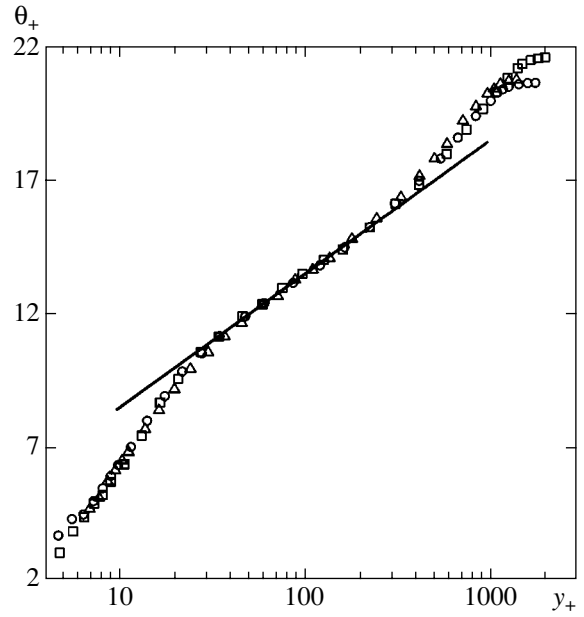


Fig. 6. Temperature profiles in turbulent boundary layers on an impermeable plate based on data from [11]: $R_x = 1.30 \times 10^6$ (○), 1.44×10^6 (□), 1.99×10^6 (△). The solid line is $2.17(\ln y_+ + 1.6)$.

for flow over an impermeable plate are in good agreement with logarithmic law (12) when $\kappa = 0.41$ and $C_0 = 2.05$. Figures 3a–3d demonstrate that the width of the logarithmic region (where scaling law (21) holds) increases with both Reynolds number and v_+. As v_+ increases, the inner boundary of the logarithmic region shifts toward the wall (see Fig. 3d, which corresponds to the strongest injection in the series of experiments reported in [15]).

Figures 3e and 3f demonstrate that the velocity profiles plotted in terms of scaling variables for flows with suction also have logarithmic intervals.

The velocity and shear-stress profiles plotted in terms of scaling variables are almost identical in the viscous sublayer and very similar in the outer part of the wall region, i.e., providing experimental evidence of the validity of the first equation in (1). It is clear from Figs. 3b and 3c that the square root of shear stress in flow with nonzero mass flux across the wall does exhibit logarithmic asymptotic behavior in the wall region.

Currently, relation (23) cannot be validated by comparison with experiment because of the lack of sufficiently accurate measurements of velocity fluctuations in the wall regions of flows with transpiration. Figure 5 shows the profile of rms transverse velocity fluctuations in the asymptotic turbulent boundary layer with suction at $v_+ = -0.0601$ and $\mathrm{Re}_{\delta^*} = 1000$ (the Reynolds number based on the displacement thickness) plotted in terms of the scaling variables used in (23) for the DNS results reported in [17, 18]. Figure 5 demonstrates that the profile has a distinct logarithmic interval even though the computations were performed at a relatively low Reynolds number, but its slope exceeds 1/κ by approximately 15%.

Relation (25) was verified against experimental data from [11], where temperature profiles and Stanton numbers were measured while skin friction was not. The skin friction coefficient was calculated by using the Reynolds number based on the momentum thickness and a universal skin-friction law [19, 20].

Figure 6 shows temperature profiles for boundary layer on impermeable plate. The experimental data points plotted in Fig. 6 are consistently described by expression (13) with $\mathrm{Pr}_t^0 = 0.89$ and $B = 1.6$.

Figure 7 shows temperature profiles for flows with transpiration plotted in terms of the scaling variables used in (25) with $C(v_+)$ determined from the experimental data presented in Fig. 4. The function $D(v_+, \mathrm{Pr})$ is defined so that the experimental data points follow linear graphs at $\mathrm{Pr}_t^0 = 0.89$ (see Fig. 8).

In agreement with scaling law (25), the experimental profiles shown in Fig. 7 have distinct logarithmic intervals, and the function plotted in Fig. 8 by using experimental data exhibits clear-cut monotonic behavior.

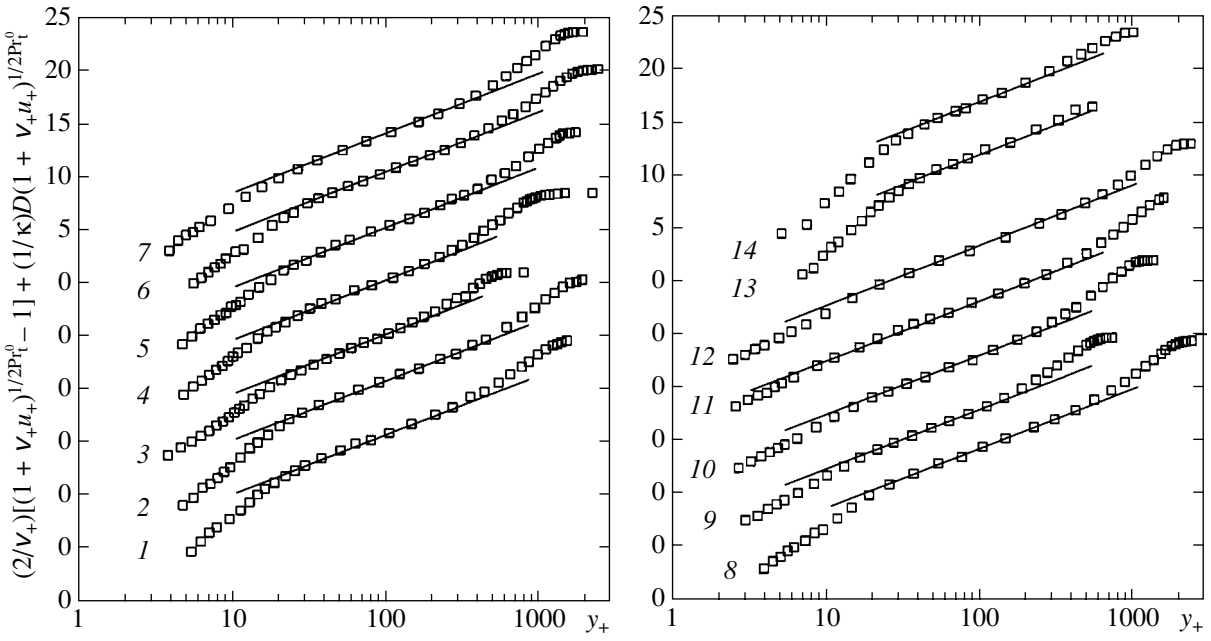


Fig. 7. Temperature profiles in boundary layers on a plate with transpiration in scaling variables, based on data from [11]: (*1*) $F = 9.6\times10^{-4}$, $R_x = 1.5\times10^6$; (*2*) 9.0×10^{-4}, 2.0×10^6; (*3*) 2.4×10^{-3}, 3.9×10^5; (*4*) 2.0×10^{-3}, 9.3×10^5; (*5*) 1.9×10^{-3}, 1.5×10^6; (*6*) 1.8×10^{-3}, 2.0×10^6; (*7*) 3.8×10^{-3}, 1.5×10^6; (*8*) 3.5×10^{-3}, 2.0×10^6; (*9*) 6.0×10^{-3}, 4.0×10^5; (*10*) 5.1×10^{-3}, 9.2×10^5; (11) 4.7×10^{-3}, 1.5×10^6; (*12*) 4.4×10^{-3}, 2.0×10^6; (*13*) -1.3×10^{-3}, 4.1×10^5; (*14*) -1.1×10^{-3}, 9.6×10^5.

According to (9), the turbulent Prandtl number for the logarithmic region is independent of the wall value of transverse velocity, in agreement with experimental evidence [21].

6. VELOCITY AND TEMPERATURE PROFILES IN FLOWS WITH STRONG INJECTION

Now, let us apply the method of matched asymptotic expansions [22] to analyze the asymptotic behavior of velocity and temperature profiles corresponding to high values of ν_+. In this case, the wall region consists of four distinct subregions.

For subregion *I* (adjoining the wall), it holds that $Y = O(1)$, turbulent shear stress can be neglected in Eq. (18), and the leading-order term in the solution is similar to that for purely laminar flow:

$$w = \frac{2}{\nu_+}e^{Y/2} + O(\nu_+^{-4}),$$
$$\ln(1+\nu_+\theta_+) = \mathrm{Pr}Y + O(\nu_+^{-3}). \tag{27}$$

The solution for subregion *II* (next to *I*) is sought in the form

$$Y = M + Y_2, \quad w = \frac{W_2(Y_2)}{M} + \ldots,$$
$$Y_2 = O(1), \quad M \longrightarrow \infty. \tag{28}$$

Substituting (28) into (18) and taking the limit as $M \longrightarrow \infty$ for $Y_2 = O(1)$ yields

$$\left(\frac{dW_2}{dY_2}\right)^2 S(\mathrm{Re},\infty) + \frac{2}{W_2}\frac{dW_2}{dY_2} = 1,$$
$$\mathrm{Re} = \frac{W_2}{2}\frac{dW_2}{dY_2}. \tag{29}$$

Thus, the turbulent and laminar components of shear stress are comparable in order of magnitude. The boundary condition for Eq. (29) and the value of *M* are

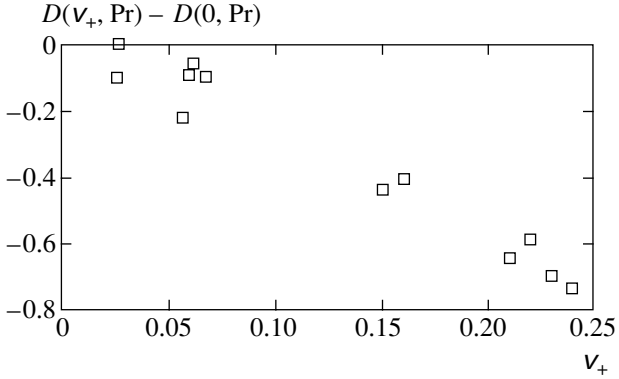


Fig. 8. Values of $D(\nu_+, \mathrm{Pr})$ based on experimental results from [11].

determined by asymptotic matching of expansions (27) and (28):

$$W_2 \longrightarrow 2e^{Y_2/2}, \quad Y_2 \longrightarrow -\infty; \quad Me^{M/2} = \nu_+. \quad (30)$$

Boundary condition (30) is then used to obtain a solution to Eq. (29) in parametric form:

$$Y_2 = \ln\mathrm{Re} + \mathrm{Re}S(\mathrm{Re}, \infty) + \int_0^{\mathrm{Re}} S(\mathrm{Re}, \infty)d\mathrm{Re},$$
$$W_2 = 2\sqrt{\mathrm{Re}^2 S(\mathrm{Re}, \infty) + \mathrm{Re}}, \quad 0 < \mathrm{Re} < \infty. \quad (31)$$

After writing the temperature distribution in subregion *II* as

$$\ln(1 + \nu_+\theta_+) = \mathrm{Pr}Y^* + \int_{Y^* - M}^{Y_2} \frac{\mathrm{Pr}dY_2}{1 + \mathrm{PrRe}T(\mathrm{Re}, \mathrm{PrRe}, \infty)} + \ldots, \quad (32)$$
$$Y^* = O(1)$$

the variable in the integrand in (32) is changed by using (31) and the limit is taken as $\mathrm{Re} \longrightarrow \infty$ to find

$$\ln(1 + \nu_+\theta_+) = \mathrm{Pr}M + \mathrm{Pr}\int_0^{\mathrm{Re}} \frac{\mathrm{Re}dS(\mathrm{Re}, \infty) + 2S(\mathrm{Re}, \infty)d\mathrm{Re}}{1 + \mathrm{PrRe}T(\mathrm{Re}, \mathrm{PrRe}, \infty)} + \mathrm{Pr}\int_1^{\mathrm{Re}} \frac{d\mathrm{Re}}{\mathrm{Re}[1 + \mathrm{PrRe}T(\mathrm{Re}, \mathrm{PrRe}, \infty)]} - \mathrm{Pr}^2\int_0^1 \frac{T(\mathrm{Re}, \mathrm{PrRe}, \infty)d\mathrm{Re}}{1 + \mathrm{PrRe}T(\mathrm{Re}, \mathrm{PrRe}, \infty)} + \ldots.$$

The condition at the outer edge of subregion *II* is obtained as $\mathrm{Re} \longrightarrow \infty$:

$$\ln(1 + \nu_+\theta_+) = 2\mathrm{Pr}_t^\infty \ln Y_2 + \mathrm{Pr}M + a_2(\mathrm{Pr}) + \ldots, \quad Y_2 \longrightarrow \infty, \quad (33)$$
$$\mathrm{Pr}_t^\infty = \mathrm{Pr}_t(\infty, \infty, \infty),$$
$$a_2(\mathrm{Pr}) = \mathrm{Pr}\int_0^\infty \frac{\mathrm{Re}dS(\mathrm{Re}, \infty)}{1 + \mathrm{PrRe}T(\mathrm{Re}, \mathrm{PrRe}, \infty)} + \mathrm{Pr}\int_1^\infty \frac{d\mathrm{Re}}{\mathrm{Re}[1 + \mathrm{PrRe}T(\mathrm{Re}, \mathrm{PrRe}, \infty)]} - \mathrm{Pr}^2\int_0^1 \frac{T(\mathrm{Re}, \mathrm{PrRe}, \infty)d\mathrm{Re}}{1 + \mathrm{PrRe}T(\mathrm{Re}, \mathrm{PrRe}, \infty)}.$$

In subregion *III*,

$$Y = M + \sqrt{M}Y_3, \quad w = \frac{W_3(Y_3)}{\sqrt{M}} + \ldots,$$
$$\frac{1}{Y_3} = O(1), \quad Y_3 = O(1).$$

In the limit of

$$M \longrightarrow \infty, \quad \frac{1}{Y_3} = O(1), \quad Y_3 = O(1),$$

Eq. (18) retains only the turbulent component of shear stress:

$$\left(\frac{dW_3}{dY_3}\right)^2 S(\infty, \beta) = 1, \quad \beta = \left(\frac{2}{W_3}\frac{dY_3}{dW_3}\right)^2. \quad (34)$$

The solution to Eq. (34) that satisfies the condition for matching with the solution for subregion *II* is

$$Y_3 = \frac{S(\infty, \beta)}{\sqrt{\beta}} + \int_\beta^\infty \frac{S(\infty, \beta)d\beta}{2\beta^{3/2}},$$
$$W_3 = 2\sqrt{\frac{S(\infty, \beta)}{\beta}}, \quad 0 < \beta < \infty. \quad (35)$$

The temperature distribution in subregion *III* is obtained by adding the integral

$$I_3 = \int_{Y_2/\sqrt{M}}^{Y_3} \frac{\sqrt{\beta}dY_3}{T(\infty, \infty, \beta)}$$

to (32). After using (35) to change the variable in this integral, it is rewritten as

$$I_3 = \int_{\beta^*}^\beta \frac{\beta dS(\infty, \beta) - S(\infty, \beta)d\beta}{\beta T(\infty, \infty, \beta)},$$
$$\beta^* = M\left(\frac{2S(\infty, \infty)}{Y_2}\right)^2 + \ldots.$$

In the limit of $M \longrightarrow \infty$,

$$I_3 = \mathrm{Pr}_t^\infty[\ln M - 2\ln Y_2 + 2\ln(2S(\infty, \infty))] + \int_\beta^1 \frac{\mathrm{Pr}_t(\infty, \infty, \beta)d\beta}{\beta} - \int_\beta^\infty \frac{dS(\infty, \beta)}{T(\infty, \infty, \beta)} + \int_1^\infty [\mathrm{Pr}_t(\infty, \infty, \beta) - \mathrm{Pr}_t^\infty]\frac{d\beta}{\beta} + \ldots.$$

As $\beta \longrightarrow 0$, the sum of (33) with this expression yields

the condition at the outer edge of subregion *III*:

$$\ln(1 + v_+\theta_+) = 2\mathrm{Pr}_t^0 \ln Y_3 + \mathrm{Pr}M + \mathrm{Pr}_t^\infty \ln M + a_2(\mathrm{Pr}) + a_3 + \ldots, \quad Y_3 \longrightarrow \infty, \tag{36}$$

$$a_3 = \int_0^1 [\mathrm{Pr}_t(\infty, \infty, \beta) - \mathrm{Pr}_t^0]\frac{d\beta}{\beta} - \int_0^\infty \frac{dS(\infty, \beta)}{T(\infty, \infty, \beta)} + \int_1^\infty [\mathrm{Pr}_t(\infty, \infty, \beta) - \mathrm{Pr}_t^\infty]\frac{d\beta}{\beta} - 2\mathrm{Pr}_t^0 \ln(2\kappa^2).$$

In the outermost subregion *IV*,

$$Y = MY_4, \quad w = W_4(Y_4) + \ldots,$$

$$1/Y_4 = O(1),$$

$$\left(\frac{dW_4}{d\ln Y_4}\right)^2 S(\infty, 0) = 1. \tag{37}$$

The solution to Eq. (37) that satisfies the condition for matching with the solution for subregion *III* is

$$W_4 = \frac{1}{\kappa}\ln Y_4. \tag{38}$$

Combined with (21), expression (38) determines the asymptotic form of $C(v_+)$:

$$C(v_+) = \frac{M}{2} + \ldots, \quad v_+ \longrightarrow \infty, \qquad M + 2\ln M = 2\ln v_+. \tag{39}$$

The temperature distribution for subregion *IV* is obtained by adding the integral

$$I_4 = 2\mathrm{Pr}_t^0 \int_{1 + Y_3/\sqrt{M}}^{Y_4} \frac{dY_4}{Y_4 \ln Y_4}$$

to (36). The resulting expression

$$\ln(1 + v_+\theta_+) = 2\mathrm{Pr}_t^0 \ln\ln Y_4 + \mathrm{Pr}M + (\mathrm{Pr}_t^\infty + \mathrm{Pr}_t^0)\ln M + a_2(\mathrm{Pr}) + a_3 + \ldots$$

is combined with (25) to find the asymptotic representation

$$D(v_+, \mathrm{Pr}) = \exp\left[-\frac{\mathrm{Pr}M}{2\mathrm{Pr}_t^0} - \frac{\mathrm{Pr}_t^0 + \mathrm{Pr}_t^\infty}{2\mathrm{Pr}_t^0}\ln M - \frac{a_2(\mathrm{Pr}) + a_3}{2\mathrm{Pr}_t^0}\right] - \frac{2\kappa}{M}e^{-M/2} + \ldots, \quad v_+ \longrightarrow \infty.$$

This function tends to zero with increasing v_+.

According to Fig. 4, the function $C_0 - C(v_w^+)$ plotted by using velocity profiles obtained experimentally in [15] and numerically in [13, 14] increases toward $v_+ \approx 0.8$. However, $C(v_+) \longrightarrow \infty$ as $v_+ \longrightarrow \infty$ by virtue of (39). Therefore, $C(v_+)$ is a nonmonotonic function, and the graph presented in Fig. 4 must have a maximum.

Subregions *I* and *II* constitute the viscous sublayer. Denote by d_+ its characteristic thickness normalized to the near-wall length scale $\rho\nu/\tau_w$. The above analysis shows that

$$d_+ = O(\ln v_+/v_+)$$

as $v_+ \longrightarrow \infty$. It should be recalled that injection reduces the viscous-sublayer thickness and widens the logarithmic region (see analysis of experimental data presented in Fig. 3).

7. CONCLUSIONS

A model based on fluid-dynamics equations, standard assumption of near-wall flow independence of external parameters, and continuity of functions *S* and *T* representing turbulent shear stress and temperature flux is formulated. Scaling relations for velocity, temperature, and Reynolds-stress components are derived without invoking any hypotheses about the mechanism of turbulent transfer. Asymptotic representations of the universal functions $C(v_+)$ and $D(v_+, \mathrm{Pr})$ contained in these relations are obtained.

The wall region of turbulent boundary layer with transpiration consists of two distinct subregions: a viscous sublayer adjoining the wall, where the turbulent and viscous stress components are comparable in order of magnitude, and an outer subregion, where molecular viscosity is negligible. In the outer subregion, the scaling laws for velocity, temperature, shear stress, and rms transverse velocity fluctuation are determined by the same logarithmic distribution, which depends only on the dimensionless transpiration velocity v_+.

As $v_+ \longrightarrow \infty$, the viscous-sublayer thickness measured in units of near-wall length scale tends to zero as $O(\ln v_+/v_+)$, and the domain where the scaling laws are valid shifts toward the wall.

ACKNOWLEDGMENTS

I thank W.M. Kays and R.J. Moffat of Stanford University for kindly providing their experimental data. This work was supported under the State Program for Support of Leading Science Schools, grant 1635.2003.1.

REFERENCES

1. A. A. Izakson, Zh. Éksp. Teor. Fiz. **7**, 919 (1937) [Tech. Phys. USSR **IV**, 155 (1937)].
2. C. B. Millikan, in *Proceedings of 5th International Congress on Applied Mechanics* (Wiley, New York, 1939), p. 386.
3. R. von Mises, in *Th. von Kármán Anniversary Volume* (Calif. Inst. Tech. Press, Pasadena, 1941), p. 317.
4. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 6: *Fluid Mechanics*, 3rd ed. (Nauka, Moscow, 1986; Pergamon, New York, 1987).
5. T. N. Stevenson, CoA Rep. Aero, No. 166 (1963).
6. T. N. Stevenson, AIAA J. **6**, 553 (1968).
7. A. E. Perry, S. Henbest, and M. S. Chong, J. Fluid Mech. **165**, 163 (1986).
8. H. H. Fernholz and P. J. Finley, Prog. Aerosp. Sci. **32**, 245 (1996).
9. J. F. Morrison, W. Jiang, B. J. McKeon, *et al.*, in *Turbulence and Shear Flow Phenomena: Second International Symposium*, Ed. by E. Lindborg *et al.* (KTH, Stockholm, 2001), Vol. 1, p. 43.
10. B. A. Kader and A. M. Yaglom, Itogi Nauki Tekh., Ser.: Mekh. Zhidk. Gaza **18**, 81 (1980).
11. D. G. Whitten, W. M. Kays, and R. J. Moffat, Rep. HMT-3 (Stanford Univ., 1967).
12. P. S. Klebanoff, NACA Rep. 1247 (1954).
13. Y. Sumitani and N. Kasagi, AIAA J. **33**, 1220 (1995).
14. N. V. Nikitin and A. A. Pavel'ev, Izv. Ross. Akad. Nauk, Mekh. Zhidk. Gaza, No. 6, 18 (1998) [Fluid Dynamics **33**, 413 (1998)].
15. P. S. Andersen, W. M. Kays, and R. J. Moffat, Rep. HMT-15 (Stanford Univ., 1972).
16. P. S. Andersen, W. M. Kays, and R. J. Moffat, J. Fluid Mech. **69**, 353 (1975).
17. P. Mariani, P. Spalart, and W. Kollmann, in *Proceedings of International Conference on Near-Wall Turbulent Flows*, Ed. by R. M. C. So *et al.* (Elsevier, Amsterdam, 1993).
18. R. A. Antonia, P. R. Spalart, and P. Mariani, Phys. Fluids **6**, 430 (1994).
19. I. I. Vigdorovich, Dokl. Akad. Nauk **337**, 39 (1994) [Phys. Dokl. **39**, 514 (1994)].
20. I. I. Vigdorovich, Dokl. Akad. Nauk **356**, 42 (1997) [Phys. Dokl. **42**, 506 (1997)].
21. R. L. Simpson, R. J. Moffat, and D. G. Whitten, Int. J. Heat Mass Transf. **13**, 125 (1970).
22. M. Van Dyke, *Perturbation Methods in Fluid Mechanics* (Academic, New York, 1964; Mir, Moscow, 1967).

Translated by A. Betev

Journal of Experimental and Theoretical Physics, Vol. 99, No. 5, 2004, pp. 1039–1043.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 5, 2004, pp. 1192–1197.

FLUIDS

Thermomechanical Coupling in a Cylindrical Hybrid-Aligned Nematic Liquid Crystal

R. S. Hakobyan[a,*], B. Ya. Zeldovich[b,], and H. Seferyan[c]**

[a] *Department of Physics, Yerevan State University, Yerevan, 375025 Armenia*
[b] *School of Optics/CREOL, University of Central Florida, P.O. Box 16-2700, Orlando, FL 32816-2700, USA*
[c] *University of California, Irvine, CA 92697-2025, USA*
*e-mail: *rhakob@server.physdep.r.am; **boris@creol.ucf.edu*

Received July 8, 2004

Abstract—A solution to the system of equations describing a cylindrical hybrid-aligned nematic liquid crystal is obtained. The rotational flow driven by vertical temperature gradient in such a cell is investigated theoretically. The cell is suggested as a new experimental setup for determining an additional relation required to measure the twelve thermomechanical coefficients. It is shown that the terms in the expressions for thermomechanical stress and heat flux obtained in [8] are equivalent to those originally proposed in [7]. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Shortly after the discovery of liquid crystals (LCs), rotation of a cholesteric drop induced by vertical temperature gradient was observed by Lehmann [1]. For cholesteric LCs, thermomechanical effects of this kind were investigated in detail both experimentally and theoretically in [2–5] and attributed specifically to the chirality of cholesterics. Consequently, effects of this kind would not be expected to occur in nonchiral LCs, such as nematics (NLCs) [2–6]. However, new thermomechanical effects were predicted for deformed nematics in [7], where the first consistent theory of thermomechanical coupling due to nonuniform director orientation under a temperature gradient was developed for uniaxial nematics. Thermomechanical effects of three basic types were considered in [7]: hydrodynamic excitation induced by temperature gradient (direct thermomechanical effect), temperature gradient arising in nonuniform flow (inverse thermomechanical effect), and additional director deflection caused by heat flow.

The validity of the thermomechanical terms in the equations obtained in [7] was questioned in [8], where these terms were written in somewhat different form. The discrepancy is elucidated in the Appendix. However, the thermomechanical coupling predicted in [7] was observed in numerous experimental studies [9–11], and the measured values of thermomechanical coefficients were in good agreement with the theoretical estimates obtained in [7]. Thermomechanical coupling, including some effects induced by external fields, was investigated experimentally in [12], where an original cylindrical hybrid-aligned nematic cell was proposed and thermally driven rotation in such a cell was observed.

In this paper, we solve the system of equations for the director distribution in a cylindrical homeotropic–planar hybrid nematic cell. A theoretical analysis of rotation of the cell induced by a temperature gradient along its axis is presented. We suggest that the cell can be used as a new experimental setup for determining an additional relation required to measure the twelve thermomechanical coefficients. To derive the lacking relations between the coefficients, alternative director orientations should be implemented by using nonuniform static or oscillating magnetic fields. The required nonuniform director orientations can be obtained by reorienting the director field through a Fréedericksz transition in an external magnetic field.

The paper is organized as follows. In Section 2, we examine the static director distribution in a cylindrical hybrid aligned nematic cell. In Section 3, we analyze thermally driven rotation of the cell. In Section 4, we discuss the reversibility of thermomechanical coupling and present our conclusions. In the Appendix, we show that the terms in the expressions for thermomechanical stress and heat flux obtained in [8] are equivalent to those originally proposed in [7].

2. DIRECTOR DISTRIBUTION IN A CYLINDRICAL HYBRID CELL

Consider a nematic layer in a cylindrical cell with the z axis directed vertically upwards. The boundary conditions on the lower and upper substrates correspond to directors oriented along the z axis (homeotropic orientation) and along the tangents to concentric circles, respectively (see Fig. 1). By virtue of symmetry, every director $\mathbf{n}$ belongs to a cylindrical surface coaxial with the z axis. It is also clear that $\mathbf{n}$ is not

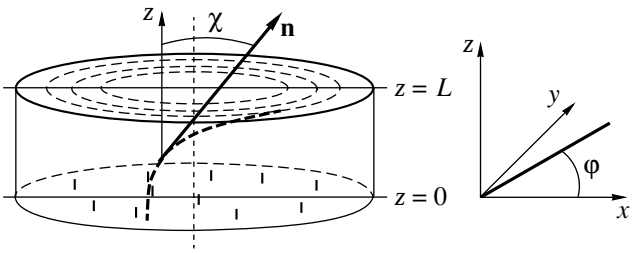


Fig. 1. Cylindrical hybrid-aligned nematic cell.

defined on the axis; i.e., the z axis is a disclination. Therefore, the director distribution can be determined only within the interval $a < r < R$, where a is the molecular size and R is the cylinder radius.

Define a cylindrical coordinate system (r, φ, z). By virtue of symmetry, the director distribution is independent of r and φ. Denoting the angle between the director **n** and the z axis by χ, we write the components of **n** as

$$n_\varphi = \sin\chi(z), \quad n_r = 0, \quad n_z = \cos\chi(z).$$

The boundary conditions for the function $\chi(z)$ are

$$\chi(0) = 0, \quad \chi(L) = \pi/2,$$

where L is the NLC layer thickness.

In the cylindrical coordinates, the Frank elastic free energy density F_d is expressed as follows [13]:

$$F_d = \frac{1}{2}K_1\chi'^2\sin^2\chi + \frac{1}{8r^2}K_2\sin^2(2\chi) + \frac{1}{2}K_3\left(\frac{1}{r^2}\sin^4\chi + \chi'^2\cos^2\chi\right), \quad (1)$$

where K_1, K_2, and K_3 are the Frank elastic constants and $\chi' = d\chi/dz$. The director distribution is determined by minimizing the elastic free energy for the cell volume:

$$F = 2\pi\int F_d r\,dr\,dz.$$

The integral over dr is logarithmically divergent. However, we can use R and a, respectively, as the upper and lower limits. Since the latter (the length scale on which the macroscopic theory fails) defines a cylinder coaxial with the outer cylinder of radius R, the integral from a to R is readily calculated. Solving the ensuing Euler–Lagrange equation, we obtain the integral of motion [7]

$$\chi'\frac{\delta F}{\delta\chi'} - F = \text{const}.$$

It can also be found by applying the Noether theorem [14]. The function $\chi(z)$ is defined implicitly by the following relation:

$$z = \int_0^\chi \left\{ \frac{\frac{R^2 - a^2}{2}[K_1\sin^2\chi + K_3\cos^2\chi]}{\text{const} + \ln\left(\frac{R}{a}\right)\left[\frac{1}{4}K_2\sin^2(2\chi) + K_3\sin^4\chi\right]} \right\}^{1/2} dx, \quad (2)$$

where the boundary condition $\chi(0) = 0$ is used. The boundary condition $\chi(L) = \pi/2$ determines the integration constant. Numerical computation of the expression (2) using Mathematica 5 shows that the function $\chi(z)$ is almost linear. In particular, it has the form $\chi(z) = \pi z/2L$ in Frank's single-constant approximation.

3. THERMALLY DRIVEN ROTATION OF A NEMATIC CELL

Suppose that external heat sources are used to keep the temperatures on the lower and upper planes ($z = 0$ and $z = L$) at $T_0 + \Delta T$ and T_0, respectively. The resulting temperature gradient gives rise to solid-body NLC rotation, which is described by the Navier–Stokes equations written in the linear steady-state approximation as follows [7]:

$$\frac{\partial\sigma_{ki}}{\partial x_k} = 0, \quad \sigma_{ki} = -p\delta_{ki} + \sigma'_{ki} + \sigma_{ki}^{\text{thm}}, \quad (3)$$

where p is pressure, σ'_{ki} denotes viscous stress components, and σ_{ki}^{thm} denotes thermomechanical stress components. The problem linearized with respect to velocity and director disturbances should be solved in the cylindrical coordinate system by using the single-constant approximation $\xi_1 = \xi_2 = \ldots = \xi_{12} = \xi$, where ξ_i are thermomechanical coefficients [7]. By virtue of symmetry, it is also clear that

$$\partial/\partial\varphi = 0, \quad v_r = v_z = 0, \quad v_\varphi = r\omega(r, z),$$

where **v** is the hydrodynamic velocity and $\omega(r, z) = d\varphi/dt$ is the angular velocity of NLC rotation. Finally, we have the following equation for the angular velocity:

$$[3\eta_3 + \eta_5\sin^2\chi]\frac{\partial\omega}{\partial r} + [(\eta_1 - \eta_2)\sin(2\chi) + 2\eta_4\sin(4\chi)]\chi' r\frac{\partial\omega}{\partial z} + [\eta_3 + \eta_6\sin^2\chi]r\frac{\partial^2\omega}{\partial r^2} + [\eta_2 + (\eta_1 - \eta_2)\sin^2\chi + \eta_4\sin^2(2\chi)]r\frac{\partial^2\omega}{\partial z^2} + \frac{1}{8}\xi\frac{\Delta T}{L}\left[(3\sin^2\chi - 2)\frac{\sin(2\chi)}{r^2}\right. \quad (4)$$

$$-(12-18\cos^2\chi+6\cos^4\chi+\sin^2(2\chi))\frac{d^2\chi}{dz^2}$$

$$-(18-12\cos^2\chi+4\cos(2\chi))\chi'^2\sin(2\chi)\Big] = 0.$$

Here,

$$\eta_1 = 0.5(\alpha_3+\alpha_4+\alpha_6),\quad \eta_2 = 0.5(\alpha_4+\alpha_5-\alpha_2),$$
$$\eta_3 = 0.5\alpha_4$$

are the Miesowicz viscosities;

$$\eta_4 = 0.25\alpha_1,\ \eta_5 = 1.5\alpha_6+0.5\alpha_3,\ \eta_6 = 0.5(\alpha_3+\alpha_6)$$

and α_1, α_2, …, α_6 are the Leslie coefficients. As $R \longrightarrow \infty$, Eq. (4) can be reduced to an equation for the velocity $v = \omega R$ of thermally driven flow in a planar hybrid cell [12].

Equation (4) is a second-order linear elliptical equation. To formulate an appropriate Dirichlet problem, we set the following boundary conditions:

$$\omega(r, z=0) = \omega(r, z=L)$$
$$= \omega(r=a, z) = \omega(r=R, z) = 0.$$

They correspond to a nematic confined between two coaxial rigid cylinders of height L with radii a and R. The Dirichlet problem can be solved by using Mathematica 5. A numerical analysis of Eq. (4) was performed for MBBA as an example of NLC. The computed results are shown in Figs. 2 and 3.

When the planar concentric orienting substrate is at the top and the cell is heated from below, the nematic is expected to rotate steadily about the z axis. When a similar nematic cell is heated from above, oscillatory rotation should be expected. These results are in qualitative agreement with recent experimental observations [12].

Moreover, our numerical analysis shows that the maximum rotation velocity (with respect to z and r) is proportional to the temperature difference between the upper and lower substrates (see Fig. 4).

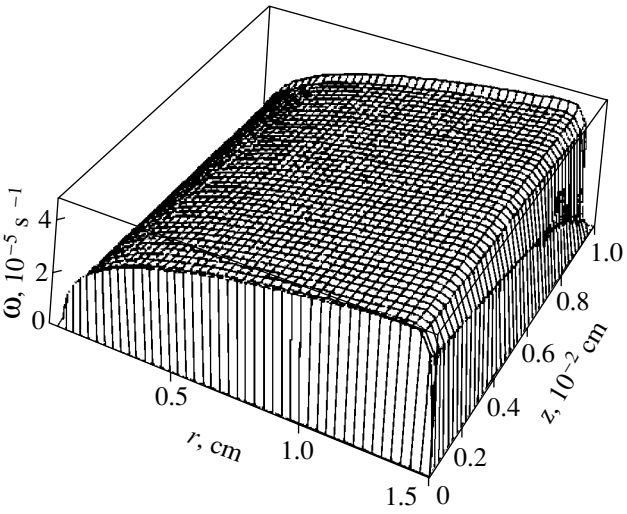


Fig. 2. Angular velocity of rotation versus z and r for MBBA at $\Delta T = 5$°C. The thermomechanical coefficient is $\xi = 0.22 \times 10^{-6}$ dyn/°C.

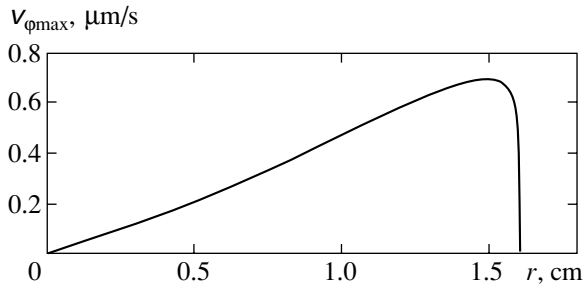


Fig. 3. Maximum rotation velocity (with respect to z) versus distance r from the common axis of cylinders for $\Delta T = 5$°C.

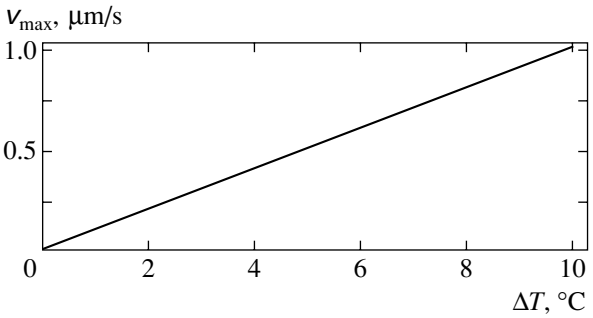


Fig. 4. Maximum rotation velocity (with respect to z and r) versus temperature difference ΔT.

4. DISCUSSION AND CONCLUSIONS

The present theoretical treatment of thermally driven rotation of a cylindrical hybrid-aligned nematic cell is in qualitative agreement with experimental observations [12]. We should note here that the thermomechanical stress tensor proposed in [7] is invariant under time reversal, as correctly reasoned in [8]. Accordingly, we concur with the authors of [8] in that thermomechanical coupling does not involve dissipation. In other words, the work done by thermomechanical forces is converted into the kinetic energy of a rotational flow (e.g., via direct thermomechanical effect) rather than dissipated into heat (internal energy of the system). Even though the change in energy density due to thermomechanical forces is caused by external sources, it is reversible. Entropy may vary, and a "dissipative" function associated with thermomechanical coupling will not be invariant under time reversal. However, such a function was used in [7] to derive exact equations describing thermomechanical effects. Consider the entropy equation

$$\partial s/\partial t + \mathrm{div}(s\mathbf{v} + \mathbf{q}/T) = 2\mathcal{R}/T,$$

where q/T is heat flux. The function $\mathcal{R}$ is antisymmetric under time reversal (i.e., "dissipative"), but this does not lead to any substantial inconsistency. Indeed, the system analyzed here is an open one, since heat is supplied by an external source, and the entropy of a thermally uninsulated system of this kind may vary.

Note also that the thermomechanical-stress and heat-flux components suggested in [8] are equivalent to the corresponding expressions in [7]. Indeed, a simple algebra can be performed to show that the expressions given in [8] can be obtained by simply rewriting those proposed in [7] in a different form (see Appendix).

In summary, reversible thermomechanical coupling does exist. It has been observed experimentally, and its analysis can provide information about molecular dynamics, elastic properties, and structure of nematic mesophase.

ACKNOWLEDGMENTS

This work was supported, in part, by CRDF, grant no. AP2-2302-YE-02.

APPENDIX

DIFFERENT REPRESENTATIONS OF THERMOMECHANICAL STRESS TENSOR

We introduce new notation for the thermomechanical coefficients used in [7]:

$$a_1 = \frac{1}{4}(2\xi_1 - \xi_3 + \xi_5 - \xi_7),$$

$$a_2 = -\frac{1}{4}(2\xi_1 - \xi_3 - \xi_5 + \xi_7),$$

$$a_3 = -\frac{1}{2}(\xi_2 + \xi_6), \quad a_4 = \frac{1}{2}(\xi_2 - \xi_6),$$

$$a_5 = \frac{1}{2}(\xi_3 + \xi_7), \quad a_6 = -\frac{1}{2}(\xi_3 - \xi_7),$$

$$a_7 = -\frac{1}{2}(\xi_3 - \xi_4 + \xi_7 - \xi_8 + \xi_{11} + 2\xi_{12}),$$

$$a_8 = \frac{1}{2}(\xi_3 - \xi_4 - \xi_7 + \xi_8 - \xi_{11} - 2\xi_{12}),$$

$$a_9 = \frac{1}{2}(-2\xi_5 + \xi_7 + 2\xi_{10} + \xi_{12}),$$

$$a_{10} = \frac{1}{2}(-2\xi_7 + 2\xi_9 + \xi_{11}), \quad a_{11} = \frac{1}{2}\xi_{11}, \quad a_{12} = \xi_{12}.$$

After some manipulation, we obtain the following expression for the thermomechanical stress tensor considered in [7]:

$$\begin{aligned}
\sigma_{ij}^{\text{thm}} &= a_1 n_j \nabla_i T \operatorname{div}\mathbf{n} + a_2 n_i \nabla_j T \operatorname{div}\mathbf{n} \\
&+ \left(a_3 + a_4 + \frac{1}{2}a_{10}\right) n_i n_j (\mathbf{n}\nabla n_q)\nabla_q T \\
&+ \left(a_3 + \frac{1}{2}a_5\right) n_j (\nabla_k n_i)(\nabla_k T) \\
&+ \left(-a_3 + \frac{1}{2}a_5\right) n_j (\nabla_i n_k)(\nabla_k T) \\
&+ \left(-a_3 + \frac{1}{2}a_7\right) n_j (\mathbf{n}\nabla n_i)(\mathbf{n}\nabla T) \\
&+ \left(a_4 + \frac{1}{2}a_6\right) n_i (\nabla_k n_j)(\nabla_k T) \\
&+ \left(-a_4 + \frac{1}{2}a_6\right) n_i (\nabla_j n_k)(\nabla_k T) \\
&+ \left(-a_4 + \frac{1}{2}a_8\right) n_i (\mathbf{n}\nabla n_j)(\mathbf{n}\nabla T) \\
&+ a_9 n_i n_j \operatorname{div}\mathbf{n}(\mathbf{n}\nabla T) + \frac{1}{2}a_{11}(\mathbf{n}\nabla n_i)\nabla_j T \\
&+ \frac{1}{2}a_{11}(\mathbf{n}\nabla n_j)\nabla_i T + \frac{1}{2}a_{12}\nabla_i n_j (\mathbf{n}\nabla T) \\
&+ \frac{1}{2}a_{12}\nabla_j n_i (\mathbf{n}\nabla T).
\end{aligned} \tag{5}$$

According to [8], the new components of thermomechanical stress have the form $\sigma_{ij}^R = a_{ijklp}(\nabla_k T)(\nabla_l n_p)$, where α_{ijklp} is expressed as follows (if no symmetry is imposed on the tensor):

$$\begin{aligned}
a_{ijkpl} &= a_1 n_i \delta_{jk}^{tr}\delta_{lp}^{tr} + a_1'' n_j \delta_{ik}^{tr}\delta_{lp}^{tr} + a_2 n_k \delta_{ij}^{tr}\delta_{lp}^{tr} \\
&+ a_3 n_i \delta_{jl}^{tr}\delta_{kp}^{tr} + a_3'' n_j \delta_{il}^{tr}\delta_{kp}^{tr} + a_4 n_l \delta_{ij}^{tr}\delta_{kp}^{tr} + a_5 n_i \delta_{jp}^{tr}\delta_{kl}^{tr} \\
&+ a_5'' n_j \delta_{ip}^{tr}\delta_{kl}^{tr} + a_6 n_i n_j n_l \delta_{kp}^{tr} + a_7 n_k \delta_{il}^{tr}\delta_{pj}^{tr} + a_7'' n_k \delta_{ip}^{tr}\delta_{jl}^{tr} \\
&+ a_8 n_i n_j n_k \delta_{pl}^{tr} + a_9 n_l \delta_{ik}^{tr}\delta_{jp}^{tr} + a_9'' n_l \delta_{jk}^{tr}\delta_{ip}^{tr} \\
&+ a_{10} n_j n_k n_l \delta_{ip}^{tr} + a_{10}'' n_i n_k n_l \delta_{jp}^{tr}.
\end{aligned}$$

If the stress tensor is symmetric, then $a_i = a_i''$ ($i = 1, 3, 5, 7, 9, 10$). Using this expression for a_{ijklp}, we perform

some manipulations to obtain the expression for thermomechanical stress σ_{ij}^R given in [8]:

$$\begin{aligned}
\sigma_{ij}^R &= a_1 n_i \nabla_j T \mathrm{div}\mathbf{n} + a_1'' n_j \nabla_i T \mathrm{div}\mathbf{n} \\
&+ [a_2(\mathbf{n}\nabla T)\mathrm{div}\mathbf{n} + a_4 \nabla_k T(\mathbf{n}\nabla n_k)]\delta_{ij} \\
&+ a_3 n_i(\nabla_j n_k)(\nabla_k T) + a_3'' n_j(\nabla_i n_k)(\nabla_k T) \\
&+ a_5 n_i(\nabla_k n_j)(\nabla_k T) + a_5'' n_j(\nabla_k n_i)(\nabla_k T) \\
&+ a_7 \nabla_i n_j(\mathbf{n}\nabla T) + a_7'' \nabla_j n_i(\mathbf{n}\nabla T) \\
&+ a_9(\mathbf{n}\nabla n_j)\nabla_i T + a_9''(\mathbf{n}\nabla n_i)\nabla_j T \\
&+ (a_8 - a_1 - a_1'' - a_2) n_i n_j \mathrm{div}\mathbf{n}(\mathbf{n}\nabla T) \\
&+ (-a_3 - a_3'' - a_4 + a_6) n_i n_j (\mathbf{n}\nabla n_q)\nabla_q T \\
&+ (-a_5 - a_7 - a_9 + a_{10}'') n_i (\mathbf{n}\nabla n_j)(\mathbf{n}\nabla T) \\
&+ (-a_5'' - a_7'' - a_9'' + a_{10}) n_j (\mathbf{n}\nabla n_i)(\mathbf{n}\nabla T).
\end{aligned} \tag{6}$$

Comparison of (5) with (6), we find that each term in (6) is identical to a corresponding term in (5) except for those containing a_2 and a_4, which correspond to the additional pressure

$$p^{\mathrm{thm}} = a_2(\mathbf{n}\nabla T)\mathrm{div}\mathbf{n} + a_4 \nabla_k T(\mathbf{n}\nabla n_k),$$

due to thermomechanical coupling (in [7], it was included in the expression for hydrodynamic pressure). Note that p^{thm} can be found by the method proposed in [7], i.e., by applying the conditional variational principle for $\mathrm{div}\mathbf{v} = 0$.

Note also that the tensor a_{ijklp} in the expression for thermomechanical stress was used in [8] to describe the heat flux $j_k^{\sigma R}$ without sufficient reason. However, the equality of the corresponding tensors can easily be proved by the method proposed in [7]. First, rewrite the "dissipative" function that describes thermomechanical coupling in terms of a_{ijklp}:

$$\mathscr{R}^{\mathrm{thm}} = a_{ijklp}(\nabla_i v_j)(\nabla_k T)(\nabla_l n_p).$$

Then, calculate the variational derivatives of $\mathscr{R}^{\mathrm{thm}}$ in $\nabla_k T$ and $\nabla_i v_j$ respectively, keeping the remaining quantities constant, to obtain

$$q_k^{\mathrm{thm}} = j_k^{\sigma R} T = -T\frac{\delta \mathscr{R}^{\mathrm{thm}}}{\delta(\nabla_k T)} = -a_{ijklp}(\nabla_i v_j)(\nabla_l n_p)T,$$

$$\sigma_{ij}^R = \frac{\delta \mathscr{R}^{\mathrm{thm}}}{\delta(\nabla_i v_j)} = a_{ijklp}(\nabla_k T)(\nabla_l n_p).$$

The resulting coefficients are obviously equal, since they are obtained by varying the same "dissipative" function.

REFERENCES

1. O. Lehmann, Ann. Phys. (Leipzig) **4**, 649 (1900).
2. F. M. Leslie, Proc. R. Soc. London, Ser. A **307**, 359 (1968).
3. I. Janossy, Mol. Cryst. Liq. Cryst. Lett. **72**, 233 (1982).
4. G. S. Ranganath, Mol. Cryst. Liq. Cryst. **92**, 105 (1983).
5. F. M. Leslie, J. Non-Equilib. Thermodyn. **11**, 23 (1986).
6. S. Chandrasekhar, *Liquid Crystals* (Cambridge Univ. Press, Cambridge, 1977; Mir, Moscow, 1980).
7. R. S. Hakobyan and B. Ya. Zel'dovich, Zh. Éksp. Teor. Fiz. **87**, 1660 (1984) [Sov. Phys. JETP **60**, 953 (1984)].
8. H. R. Brand and H. Pleiner, Phys. Rev. A **35**, 3122 (1987).
9. O. D. Lavrentovich and Yu. A. Nastishin, Ukr. Fiz. Zh. **32**, 710 (1987).
10. R. S. Hakobyan, R. B. Alaverdyan, É. A. Santrosyan, and Yu. S. Chilingaryan, Pis'ma Zh. Tekh. Fiz. **23** (17), 77 (1997) [Tech. Phys. Lett. **23**, 690 (1997)].
11. R. S. Hakobyan, R. B. Alaverdyan, S. Ts. Nersisyan, *et al.*, Zh. Tekh. Fiz. **69** (5), 122 (1999) [Tech. Phys. **44**, 462 (1999)].
12. R. S. Akopyan, R. B. Alaverdyan, Yu. S. Chilingaryan, and E. A. Santrosyan, J. Appl. Phys. **90**, 3371 (2001).
13. P. G. de Gennes and J. Prost, *The Physics of Liquid Crystals*, 2nd ed. (Clarendon Press, Oxford, 1993; Mir, Moscow, 1982).
14. R. S. Hakobyan and B. Ya. Zel'dovich, Zh. Éksp. Teor. Fiz. **83**, 2137 (1982) [Sov. Phys. JETP **56**, 1239 (1982)].

Translated by A. Betev

Journal of Experimental and Theoretical Physics, Vol. 99, No. 5, 2004, pp. 1044–1053.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 5, 2004, pp. 1198–1208.

SOLIDS
Structure

Inhomogeneity-Induced Second-Order Phase Transitions in the Potts Model on Hierarchical Lattices

P. N. Timonin

Physics Research Institute, Rostov State University, Rostov-on-Don, 344090 Russia
e-mail: timonin@aaanet.ru

Received February 9, 2004

Abstract—Thermodynamics of the Potts model with an arbitrary number of states is analyzed for a class of hierarchical lattices of fractal dimension $d > 1$. In contrast to the case of crystal lattice, it is shown that all phase transitions on lattices of this type are of the second order. Critical exponents are determined, their dependence on structural parameters is examined, and scaling relations between them are established. A structural criterion for change in transition order is discussed for inhomogeneous systems. Application of the results to critical phenomena in phase transitions in dilute crystals and porous media is discussed. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Analysis of phase transitions in spin models on hierarchical lattices dates back to the Migdal–Kadanoff renormalization group method [1, 2], where such lattices were introduced as approximations of crystal lattices [3]. The variety of hierarchical lattices of noninteger dimension proposed in further studies [4, 5] can be used as models of fractal systems. Analysis of phase transitions on fractals is of great interest, since fractals are frequently used in models of disordered systems, such as porous media or percolation clusters in disordered crystals [6]. Hierarchical lattices are not random, but they have wide distributions of coordination number and characteristic length and can therefore be used as models of random inhomogeneous media. Indeed, studies of Ising models on hierarchical lattices have shown that the critical exponents for second-order phase transitions are different for fractals and depend on their structural characteristics [7, 8]. This behavior is analogous to change in critical exponents for disordered crystals with degree of disorder [9].

Furthermore, dependence of first-order transitions on structural inhomogeneity in hierarchical fractal models is qualitatively similar to that observed in numerical studies of such transitions in models of diluted crystals [10–13] and porous media [14, 15]. In particular, the inhomogeneity-induced crossover from first- to second-order transition discussed in [10–15] occurs in the q-state Potts model (with $q = 4$ and 10) on hierarchical lattices of fractal dimension $d > 2$ [16]. This remarkable observation cannot be explained in the framework of the standard phenomenology that attributes "suppressed" first-order transitions in inhomogeneous systems (partial or complete suppression of jumps in thermodynamic state variables) to the formation of an inhomogeneous two-phase state near the critical point [17]. This explanation is inconsistent with the physical nature of the instability that develops in the inhomogeneous system and manifests itself by divergence of correlation length and critical susceptibility [10–15].

However, experiments on transitions in liquid crystals [18, 19] and the antiferromagnet MnO [20] confined in porous matrices demonstrate the possibility of change in order of transition due to inhomogeneity. A change from the first-order structural transition $O_h \longrightarrow D_{4h}$ to a symmetry-forbidden second-order transition occurs for a magnetite (Fe_3O_4) under doping with zinc [21]. An analogous change is also observed for phase transition from cubic to orthorhombic phase in mixed crystals $(KBr)_{1-x}(KCN)_x$ [22, 23]. Whereas transition of this kind in ideal cubic lattices are of the first order [24], it turns into a second-order one for $x =$ 0.65, 0.7 [22], and 0.73 [23] as the elastic modulus C_{44} vanishes at the critical point [23].

Thus, both experiments and numerical analysis of realistic models demonstrate that inhomogeneities not only suppress jumps in thermodynamic state variables, but also lead to second-order singular behavior. Elucidation of the nature of this phenomenon and development of its quantitative models will improve understanding of phase-transition mechanisms and stimulate progress in many applications of inhomogeneous materials. One goal of theoretical analysis is to discriminate between systems characterized by change from first- to second-order phase transition and those in which first-order jumps are merely suppressed by the formation of an intermediate inhomogeneous phase [17, 25]. Another goal is to determine the critical exponents for inhomogeneity-induced second-order transitions and find their relationship to characteristics of inhomogeneous structure.

1063-7761/04/9905-1044$26.00

Studies of phase transitions in simplified spin models of inhomogeneous systems, such as hierarchical lattices, can be very helpful in achieving these goals, because thermodynamics of certain models of this kind admits exact analytical treatment [16]. Indeed, analytical evidence of change from first- to second-order phase transition in inhomogeneous systems can currently be found only in the results reported in [16] and in analyses of the Potts model for random graphs with power-law distributions of coordination number in [26, 27].

The Potts model was examined in [10–16, 26, 27] as the simplest one that admits first-order transitions on translationally invariant lattices of dimension $d = 2$ or 3 and represents various physical phenomena, including structural transitions in adsorbate layers ($q = 3, 4$) and transitions in cubic ferromagnets placed in magnetic field or in liquid mixtures ($q = 3$) (see review in [28] and references therein). The Potts model can also be used to describe various ferroelastic transitions, such as the $O_h \longrightarrow D_{4h}$ transitions in the spinels exemplified by $NiCr_2O_4$ and Fe_3O_4 and in the superconductors Nb_3Sn and V_3Si ($q = 3$) or the charge-ordering transitions in Yb_4As_3 crystals [29] and Mg_3Cd-type alloys ($q = 4$) (see Table IV.4 in [24]).

The analysis presented in this paper is focused on the q-state Potts model with arbitrary q on a two-parameter family of hierarchical lattices characterized by fractal dimensions $d > 1$ and average coordination numbers $2 < \bar{z} < 4$. An analytical approach different from that employed in [16] is used to show that second-order phase transitions can occur for all of these q, d, and $\bar{z}$; to obtain expressions for critical exponents; to examine their dependence on structural parameters; and to find scaling relations between them. The results obtained here make it possible to suggest a structural criterion for inhomogeneity-induced change in order of transition. Finally, these results are discussed with regard to their applicability to critical phenomena in phase transitions in dilute crystals and porous media.

2. GEOMETRIC CHARACTERIZATION OF HIERARCHICAL LATTICES

Figure 1 schematizes the construction of a hierarchical lattice of the family specified above by replacing each bond with $n \geq 2$ chains containing $m \geq 2$ bonds. The total number of bonds increases by a factor of $B = mn$ at each step of the procedure, amounting to B^k after the kth step. The number N_k of sites is given by the recursion relation

$$N_k = BN_{k-1} - n - B + 2.$$

Therefore, if $N_0 = 2$, then

$$N_k = \frac{B-n}{B-1}B^k + \frac{n-1}{B-1} + 1. \quad (1)$$

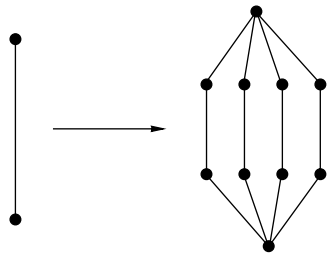


Fig. 1. Iterative construction of hierarchical lattice with $m = 3$ and $n = 4$.

Since the average coordination number for an infinite lattice is $\bar{z} \equiv \lim_{k\to\infty}(2B^k/N_k)$, it follows from (1) that

$$\bar{z} = 2\frac{B-1}{B-n}; \quad (2)$$

i.e., $\bar{z}$ varies from 2 ($m \longrightarrow \infty$) to 4 ($m = 2$, $n \longrightarrow \infty$). After k steps, the largest distance between sites is m^k. Therefore, $N_k \propto (m^k)^d$ as $k \longrightarrow \infty$, where

$$d = \frac{\ln B}{\ln n}$$

is the fractal dimension of the lattice. It is obvious that $1 < d < \infty$.

If $\bar{z}$ and d are treated as independent parameters, then the conditions $m \geq 2$ and $n \geq 2$ imply that

$$2 < \bar{z} < 4(1 - 2^{-d}). \quad (3)$$

To find the distribution of coordination number, note that $z_k = 2n^k$ and the number of such sites in the lth-level lattice is

$$s_k = (B-n)B^{l-k-1}, \quad 0 \leq k \leq l-1.$$

Since $s_l = 2$ ($z_l = n^l$ for the two basic sites), it follows that $\sum_{k=0}^{l} s_k = N_l$ and $\sum_{k=0}^{l} s_k z_k = 2B_l$. The resulting coordination-number distribution is

$$W(z) = \lim_{l\to\infty}\sum_{k=0}^{l}\frac{s_k}{N_l}\delta(z - z_k)$$

$$= \frac{B-1}{B}\left(\frac{2}{z}\right)^{\frac{d}{d-1}}\sum_{k=0}^{\infty}\delta(z - 2n^k);$$

i.e., a power law with an exponent greater than unity. Analogous distributions with denser sequences of z_k ($z_k = k$) are used in models based on iterative scale-free

random graphs [26, 27]. For these graphs, Potts models with $q \geq 1$ admit second-order transitions, with effective-field singularities due to a divergent moment $\langle z^3 \rangle$, if the power exponent in $W(z)$ is less than 3 [26]. In the present model, $\langle z^r \rangle$ is divergent if $r \geq d/(d-1)$; otherwise,

$$\langle z^r \rangle = 2^r \frac{B-1}{B-n^r}.$$

However, it is shown below that all transitions on the hierarchical lattices of the type considered here are of the second order with anomalous scaling, and the form of $W(z)$ determines only the values of critical exponents.

3. RECURSION RELATIONS, GIBBS FREE ENERGY, AND CORRELATIONS IN THE POTTS MODEL

The partition function for the q-state Potts model on the lattices under analysis can be calculated by assigning the factor

$$Z_0(\sigma, \sigma') = \exp\left[K\delta_{\sigma,\sigma'} + \frac{h}{2}(\delta_{\sigma,1} + \delta_{\sigma',1})\right] \quad (4)$$

to each bond and performing the sum of the resulting expressions over $\sigma = \{1, 2, \ldots, q\}$. (Here, σ and σ' are the Potts spins at the bonded sites, $K = J/T$, and h is the external field.) The "partial" partition function $Z_l(\sigma, \sigma')$ calculated for the lth-level lattice as a sum over all spins except those at the basic sites can be shown to satisfy the recursion relation [1–3]

$$Z_{l+1}(\sigma, \sigma') = [(\hat{Z}_l)^m_{\sigma,\sigma'}]^n \exp\frac{h}{2}(1-n)(\delta_{\sigma,1} + \delta_{\sigma',1}). \quad (5)$$

Thus, the next-level partial partition function is obtained by raising $Z_l(\sigma, \sigma')$ to the power m and raising each element of the resulting matrix to the power n. The exponential factor in (5) is introduced to remove spurious powers of $\exp(h/2)(\delta_{\sigma,1} + \delta_{\sigma',1})$. Relations (4) and (5) can be used to find the partition function in the thermodynamic limit as $l \longrightarrow \infty$.

By virtue of (4) and (5), $Z_l(\sigma, \sigma')$ can be represented as

$$Z_l(\sigma, \sigma') = a_{1l}\delta_{\sigma,1}\delta_{\sigma',1} + a_{2l}\frac{(1-\delta_{\sigma,1})(1-\delta_{\sigma',1})}{q-1} + b_l(\delta_{\sigma,1} + \delta_{\sigma',1} + \delta_{\sigma,1}\delta_{\sigma',1}) + c_l\left[\delta_{\sigma,\sigma'} - \delta_{\sigma,1}\delta_{\sigma',1} - \frac{(1-\delta_{\sigma,1})(1-\delta_{\sigma',1})}{q-1}\right]. \quad (6)$$

The matrix $Z_l(\sigma, \sigma')$ given by (6) has two nondegenerate eigenvalues,

$$\lambda_{\pm l} = \frac{1}{2}(a_{1,l} + a_{2,l}) \pm \sqrt{\frac{1}{4}(a_{1,l} - a_{2,l})^2 + (q-1)b_l^2}, \quad (7)$$

and an eigenvalue c_l of multiplicity $q-2$. Expression (7) can be used to represent the coefficients obtained when $Z_l(\sigma, \sigma')$ is raised to the power m as follows:

$$a'_{1l} = \frac{1}{2}(\lambda^m_{+l} + \lambda^m_{-l}) + \frac{\zeta_1}{2}(a_{1l} - a_{2l}),$$
$$a'_{2l} = \frac{1}{2}(\lambda^m_{+l} + \lambda^m_{-l}) - \frac{\zeta_1}{2}(a_{1l} - a_{2l}), \quad (8)$$
$$b'_l = \zeta_l b_l, \quad c'_l = c^m_l, \quad \zeta_l \equiv \frac{\lambda^m_{+l} - \lambda^m_{-l}}{\lambda_{+l} - \lambda_{-l}}.$$

Accordingly, the coefficients satisfy recursion relations corresponding to (5):

$$a_{1,l+1} = e^{-h(n-1)}(a'_{1l})^n, \quad b_{l+1} = e^{-h(n-1)/2}(b'_l)^n,$$
$$a_{2,l+1} = \left(\frac{a'_{2l} + (q-2)c'_l}{q-1}\right)^n + (q-2)\left(\frac{a'_{2l} - c'_l}{q-1}\right)^n, \quad (9)$$
$$c_{l+1} = \left(\frac{a'_{2l} + (q-2)c'_l}{q-1}\right)^n - \left(\frac{a'_{2l} - c'_l}{q-1}\right)^n.$$

The starting values of the coefficients are derived from (4):

$$a_{1,0} = e^{K+h}, \quad a_{2,0} = e^K + q - 2, \quad b_0 = e^{h/2}, \quad c_0 = e^K - 1. \quad (10)$$

The partition functions for the Potts models with particular boundary conditions at the basic sites are found by solving system (8), (9). In the case of free boundary conditions, the following result is obtained by adding the required fields $h/2$ at the basic sites:

$$Z^{(f)}_l = \sum_{\sigma,\sigma'} e^{h(\sigma+\sigma')/2} Z_l(\sigma, \sigma') = e^h a_{1l} + (q-1)(a_{2l} + 2e^{h/2}b_l). \quad (11)$$

If periodic boundary conditions are set at the basic sites, then

$$Z^{(p)}_l = \sum_{\sigma} Z_l(\sigma, \sigma) = a_{1l} + a_{2l} + (q-2)c_l. \quad (12)$$

If boundary conditions specify that $\sigma = 1$ at the basic sites, then

$$Z^{(1)}_l = Z_l(1, 1) = a_{1l}. \quad (13)$$

In the absence of long-range ordering, all of these partition functions must yield the same Gibbs free-

energy density in the thermodynamic limit. They are very difficult to calculate for $h \neq 0$. However, both Gibbs free energy and its field derivatives can be obtained by analytical methods for $h = 0$ near the critical point, and this will suffice to determine the order of transition and the critical exponents.

In this section, an expression for the Gibbs free energy for $h = 0$ is derived. In this case, it follows from (8) and (9) that

$$a_{1l} = b_l + c_l, \quad a_{2l} = a_{1l} + (q-2)c_l,$$

i.e., only two coefficients are independent. By introducing

$$K_l \equiv \ln\frac{a_{1l}}{b_l},$$

a well-known relation is obtained [30]:

$$e^{K_{l+1}} = f(e^{K_l}),$$
$$f(x) \equiv \left[1 + \frac{q(x-1)^m}{(x-1+q)^m - (x-1)^m}\right]^n. \tag{14}$$

Another recursion relation valid for $h = 0$ is

$$b_{l+1} = g_{l+1} b_l^B, \quad g_l \equiv \frac{(e^{K_{l-1}} - 1)^B}{(e^{K_l/n} - 1)^n}. \tag{15}$$

By combining (15) with the definition of K_l, the following expression for the Gibbs free-energy density is obtained:

$$F = -T\lim_{l\to\infty} N_l^{-1} \ln Z_l^{(f)} = -\frac{\bar{z}}{2} T \sum_{k=1}^{\infty} B^{-k} \ln g_k, \tag{16}$$

where use is made of the fact that $K_l B^{-l} \longrightarrow 0$ as $l \longrightarrow \infty$.

The fixed point $K = K_c$ of relation (14),

$$e^{K_c} = f(e^{K_c}), \tag{17}$$

is the critical point. If $K > K_c$, then $K_l \longrightarrow \infty$. If $K < K_c$, then $K_l \longrightarrow 0$. If $|K - K_c| \ll K_c$ and l is sufficiently small, then K_l varies slowly,

$$e^{K_l - K_c} \approx 1 + \kappa^l (K - K_c), \tag{18}$$

$$\kappa \equiv f'(e^{K_c}) = B\frac{(e^{K_c} - e^{K_c/n})(e^{K_c/n} + q - 1)}{(e^{K_c} - 1)(e^{K_c} + q - 1)} < B, \tag{19}$$

until it becomes much greater or much smaller than K_c.

An applicability condition for (18) can be written as

$$l < l_c \equiv \ln\frac{\text{const}}{|K - K_c| \ln\kappa}, \tag{20}$$

where the constant is determined from (18) by using requiring that $K_{l_c} \sim K_c$ at $K < K_c$ and $K_{l_c} \gg K_c$ at $K > K_c$.

For $l > l_c$, (14) yields

$$K_{l+1} \approx nq\left(\frac{K_l}{q}\right)^m,$$

$$K_l \approx qn^{-1/(m-1)}\left(\frac{K_{l_c} n^{1/(m-1)}}{q}\right)^{m^{l-l_c}}, \quad K < K_c,$$

$$\exp K_{l+1} \approx \left(\frac{\exp K_l}{m}\right)^n, \tag{21}$$

$$\exp K_l \approx m^{n/(n-1)} (m^{-n/(n-1)} \exp K_{l_c})^{n^{l-l_c}}, \quad K > K_c.$$

By virtue of (18)–(21), the Gibbs free-energy density F given by (16) has a singular part that scales as $B^{-l_c} \propto |K - K_c|^{2-\alpha}$, with the specific-heat critical exponent

$$\alpha = 2 - \frac{\ln B}{\ln\kappa}. \tag{22}$$

In particular, if $K > K_c$ and $|K - K_c| \sim K_c$, then expression (16) with

$$g_k \approx \begin{cases} g_c \equiv (e^{K_c} - 1)^B / (e^{K_c/n} - 1)^n, & l < l_c, \\ g_\infty \equiv (q^{m-1}/n)^n, & l > l_c \end{cases}$$

yields

$$F \approx -\frac{\bar{z}}{2(B-1)} T[g_c + B^{-l_c}(g_\infty - g_c)].$$

For $h = 0$, the correlation function

$$G = \langle \delta_{\sigma,1}\delta_{\sigma',1}\rangle - \langle \delta_{\sigma,1}\rangle\langle \delta_{\sigma',1}\rangle$$

is readily obtained for the basic spins. For the lth-level lattice,

$$G_l = \frac{Z_l(1,1)}{Z_l^f} - \left(\frac{\sum_\sigma Z_l(\sigma,1)}{Z_l^f}\right)^2 = \frac{q-1}{q^2}\frac{e^{K_l} - 1}{e^{K_l} + q - 1}.$$

If $K < K_c$ and $l > l_c$, then the first equation in (21) yields

$$G_l \approx \frac{K_l}{q^3} \approx q^{-2} n^{-1/(m-1)}\left(\frac{K_{l_c} n^{1/(m-1)}}{q}\right)^{m^{l-l_c}} \sim \exp\left(-\frac{m^l}{\xi}\right),$$

where the correlation length is

$$\xi \sim m^{l_c} \sim (K_c - K)^{-\nu}, \quad \nu = \frac{\ln m}{\ln \kappa}. \tag{23}$$

The critical exponent for correlation length satisfies the scaling relation

$$d\nu = 2 - \alpha.$$

If $K > K_c$, then the exponential scaling is not valid, and the second relation in (21) implies that G_l is asymptotically constant:

$$G_l \approx \frac{q-1}{q^2} + A\exp(-\text{const} n^{l-l_c}).$$

However, the characteristic change in G_l scales as m^{l_c}, because $n^{l-l_c} = (m^l/m^{l_c})^{d-1}$.

Thus, the specific heat and correlation length at the critical point of the Potts model under analysis exhibit power-law singular behavior satisfying a standard scaling relation. Anomalous scalings of order parameter and susceptibility are considered in the next section.

4. ORDER PARAMETER AND CRITICAL SUSCEPTIBILITY

In the present Potts model, the spontaneous order parameter is expressed as

$$\mu = \lim_{l\to\infty} \frac{qN_l^{-1}\sum_{i=1}^{N_l-1}\langle\delta_{\sigma_i,1}\rangle - 1}{q-1}, \tag{24}$$

where $\langle\delta_{\sigma,1}\rangle$ is the average calculated under the boundary conditions with $\sigma = 1$ at the basic sites, i.e.,

$$\sum_{i=1}^{N_l-1}\langle\delta_{\sigma_i,1}\rangle = \frac{1}{Z_l^{(1)}}\frac{\partial Z_l^{(1)}}{\partial h}\bigg|_{h=0} \equiv \frac{\dot{Z}_l^{(1)}}{Z_l^{(1)}}. \tag{25}$$

Note that nonzero μ at $K > K_c$ and $h = 0$ can be obtained only under these symmetry-breaking boundary conditions. Indeed, $\langle\delta_{\sigma_i,1}\rangle = 1/q$ for zero field under the free or periodic boundary conditions by virtue of the symmetry with respect to permutation of the values of σ. Hence,

$$\dot{Z}_l^{(f)} = \frac{N_l Z_l^{(f)}}{q}, \quad \dot{Z}_l^{(p)} = \frac{(N_l-1)Z_l^{(p)}}{q}.$$

Substituting (11) and (12) into these equations, one obtains

$$\begin{aligned}\dot{\lambda}_{+l} &= \frac{(N_l-1)\lambda_{+l}}{q},\\ \dot{\lambda}_{-l} + (q-2)\dot{c}_l &= \frac{(N_l-1)(q-1)\lambda_{-l}}{q},\end{aligned} \tag{26}$$

where $\dot{\lambda}_{\pm l}$ are the field derivatives of eigenvalues (7) for $h = 0$,

$$\dot{\lambda}_{\pm l} = \frac{1}{2}(\dot{a}_{1l} + \dot{a}_{2l}) \pm \frac{1}{2q}[(2-q)(\dot{a}_{1l} - \dot{a}_{2l}) + 2(q-1)\dot{b}_l].$$

It follows from (13) and (24)–(26) that

$$\begin{aligned}\mu &= \lim_{l\to\infty}\frac{qN_l^{-1}\varphi_{+l} + q - 2}{2(q-1)},\\ \varphi_{+l} &= \frac{\dot{a}_{1l} - \dot{a}_{2l} - (q-2)\dot{c}_l}{a_{1l}}.\end{aligned} \tag{27}$$

Define another combination of derivatives independent of φ_{+l} and the left-hand sides of (26):

$$\varphi_{-l} = \frac{\dot{a}_{1l} - \dot{a}_{2l} + q\dot{c}_l}{a_{1l}}.$$

Using the derivatives of (9) with respect to h and Eqs. (26), one obtains the following recursion relation for the vector $\boldsymbol{\varphi}_l = (\varphi_{+l}, \varphi_{-l})$:

$$\boldsymbol{\varphi}_l = \hat{T}_l\boldsymbol{\varphi}_{l-1} + \mathbf{u}_l, \tag{28}$$

where

$$\begin{aligned}u_{+l} &= \frac{q-2}{q^2}n(N_{l-1}-1)[(q-2)(m-e_l)\\ &\quad -2(q-1)(m-1)e^{-K_l/n}] - n + 1,\end{aligned} \tag{29}$$

$$u_{-l} = (N_l - 1)\left(1 - \frac{e_l'}{e_l}\right) + \frac{e_l'}{e_l}u_{+l}, \tag{30}$$

$$\hat{T}_l = \frac{n}{q}\begin{pmatrix} e_l[2 + (q-2)m\vartheta_l] & e_l(q-2)(1-m\vartheta_1)\\ 2e_l'(1-m\vartheta_l) & e_l'(q-2+2m\vartheta_l)\end{pmatrix}, \tag{31}$$

$$e_l \equiv \exp\left(K_{l-1} - \frac{K_l}{n}\right),$$

$$e_l' \equiv \exp(K_{l-1} - K_l), \tag{32}$$

$$\vartheta_l \equiv \left[\exp\left(\frac{K_l}{n}\right) - 1\right]/[\exp(K_{l-1}) - 1].$$

The solution to Eq. (28) has the form

$$\begin{aligned}\varphi_l &= \hat{T}_l\hat{T}_{l-1}\ldots\hat{T}_1\varphi_0\\ &\quad + \sum_{k=1}^{l-1}\hat{T}_l\hat{T}_{l-1}\ldots\hat{T}_{k+1}\mathbf{u}_k + \mathbf{u}_l,\end{aligned} \tag{33}$$

where $\varphi_0 = (1, 1)$.

To examine the asymptotic behavior of φ_l as $l \longrightarrow \infty$ near the critical point, represent $\hat{T}_l$ approximately as

$$\hat{T}_l \approx \begin{cases} \hat{T}_c \equiv \lim\limits_{K_l \to K_c} \hat{T}_l, & l < l_c, \\ \hat{T}_\infty \equiv \lim\limits_{l \to \infty} \hat{T}_l, & l > l_c. \end{cases} \tag{34}$$

Then,

$$\begin{aligned} \varphi_l &\approx \hat{T}_\infty^{l-l_c} \hat{T}_c^{l_c} \left[\varphi_0 + \sum_{k=1}^{l_c} (B\hat{T}_c^{-1})^k \mathbf{u}_c \right] \\ &+ \sum_{k=l_c+1}^{l} B^k \hat{T}_\infty^{l-k} \mathbf{u}_\infty \\ &= \hat{T}_\infty^{l-l_c} \hat{T}_c^{l_c} (\varphi_0 - \varphi_c) + B^{l_c} \hat{T}_\infty^{l-l_c} \varphi_c \\ &+ B^l [\hat{I} - (B^{-1}\hat{T}_\infty)^{l-l_c}] \varphi_\infty, \end{aligned} \tag{35}$$

where

$$\mathbf{u}_c = \lim_{l \to \infty} \lim_{K_l \to K_c} \mathbf{u}_l B^{-l}, \quad \mathbf{u}_\infty = \lim_{l \to \infty} \mathbf{u}_l B^{-l},$$

$$\varphi_c \equiv (\hat{I} - B^{-1}\hat{T}_c)^{-1} \mathbf{u}_c, \quad \varphi_\infty \equiv (\hat{I} - B^{-l}\hat{T}_\infty)^{-1} \mathbf{u}_\infty.$$

It follows from (29)–(32) that

$$\hat{T}_c = \frac{n}{q} \begin{pmatrix} e_c[2 + (q-2)m\vartheta_c] & e_c(q-2)(1 - m\vartheta_c) \\ 2(1 - m\vartheta_c) & (q - 2 + 2m\vartheta_c) \end{pmatrix}, \tag{36}$$

$$\begin{aligned} e_c &= \exp\left[K_c \frac{n-1}{n} \right], \\ \vartheta_c &= \left[\exp\left(\frac{K_c}{n} \right) - 1 \right] / [\exp(K_c) - 1], \\ \varphi_{c+} &= -\frac{2}{\bar{z}} \frac{q-2}{q}, \\ \varphi_{c-} &= -\frac{2}{\bar{z}} \left(1 - 2e^{-K_c} \frac{q-1}{q} \right). \end{aligned} \tag{37}$$

For the analysis that follows, it is important that the eigenvalues of the matrix $\hat{T}_c$ are real and less than B for $m \geq 2$, $n \geq 2$, $q > 0$. Indeed, the largest eigenvalue of $\hat{T}_c$ can be expressed as

$$\begin{aligned} \frac{2\lambda_{\max}}{n} &= m + \varepsilon - \rho + \sqrt{(m + \varepsilon - \rho)^2 - 4m\varepsilon}, \\ \varepsilon &\equiv e_c \vartheta_c < 1, \\ \rho &\equiv \frac{2}{q}(1 - \varepsilon) \left[m e^{K_c/n} - e^{K_c} + \frac{q-2}{2}(m-1) \right]. \end{aligned} \tag{38}$$

Equation (17) for K_c can be used to show that $0 < \rho < (m-1)(1-\varepsilon)$ for $m \geq 2$, $n \geq 2$, and $q > 0$, which implies that the eigenvalues of $\hat{T}_c$ are real and $\lambda_{\max} < B$.

The expressions for $\hat{T}_\infty$ and $\mathbf{u}_\infty$ are different for the ordered and disordered phases. In particular, when $K < K_c$, it follows from (21), (31), (32), and (34) that

$$\hat{T}_\infty = \frac{n}{q} \begin{pmatrix} 2 & q-2 \\ 2 & q-2 \end{pmatrix},$$

$$\mathbf{u}_\infty = -\frac{q-2}{q} \frac{m-1}{m} \frac{2}{\bar{z}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Since the eigenvalues of $\hat{T}_\infty$ (0 and n) and $\hat{T}_c$ are less than B, expression (35) reduces to the following as $l \longrightarrow \infty$:

$$\varphi_l \approx B^l \varphi_\infty \approx -\frac{q-2}{q} N_l \varphi_0.$$

Therefore, the order parameter given by (27) vanishes for the disordered phase.

When $K > K_c$, it holds that

$$\hat{T}_\infty = \begin{pmatrix} B & 0 \\ 0 & 0 \end{pmatrix}, \quad u_\infty = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Therefore, $\mathbf{u}_\infty(\varphi_\infty)$ vanishes in (35), and

$$\begin{aligned} \varphi_{+l} &= B^{l-l_c} e_+ \hat{T}_c^{l_c} (\varphi_0 - \varphi_c) + B^l \varphi_{c+} \\ &\approx N_l \left[\text{const} \left(\frac{\lambda_{\max}}{B} \right)^{l_c} - \frac{q-2}{q} \right]. \end{aligned}$$

Accordingly, the order parameter given by (27) is

$$\begin{aligned} \mu &\sim \left(\frac{\lambda_{\max}}{B} \right)^{l_c} \sim (K - K_c)^\beta, \\ \beta &= \frac{\ln(B/\lambda_{\max})}{\ln \kappa} > 0. \end{aligned} \tag{39}$$

To examine the behavior of critical susceptibility for $h = 0$ near the critical point, consider the expression

$$\begin{aligned} \chi &\equiv \lim_{l \to \infty} N_l^{-1} \left[\frac{\ddot{Z}_l^{(f)}}{Z_l^{(f)}} - \left(\frac{\dot{Z}_l^{(f)}}{Z_l^{(f)}} \right)^2 \right] \\ &= \lim_{l \to \infty} \left(N_l^{-1} \psi_{+l} - \frac{N_l}{q^2} \right) + \frac{2}{q}, \end{aligned} \tag{40}$$

which is obtained by using Eqs. (26) and introducing

$$\psi_{+l} = [\ddot{a}_{1l} + (q-1)(\ddot{a}_{2l} + 2\ddot{b}_l)]/q\lambda_{+l}.$$

Define another linear combination of second-order derivatives,

$$\psi_{-l} = [(q-1)(\ddot{a}_{1l} - 2\ddot{b}_l) + \ddot{a}_{2l} + q(q-2)\ddot{c}_l]/q\lambda_{-l},$$

and differentiate Eqs. (9) with respect to h to obtain the following equation for the vector $\boldsymbol{\psi}_l = (\psi_{+l}, \psi_{-l})$:

$$\boldsymbol{\psi}_l = \hat{P}_l \boldsymbol{\psi}_{l-1} + \mathbf{v}_l, \tag{41}$$

where

$$\hat{P}_l = B\begin{pmatrix} x_l & (1-x_l)/(q-1) \\ (q-1)y_l & 1-y_l \end{pmatrix},$$

$$x_l = (e^{K_l/n} + q - 1)\frac{e^{K_l(n-1)/n} + q - 1}{e^{K_l} + q - 1}, \tag{42}$$

$$y_l = (e^{K_l/n} + q - 1)\frac{e^{K_l(n-1)/n} - 1}{e^{K_l} - 1},$$

and

$$\mathbf{v}_l = \frac{B-1}{B}\frac{N_l^2}{q^2}\begin{pmatrix} 1 \\ q-1 \end{pmatrix} + \mathbf{c}\lambda_{\max}^{2l} + O(N_l) \tag{43}$$

for $|K - K_c| \sim K_c$ and large l. The matrix $\hat{P}_l$ has the eigenvalues B and $B(x_l - y_l) < B$. In the expression for $\mathbf{v}_l$, the term proportional to N_l^2 is the right eigenvector of each $\hat{P}_l$ that corresponds to B. Combined with an expression for $\hat{P}_l$ analogous to (34), this fact is used to derive the critical scaling from (40)–(43):

$$\chi \sim \left(\frac{\lambda_{\max}^2}{B}\right)^{l_c} \sim |K - K_c|^{-\gamma}, \qquad \gamma = \frac{2\ln\lambda_{\max}\ln B}{\ln\kappa}. \tag{44}$$

It is obvious that the standard scaling relation is valid:

$$\alpha + 2\beta + \gamma = 2.$$

Thus, the Potts model admits only second-order phase transition with power-law anomalous scaling of thermodynamic parameters.

5. CRITICAL EXPONENTS

Let us consider the dependence of the critical exponents on lattice parameters and q. The inequalities $\kappa < B$ and $\lambda_{\max} < B$ (see (19) and (38)) combined with the scaling relations lead to the following constraints for the exponents in (22), (23), (39), and (44):

$$\nu > 1/d, \quad \alpha < 1, \quad \beta > 0, \quad \gamma > 1.$$

As $q \longrightarrow \infty$, the exponents approach their limit values in these inequalities. Indeed, if $q \gg 1$, then (17), (19), and (38) yield $K_c \approx 2/\bar{z}\ln q$, $\kappa \approx B$, and $\lambda_{\max} \approx B_\varepsilon$. Accordingly,

$$\nu \approx 1/d, \quad \alpha \approx 1, \quad \beta \approx 1/q^{2/\bar{z}n}\ln B, \quad \gamma \approx 1.$$

As $n \longrightarrow \infty$ ($d \longrightarrow \infty$), it holds that $K_c \approx qn^{-1/(m-1)}$, $\kappa \approx m$, and $\lambda_{\max} \approx n$ and the expressions for the exponents reduce to

$$\nu \approx 1, \quad \alpha \approx 2 - d, \quad \beta \approx 1, \quad \gamma \approx d - 2.$$

In the limit of $m \longrightarrow \infty$ ($d \longrightarrow 1$), one obtains $K_c \approx n\ln m/(n-1)$, $\kappa \approx n$, $\rho \approx qm^{(n-3)/(n-1)}/6$, and $\lambda_{\max} \approx B(1 - \rho/m)$; hence,

$$\nu \approx \frac{1}{d-1}, \quad \alpha \approx \frac{d-2}{d-1}, \quad \beta \approx \frac{q}{6m^{2/(n-1)}\ln n},$$

$$\gamma \approx \frac{d}{d-1}.$$

Note that only the exponent β depends on q in these limit cases, and only if it is small. In the general case, the exponents (considered as functions of the measurable parameters of real fractals, d and $\bar{z}$) weakly depend on the average coordination number if $2.5 < \bar{z} < 4$. Figure 2 shows the critical exponents as functions of d on the physical interval $1.5 < d < 3$ for several values of q and $\bar{z} = 2.5$. The case of $q = 1$ corresponds to the anomalous scaling at the threshold for percolation over randomly distributed bonds with probability density $p = 1 - e^{-K}$ [31]. The equation for the percolation threshold, $p_c = 1 - e^{-K_c}$, follows from (17):

$$p_c = 1 - (1 - p_c^m)^n.$$

It should be noted that, whereas the critical exponents in the Potts model on fractals illustrated by Fig. 2 vary significantly, the actually observed values of the exponents exhibit certain trends. In particular, the specific-heat exponent α is usually negative, the susceptibility exponent is anomalously large ($\gamma \geq 1.7$), and the correlation-length exponent monotonically decreases with increasing d while β increases. Furthermore, all of them are monotonic functions of q, except for β. These characteristics of the critical exponents may also hold in other spin models on fractal lattices with sufficiently small d.

Note also that the numerical values of α and β corresponding to $m = 2$ and certain n and q have been cal-

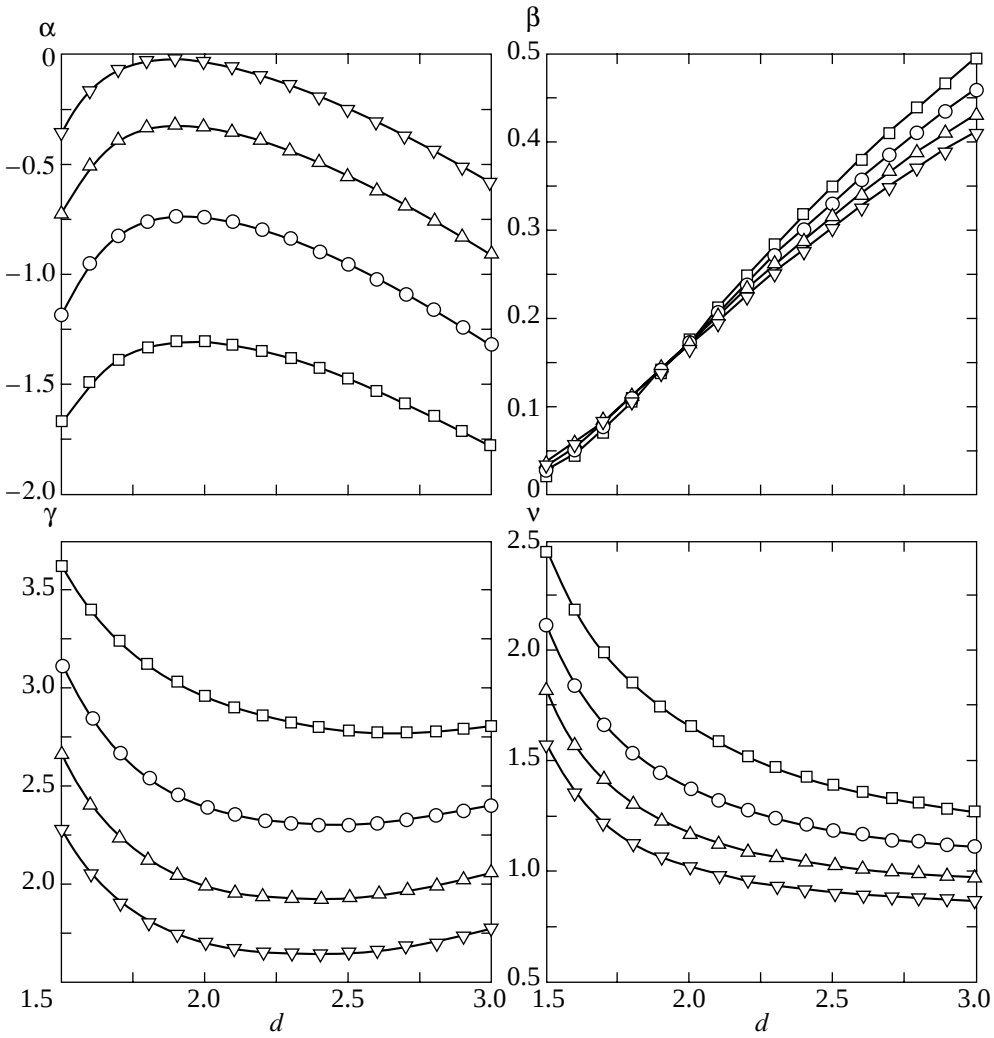


Fig. 2. Critical exponents versus fractal dimension for Potts models on hierarchical lattices with $\bar{z} = 2.5$ and $q = 1$ (□), 2 (○), 4 (△), and 8 (▽).

culated in an earlier study [16]. Similar values of α and β are given by analytical expressions (22) and (39), respectively, within numerical accuracy.

6. CRITERION FOR CHANGE IN ORDER OF TRANSITION

If hierarchical lattices are considered as models of inhomogeneous systems, such as percolation clusters in dilute crystals or materials confined in porous matrices, then certain conclusions can be made about the geometric characteristics of real inhomogeneous media that are responsible for change in the order of phase transition. Indeed, this phenomenon is independent of the fractal dimension for lattices of the type considered here in the entire interval $1 < d < \infty$. It can be speculated that the occurrence of a second-order transition is due to the small average coordination number ($\bar{z} < 4$). Analogous behavior is obtained for diluted models on square lattices: $\bar{z} < 4$ at any dilution while the dimension $d = 2$ of the largest cluster holds until the percolation threshold is reached [6]. Therefore, the change from first- to second-order transition occurs even if the impurity concentration is arbitrarily small [10, 11, 32]. On the other hand, first-order transition is suppressed in the model of an inhomogeneous system with $\bar{z} \geq 4$ instead of changing to second-order transition [25].

Apparently, the condition $\bar{z} < 4$ for the average coordination number can be suggested as a criterion for change from first- to second-order phase transition in a variety of inhomogeneous systems with short-range interactions. In particular, for diluted models on a simple cubic lattice ($z = 6$) with a vacancy concentration of $1 - x$, the average coordination number in the largest cluster can be roughly estimated as $\bar{z} = 6x$. By condition $\bar{z} < 4$, this implies that second-order phase transi-

tion must occur if $x < 2/3$, in agreement with the numerical estimate $x < 0.7 \pm 0.05$ [12, 13].

The existence of a threshold value of the average coordination number can be qualitatively explained as follows. An inhomogeneous lattice with low connectivity (small $\bar{z}$) does not contain macroscopic regions of maximal coordination number. Therefore, as the critical point is approached, macroscopic regions of ordered phase are unlikely to form for energy reasons, while correlations of order-parameter fluctuations grow, as in second-order transition, which leads to singularities in thermodynamic parameters. Conversely, a progressively increasing number of macroscopic regions of ordered phase can form in systems with large $\bar{z}$, as in the phenomenological scenario discussed in [17]. In this case, the correlation length of order-parameter fluctuations remains finite in both ordered and disordered phases, and no second-order singularity arises, even though first-order jumps may completely vanish in highly inhomogeneous systems [25].

7. CONCLUSIONS

The results obtained in this study can be used to explain anomalous scaling in systems described by the Potts model, such as percolation clusters in dilute crystals or materials confined in porous matrices with fractal properties similar to those of the hierarchical lattices considered here. This is corroborated by the fact that $\alpha < 0$ for $q \le 10$, which corresponds to an inequality established rigorously for random inhomogeneous systems [33]. The analytical expressions obtained above can be compared with experiment for particular values of m and n or d and $\bar{z}$.

The critical exponents weakly depend on $\bar{z}$ ($2.5 < \bar{z} < 4$) and strongly depend on the fractal dimension. This explains their slow variation with impurity concentration for dilute crystals [10–13], since the fractal dimension of the percolation clusters that undergo inhomogeneity-induced phase transition remains almost constant until the percolation threshold is reached [6]. However, the exponents for materials confined in porous matrices should be expected to vary more significantly with the fractal dimensions of porous media [6].

The predicted exponents cannot be compared with experiment. No detailed data are currently available on anomalous scaling in inhomogeneity-induced second-order transitions, mainly because their fundamental nature has been demonstrated only in recent theoretical studies [10–15]. However, numerous examples of second-order transitions have long since been discovered experimentally in crystals for which first-order transitions are predicted by the Landau theory of phase transitions [24]. According to [10–15], these findings should be attributed to the presence of impurities or defects. For example, the ferroelastic transitions in Nb_3Sn and V_3Si crystals [34] were distinguished from second-order transitions only for sufficiently pure specimens [35].

Finally, note that inhomogeneity-induced second-order transition may not be specific to the Potts model analyzed in [10–16] and the present study. It should be associated with every first-order transition to a subgroup of the high-symmetry phase. (For ideal cubic lattices, transitions of this type are described by the Landau potential with a cubic invariant [24].) However, the feasibility of this phenomenon in the absence of group–subgroup relation between the symmetries of the ordered and disordered phases (i.e., for reconstructive first-order phase transitions [24]) cannot be corroborated by any experimental or theoretical evidence. Apparently, second-order singular behavior induced by structural inhomogeneity is impossible in this case.

ACKNOWLEDGMENTS

This work was supported by INTAS, grant no. 2001-0826, and the Russian Foundation for Basic Research, project no. 04-02-16228. I thank V.P. Sakhnenko, V.I. Torgashev, V.B. Shirokov, M.P. Ivliev, and E.I. Gutlyanskiĭ for helpful discussions.

REFERENCES

1. A. B. Migdal, Zh. Éksp. Teor. Fiz. **69**, 1457 (1975) [Sov. Phys. JETP **42**, 743 (1975)].
2. L. P. Kadanoff, Ann. Phys. (N.Y.) **100**, 359 (1976).
3. A. N. Berker and S. Ostlund, J. Phys. C **12**, 4961 (1979).
4. M. Kaufman and R. B. Griffiths, Phys. Rev. B **24**, 496 (1981).
5. R. B. Griffiths and M. Kaufman, Phys. Rev. B **26**, 5022 (1982).
6. J. Feder, *Fractals* (Plenum, New York, 1988; Mir, Moscow, 1991).
7. M. Kaufman and R. B. Griffiths, Phys. Rev. B **30**, 244 (1984).
8. Z. R. Yang, Phys. Rev. B **38**, 728 (1988).
9. R. Folk, Y. Holovatch, and T. Yavors'kii, cond-mat/0106468; Usp. Fiz. Nauk **173**, 175 (2003) [Phys. Usp. **46**, 169 (2003)].
10. J. Cardy and J. L. Jacobsen, Phys. Rev. Lett. **79**, 4063 (1997).
11. B. Berche and C. Chatelain, Phys. Rev. Lett. **80**, 1670 (1998); cond-mat/0207421.
12. C. Chatelain, B. Berche, W. Janke, and P.-E. Berche, Phys. Rev. E **64**, 036120 (2001); W. Janke, P.-E. Berche, C. Chatelain, and B. Berche, cond-mat/0304642.
13. H. G. Ballesteros, L. A. Fernandez, V. Martin-Mayor, *et al.*, Phys. Rev. B **61**, 3215 (2000).
14. K. Uzelac, A. Hasmy, and R. Jullien, Phys. Rev. Lett. **74**, 422 (1995).
15. K. Venul, V. S. S. Sastry, and K. P. N. Murthy, cond-mat/0106482.

16. L. da Silva, E. M. F. Curado, S. Coutinho, and W. A. M. Morgado, Phys. Rev. B **53**, 6345 (1996).
17. Y. Imry and M. Wortis, Phys. Rev. B **19**, 3580 (1979).
18. A. Golemme, S. Zumer, D. W. Allender, and J. W. Doane, Phys. Rev. Lett. **61**, 2937 (1988).
19. G. S. Iannachione, G. P. Crawford, S. Zumer, *et al.*, Phys. Rev. Lett. **71**, 2595 (1993).
20. I. V. Golosovsky, I. Mirebeau, G. Andre, *et al.*, Phys. Rev. Lett. **86**, 5783 (2001).
21. Z. Kakol, R. Zalecky, K. Knight, *et al.*, J. Phys.: Condens. Matter **11**, 2749 (1999).
22. K. Knorr and A. Loidl, Phys. Rev. B **31**, 5387 (1985).
23. K. Knorr, A. Loidl, and J. K. Kjems, Phys. Rev. Lett. **55**, 2445 (1985).
24. J.-C. Toledano and P. Toledano, *The Landau Theory of Phase Transitions* (World Sci., Singapore, 1987; Mir, Moscow, 1994), Chap. 4.
25. P. N. Timonin, Phys. Rev. B **69**, 092102 (2004); cond-mat/0308422.
26. S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, cond-mat/0310693.
27. F. Igloi and L. Turban, Phys. Rev. E **66**, 036140 (2002).
28. F. Y. Wu, Rev. Mod. Phys. **54**, 235 (1982).
29. T. Goto, Y. Nemoto, A. Ochiai, and T. Suzuki, Phys. Rev. B **59**, 269 (1999).
30. Y. Qin and Z. R. Yang, Phys. Rev. B **43**, 8576 (1991).
31. C. M. Fortuin and P. W. Kasteleyn, Physica (Amsterdam) **57**, 536 (1972).
32. M. Aizenmann and J. Wehr, Phys. Rev. Lett. **62**, 2503 (1989).
33. J. T. Chayes, L. Chayes, D. S. Fisher, and T. Spencer, Phys. Rev. Lett. **57**, 2999 (1986).
34. L. R. Testardi, Rev. Mod. Phys. **47**, 637 (1975).
35. B. S. Chandrasekhar, H. R. Ott, and B. Seeker, Solid State Commun. **39**, 1265 (1981).

Translated by A. Betev

Journal of Experimental and Theoretical Physics, Vol. 99, No. 5, 2004, pp. 1054–1064.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 5, 2004, pp. 1209–1220.

SOLIDS
Structure

The Role of Defects in the Formation of the Multidomain State of Easy-Plane Antiferromagnets with Magnetoelastic Interaction

V. M. Kalita, A. F. Lozenko, S. M. Ryabchenko, and P. A. Trotsenko

Institute of Physics, National Academy of Sciences of Ukraine, pr. Nauki 144, Kiev, 03028 Ukraine
e-mail: lozenko@iop.kiev.ua

Received June 18, 2004

Abstract—We analyzed the field dependences of forced magnetostriction in the multidomain state of the easy-plane antiferromagnet $CoCl_2$ obtained in the following cycles: the introduction–removal of a magnetic field lying in the easy plane, the introduction–removal of a magnetic field lying in the easy plane and directed normally to that introduced earlier, etc. The magnetostriction of the crystal in the multidomain state was shown to contain two components. First, the component reversible in the cycle magnetic field introduction–removal, which makes the major contribution in the crystal under consideration, and, second, a comparatively small irreversible component, that is, the contribution retained after magnetic field removal. In low fields, the reversible magnetostriction component was proportional to the square of the applied magnetic field. Field-induced rearrangement of the multidomain antiferromagnetic state was found to be responsible for singularities of the field dependence of crystal magnetization. In particular, in a near-zero field that lay in the easy plane, the transverse susceptibility decreased twofold compared with its value in fields in which the crystal is already in the monodomain state. At the same time, close to the "monodomainization" field, transverse magnetic susceptibility was maximum. Defects were shown to favor the formation of the reversible multidomain state. Determining factors in this process were elastic and magnetoelastic interactions. The multidomain state of antiferromagnets was described using the domain distribution function over the orientations of domain antiferromagnetic vectors with respect to the magnetic field direction and the magnetic field dependence of this function. The results of our analysis were in close agreement with the experimental data on $CoCl_2$. *© 2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Magnetic ordering results in the formation of an inhomogeneous multidomain antiferromagnetic state in many antiferromagnets [1]. Antiferromagnetic domains have different antiferromagnetic vector **L** directions. The formation of the multidomain state of antiferromagnets is not related to demagnetizing fields, which are dipole in their nature [2]. Exchange energy loss in domain walls should make this state energetically unfavorable. Therefore, the main problem in describing the multidomain antiferromagnetic state is primarily the explanation of the reasons for its formation [1, 3].

Domains in antiferromagnets were first studied for orientation phase transitions in uniaxial antiferromagnets [4, 5]. In the field region of the transition from the state with the easy-axis antiferromagnetic orientation of spins in sublattices to the state with spins canted with respect to the field and oriented almost normally to the easy axis, both these states are observed simultaneously; that is, antiferromagnets are then in an intermediate state [5]. A description of such a state formed in magnetic fields that magnetize antiferromagnets was performed taking into account magnetostatic energy [6].

In the absence of a magnetic field, antiferromagnets are not magnetized. Various mechanisms have been suggested to explain the formation of the multidomain state as a result of their ordering [1]. However, it appears that there is no unique mechanism applicable to all antiferromagnets. At first sight, the entropy mechanism [7] is most universal. According to this mechanism, exchange energy loss on domain walls is balanced by a decrease in free energy, $-T\Delta S$, because of an increase in the entropy of the multidomain state. Here, T is the temperature and ΔS is the entropy increment. This contribution depends on temperature and vanishes as $T \longrightarrow 0$. In certain instances, the entropy mechanism explains the formation of the multidomain state only close to the Néel temperature [1].

In many antiferromagnets, the equilibrium or almost equilibrium multidomain state exists over the entire temperature interval from T_N to $T \longrightarrow 0$ and is related to defects of various kinds. Domains are formed in crystal twinning, when the domain structure is a combination of antiferromagnetic and structural domains. The formation of the multidomain state is possible when the sequence of sublattices is disarranged, for instance, in the presence of edge dislocations, when the

defect is a half-plane of atoms with magnetic moments related to one of the magnetic sublattices. It was shown in [8, 9] that screw dislocations in antiferromagnets caused the formation of spin dislocations, which, if anisotropy was taken into account, led to the formation of the multidomain antiferromagnetic state. Defects can also stabilize multidomain states of kinetic origin that accompany antiferromagnetic ordering [10].

The simplest example of the influence of defects on the formation of the multidomain state is provided by so-called "metallurgical defects" near which the lattice is distorted in such a way that anisotropy field directions locally change. The local orientations of the antiferromagnetic vector are also determined by inhomogeneities. We will not consider such effects, and our object of study will be fairly perfect crystals with inevitable defects. Attention will be focused on the multidomain states formed as a result of magnetoelastic interaction between the magnetic subsystem and elastic fields of defects. This interaction has been studied comparatively poorly.

The field dependences of magnetostriction in easy-plane antiferromagnets, namely, iron family metal dihalides, were studied experimentally in [11]. The presence of domains in these substances was, in particular, substantiated by neutron diffraction [12]. According to the results obtained in [11], the multidomain state of these antiferromagnets is magnetoelastic in nature. At the same time, the crystal as a whole in the multidomain antiferromagnetic state does not experience spontaneous anisotropic magnetostriction in the easy plane, which accompanies antiferromagnetic ordering, whereas such strain does exist in domains, judging from the presence of a gap in the low-frequency branch of antiferromagnetic resonance. Spontaneous magnetostriction is restored in the crystal as a whole in a magnetic field, as the transition to the monodomain state occurs. When the field is removed, the initial undeformed state of the crystal and, therefore, its multidomain state are restored almost fully. Such a quasi-equilibrium character of the multidomain state of antiferromagnets can, in our view, be stabilized by interactions between spontaneously strained domains and elastic stress fields of defects. For instance, it was shown in [13] that three domains converging at one vertex (a triad of domains) with their vectors **L** rotated through 120° with respect to each other created elastic stresses that balanced lattice stresses caused by defects of a certain type situated at this vertex.

It follows that elastic interactions of domains and defects can provide the energy gain of the multidomain state. On the other hand, the magnetic field dependence of the rearrangement of the multidomain state of antiferromagnets under magnetic field actions should be determined by the specific mechanism of the stabilization of this state. Attention in this work will be focused precisely on this problem.

2. THE REARRANGEMENT OF THE MULTIDOMAIN STATE OF THE $CoCl_2$ ANTIFERROMAGNET UNDER MAGNETIC FIELD ACTION (ACCORDING TO THE MAGNETOSTRICTION DATA)

Two-sublattice antiferromagnets $CoCl_2$ are layered dihalides with D_{3d}^5 symmetry. They consist of sandwiches formed by Co^{2+} ion layers, which are surrounded by Cl^- layers on both sides. The bonds are ion-covalent within the sandwiches and van der Waals between them [14]. The Co^{2+} ions within layers are coupled by comparatively strong ferromagnetic exchange. Exchange interactions of Co^{2+} ions between the nearest sandwiches (that is, between the nearest Co^{2+} layers) are antiferromagnetic and very weak (much weaker than intralayer ferromagnetic exchange). The Néel temperature of $CoCl_2$ is $T_N = 24.7$ K [15].

Uniaxial easy-plane anisotropy in $CoCl_2$ is substantial (the anisotropy field at $T = 4.2$ K is approximately 150 kOe), whereas symmetry-allowed intralayer anisotropy could not be observed. The weakness of intralayer anisotropy results in the degeneracy of vector $\mathbf{L} = \mathbf{s}_1 - \mathbf{s}_2$ directions in the plane, where $\mathbf{s}_1$ and $\mathbf{s}_2$ are the magnetizations of the sublattices. When the crystal is cooled below T_N in the absence of an external magnetic field, it experiences the transition to the multidomain antiferromagnetic state [12]. And when a magnetic field lying in the easy plane is applied, the crystal can be converted to the homogeneous state, in which the **L** vectors of all domains uniformly acquire an orientation normal to the applied field while remaining in the easy plane. The $\mathbf{s}_1$ and $\mathbf{s}_2$ vectors become canted toward the field direction, while remaining in the easy plane. Upon the attainment of the spin flip field (it is $H_{ff} \approx 32$ kOe at $T = 4.2$ K for $CoCl_2$ [15]), the magnetizations of the sublattices become parallel to each other and the field.

Measurements of the forced magnetostriction of $CoCl_2$ single crystals in external magnetic fields were performed using a capacitance dilatometer [16]. An assembly of magnets that created fields perpendicular to each other was used. Both fields were directed in the plane of the crystal. The sample had the shape of a $5 \times 5 \times 1$ mm^3 plate, and the symmetry axis of the crystal was perpendicular to the plate plane. Crystal elongation was measured in the direction lying in the plane along one of its sides. The forced striction value (the relative elongation) is $\varepsilon(H) = \Delta l(H)/l$, where l is the length of the crystal in the direction of measurements and $\Delta l(H)$ is its increment in the field. Crystallographic directions in the plane of the sample plate were not controlled. The samples had fortuitous orientations with respect to these directions.

The $\varepsilon(H)$ dependences of the $CoCl_2$ crystal in the cycle of applying–removing crossed magnetic fields

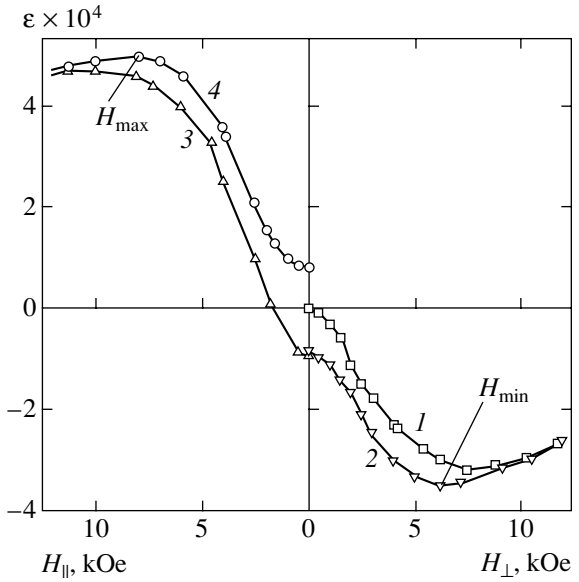


Fig. 1. Dependences of the relative elongation of the $CoCl_2$ crystal on magnetic field strength at $T = 4.2$ K. Curves *1* and *2* were obtained for the introduction–removal of magnetic field $H_\perp$ perpendicular to the direction of elongation measurements, and curves *3* and *4*, when field $H_\parallel$ was parallel to the direction of measurements.

are shown in Fig. 1. Magnetostriction $\varepsilon_\perp(H)$ measurements under field applying and removing conditions when the field H was perpendicular to the direction of measurements correspond to curves *1* and *2*, respectively, and $\varepsilon_\parallel(H)$ values obtained when field H was subsequently applied and removed along the direction of measurements, to curves *3* and *4*, respectively. We see that the field dependences of magnetostriction have the form of a hysteresis loop with residual striction whose sign is determined by the direction of the field being removed. Note that this hysteresis is related to field rotation through 90° rather than the change in its sign. If a field of the same direction is repeatedly applied and removed, with or without changes in its sign, no hysteresis is observed, and the residual striction is only fixed after the first field applying–removing cycle.

The relative elongations caused by the rearrangement of the multidomain state of the $CoCl_2$ crystal (in low fields up to 10 kOe) reach values fairly large for antiferromagnets, on the order of 5×10^{-4}, which actually corresponds to the restoration of spontaneous magnetostriction of the single-domain state under field action. As follows from the dependences shown in Fig. 1 and obtained with switching field directions, this spontaneous magnetostriction in domains in the multidomain state is balanced to within a comparatively small residual striction value, and even this residual deformation is absent until a magnetic field is applied for the first time. At the same time, as follows from the antiferromagnetic resonance data [17], domains in this state retain their spontaneous deformation and remain virtually unstressed, which manifests itself by the presence of a gap in the spectrum of the low-frequency antiferromagnetic resonance branch caused by spontaneous magnetostriction. A multidomain state model should therefore combine the absence of crystal striction as a whole and a mutually-domain configuration when the domains remain free and, therefore, exhibit spontaneous magnetostriction and virtually do not create stresses on one another.

Figure 1 shows that the $CoCl_2$ crystal experiences positive deformation in the region of the transition from the multidomain to homogeneous state; that is, it elongates along the field and shrinks in the direction normal to the field. In fields of both orientations higher than 10 kOe, the crystal is in the uniform (single-domain) state. Close to the transition to this state, $\varepsilon_\parallel(H)$ is maximum and $\varepsilon_\perp(H)$, minimum. The maximum field H_{max} is not equal to (higher than) the minimum field H_{min}. In addition, $\varepsilon_\parallel(H_{max}) > |\varepsilon_\perp(H_{min})|$. This asymmetry of the $\varepsilon_\parallel(H)$ and $\varepsilon_\perp(H)$ dependences will be shown to be related to the special features of the field dependence of the magnetostriction of the homogeneous state.

Our consideration will be restricted to anisotropic magnetoelastic interactions in the easy plane. Let us determine the behavior of magnetostriction in the homogeneous state. Taking into account hexagonal symmetry of the $CoCl_2$ crystal, the sum of the magnetoelastic and elastic energies can be written in the form [13]

$$\begin{aligned} E &= \sum_{\alpha\beta} \gamma_{\alpha\beta}(T)(n_{\alpha x}n_{\beta x} - n_{\alpha y}n_{\beta y})(U_{xx} - U_{yy}) \\ &+ \sum_{\alpha\beta} \lambda_{\alpha\beta}(T)(n_{\alpha x}n_{\beta y} + n_{\beta x}n_{\alpha y})U_{xy} \\ &+ \sum_{\alpha\beta} \delta_{\alpha\beta}(T)(n_{\alpha x}^2 + n_{\alpha y}^2)(U_{xx} + U_{yy}) \\ &+ \frac{1}{2}C_{11}(U_{xx}^2 + U_{yy}^2) + C_{12}U_{xx}U_{yy} + (C_{11} - C_{12})U_{xy}^2, \end{aligned} \tag{1}$$

where $\alpha, \beta = 1, 2$ are sublattice numbers, $\alpha \geq \beta$; γ, λ, and δ are the temperature-dependent parameters of magnetoelastic interactions; $n_{\alpha x}$, $n_{\alpha y}$ and $n_{\beta x}$, $n_{\beta y}$ are the direction cosines of the magnetization vectors of the sublattices $\mathbf{s}_\alpha$ and $\mathbf{s}_\beta$; U_{ij} are the components of the strain tensor; and the x and y axes are oriented in the easy plane. The terms quadratic in U_{ij} describe the elastic contribution to the free energy of the crystal. The first and second sums in (1) are related to anisotropic magnetoelastic interactions determined by the directions of sublattice magnetic moments. The terms with $\alpha = \beta$ and $\alpha \neq \beta$ describe intra- and intersublattice contributions, respectively. The third sum corresponds to magnetoelastic interactions isotropic in the easy plane.

In the homogeneous state, the spins of both sublattices are equally canted with respect to the field in the easy plane, and the cosine of the cant angle is proportional to the ratio between the external field H and the spin flip field H_{ff}. The minimization of (1) with respect to the U_{ij} components taking into account this spin orientation in the field yields equations that describe the applied field dependences of strain values. For the strain U_{yy} when $\mathbf{H}$ is oriented either along y or along x, we have

$$U_{yy}(\mathbf{H} \parallel y, x) = -\frac{2\delta_{11}}{C_{11}+C_{12}} \mp \frac{\gamma_{12}}{C_{11}-C_{12}} + \left(\frac{\delta_{12}}{C_{11}+C_{12}} \pm \frac{2\gamma_{11}}{C_{11}-C_{12}}\right)\left(1-2\frac{H^2}{H_{ff}^2}\right), \quad (2)$$

where the upper sign corresponds to the field orientation $\mathbf{H} \parallel y$ (the field is parallel to the measure strain U_{yy}) and the lower sign, to the $\mathbf{H} \parallel x$ orientation (the field is perpendicular to the measurement direction). The magnetostriction of the $CoCl_2$ crystal in the homogeneous state was analyzed in [18], however, only along the field; for this reason, the asymmetry of the field dependences of magnetostriction mentioned above was not discussed in [18].

Magnetostriction measurements are performed with respect to the size of the crystal at $H = 0$; that is, the isotropic contributions to strain present in (2) are included in this size at the temperature of measurements. As a result, the equation for the magnetostriction of the $CoCl_2$ crystal in the directions along and normal to the field can be written in the form

$$\varepsilon_{d\parallel,\perp} = \pm\varepsilon^{(S)}\left(1-2(1+\eta_{\parallel,\perp})\frac{H_{\parallel,\perp}^2}{H_{ff}^2}\right). \quad (3)$$

Here, $\varepsilon^{(S)}$ is the spontaneous magnetostriction of the single-domain state with $\mathbf{L} \perp \mathbf{H}$ as $\mathbf{H} \longrightarrow 0$ (its values are determined by extrapolating $\varepsilon_{d\parallel,\perp}(H^2)$ from the region of high fields, in which the antiferromagnetic state is homogeneous, to $H \longrightarrow 0$); $H_\parallel$ and $H_\perp$ are the magnetic fields directed parallel and normally to the direction of size measurements; and $\eta_{\parallel,\perp}$ is the parameter that depends on the direction of the field and determines the rate at which magnetostriction increases in the field. According to (3), the magnetostriction of the homogeneous state at $H = 0$ should satisfy the equalities $\varepsilon_{d\parallel}(H_\parallel = 0) = \varepsilon^{(S)}$ and $\varepsilon_{d\perp}(H_\perp = 0) = -\varepsilon^{(S)}$. In addition,

$$\varepsilon^{(S)} = \frac{2\gamma_{11}-\gamma_{12}}{C_{11}-C_{12}}.$$

The $\eta_{\parallel,\perp}$ values are also determined by the parameters of magnetoelastic interactions,

$$\eta_\parallel = \frac{\gamma_{12}}{2\gamma_{11}-\gamma_{12}} + \frac{\delta_{12}(C_{11}-C_{12})}{(2\gamma_{11}-\gamma_{12})(C_{11}+C_{12})},$$

$$\eta_\perp = \frac{\gamma_{12}}{2\gamma_{11}-\gamma_{12}} - \frac{\delta_{12}(C_{11}-C_{12})}{(2\gamma_{11}-\gamma_{12})(C_{11}+C_{12})}.$$

We see that the $\varepsilon_{d\parallel}(H^2)$ and $\varepsilon_{d\perp}(H^2)$ dependences have different slopes at $\delta_{12} \neq 0$ and $\gamma_{12} \neq 0$, because $\eta_\parallel \neq \eta_\perp$.

Our experimental unit allowed us to apply fields $H_\parallel$ up to H_{ff}. In such high fields, the magnetostriction of the single-domain state quadratically depends on H and satisfies (3). These measurements were used in [18] to determine the $\varepsilon^{(S)} = 6.3 \times 10^{-4}$ and $\eta_\parallel = -0.2$ values. The fields introduced normally to the direction of measurements were limited from above by the value $H_\perp$ = 13 kOe. In this field interval, the region of the existence of the homogeneous state was too narrow and insufficient for determining $\eta_\perp$ from the magnetostriction data. It will, however, be shown that the $\eta_\perp$ value can be found by analyzing the $\varepsilon_\perp(H_\perp)$ dependence in the field region of multidomain state rearrangements. The case when $H_{max} > H_{min}$ and $\varepsilon_\parallel(H_{max}) > |\varepsilon_\perp(H_{min})|$ corresponds to the above-mentioned asymmetry of the dependences $\varepsilon_\parallel(H)$ and $\varepsilon_\perp(H)$.

3. DISCUSSION OF THE FIELD DEPENDENCES OF THE MAGNETOSTRICTION OF THE $CoCl_2$ CRYSTAL IN THE MULTIDOMAIN STATE

Let us consider the behavior of magnetostriction when the multidomain state experiences rearrangement in low fields. The dependences of the magnetostriction of the multidomain state on the square of magnetic field intensity shown in Fig. 2 correspond to field withdrawal only, that is, to curve *2* and *4* portions in Fig. 1. We see that, at $H < 2.5$ kOe, the dependence of the relative crystal elongation on magnetic field intensity under field removal conditions is described by the equation

$$\varepsilon_{\parallel,\perp} = \varepsilon_{r\parallel,\perp} + \alpha_{\parallel,\perp}H_{\parallel,\perp}^2, \quad (4)$$

where $\varepsilon_{r\parallel,\perp}$ is the residual magnetostriction in measurements along ($\parallel$) and normally to ($\perp$) the field being removed and $\alpha_\parallel$ and $\alpha_\perp$ are the empirical parameters. The $\varepsilon_{r\parallel}$ and $\varepsilon_{r\perp}$ values are almost equal in magnitude ($\varepsilon_{r\parallel} \approx \varepsilon_{r\perp}$). Let us rewrite (4) in the form

$$\varepsilon_{\parallel,\perp} = \varepsilon_{r\parallel,\perp} \pm \varepsilon^{(S)}\frac{H_{\parallel,\perp}^2}{H_d^2}, \quad (5)$$

where H_d is a parameter with the magnetic field intensity dimensions. Agreement with the results of mea-

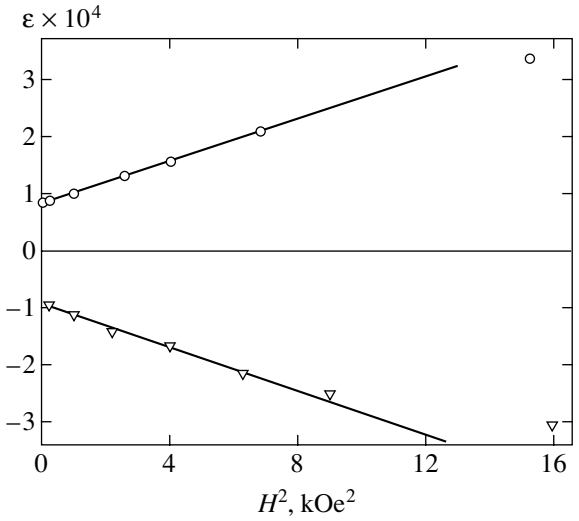


Fig. 2. Dependences of the relative elongation of the $CoCl_2$ crystal in the multidomain state on the square of magnetic field intensity measured when magnetic field was removed from the substance (correspond to curves *2* and *4* in Fig. 1).

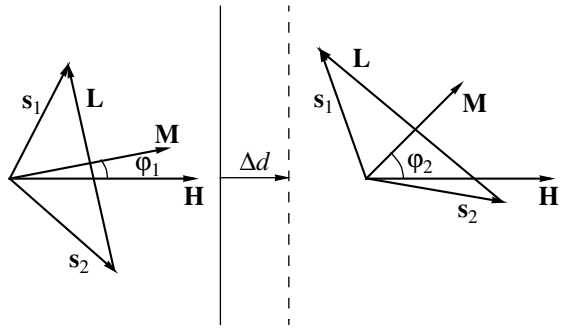


Fig. 3. Orientation of the spin vectors of sublattices $\mathbf{s}_1$ and $\mathbf{s}_2$ and magnetization **M** and domain antiferromagnetic **L** vectors with respect to the magnetic field vector **H**. Neighboring domains with different vector **L** orientations are shown.

surements is attained for $H_d = 5.7 \pm 0.5$ kOe. When field $H < 2.5$ kOe is applied for the first time, the $\varepsilon(H)$ dependence is described by (4), but with $\varepsilon_r = 0$. If the formation of the multidomain state is related to defects, the attainment of the equilibrium state after every change in the magnetic field should occur after the elapse of relaxation time, which, generally, depends on many factors and can be fairly long. It is assumed in (4) that all relaxation processes are completed and $\varepsilon(H)$ is a quasi-equilibrium quantity.

As has been mentioned, the number of domains with the vector **L** orientation $\mathbf{L} \perp \mathbf{H}$ increases and that with $\mathbf{L} \parallel \mathbf{H}$ decreases as a magnetic field is introduced. In the state prior to applying the field for the first time, the relative volumes of domains with arbitrary orientations should be equal. The multidomain state can be described in terms of the domain distribution function $p(\varphi)$ determined from the ratio between the volume of domains whose vector **L** makes angle φ with the field direction and the volume of the whole crystal. Let φ be the angle between the normal to **L** and the direction of **H**. Its changes should be considered in the interval from $-\pi/2$ to $\pi/2$, because **L** is a director rather than a vector. For definiteness, the angle φ can be defined as the angle between $\mathbf{M} = \mathbf{s}_1 + \mathbf{s}_2$ and **H** (see Fig. 3).

It follows from the dependences shown in Fig. 2 that only the anisotropic component of the magnetostriction of domains manifests itself at $H < 3$ kOe at the beginning of multidomain state rearrangement. The field-dependent contributions to the magnetostriction of the single-domain state, which are responsible for its asymmetry, appear in fields $H > 3$ kOe. For this reason, the relative crystal elongation along the direction of the field being introduced can in low fields be written as

$$\varepsilon_{\parallel} = \varepsilon^{(S)} \int_{-\pi/2}^{\pi/2} 2\left(\cos^2\varphi - \frac{1}{2}\right) p(\varphi)\, d\varphi. \qquad (6)$$

It is taken into account in (6) that the spontaneous anisotropic magnetostriction deformation of a separate single domain along and normally to **L** is $\mp\varepsilon^{(S)}$, respectively. The distribution density of domains is normalized according to the condition

$$\int_{-\pi/2}^{\pi/2} p(\varphi)\, d\varphi = 1.$$

In low fields ($H < H_d$) [19], the equation for the distribution density of domains that satisfies (3) under magnetic field removal conditions has the form

$$p(\varphi) = \frac{1}{\pi}\left\{1 + 4\left(\frac{\varepsilon_r}{\varepsilon^{(S)}} + \frac{H^2}{H_d^2}\right)\left(\cos^2\varphi - \frac{1}{2}\right)\right\}, \qquad (7)$$

where ε_r is the modulus of the residual magnetostriction in (4) obtained when the field is removed. Distribution (7) only depends on the angle φ. The ratio between residual and anisotropic magnetostriction in (7) determines the fraction of the domains that retain the field-induced orientation with $\mathbf{L} \perp \mathbf{H}$ after the field is removed. For the first magnetic field introduction, we must set $\varepsilon_r = 0$ in (7) [19].

According to (7), the distribution of domains during magnetic field removal contains two terms one of which is related to the reversible component of the rearrangement of the multidomain state and the other, to the irreversible component. The second term is character-

ized by residual magnetostriction. For this reason, $\varepsilon(H)$ for a crystal can be written as the sum of two terms,

$$\varepsilon(H) = \varepsilon_{rev}(H) + \varepsilon_{irrev}(H), \quad (8)$$

where $\varepsilon_{rev}(H)$ is the reversible contribution to the resultant striction and $\varepsilon_{irrev}(H)$ is the irreversible contribution. Generally, both terms of (8) depend on the field applied and the stage of the cycle of its introduction, although $\varepsilon_{irrev}(H)$, probably starting with some introduced field value, should not depend on field intensity. We assume that the irreversible striction component remains constant as the magnetic field is removed over the entire field interval of the existence of the multidomain state; that is, $\varepsilon_{irrev}(H) = \varepsilon_r = \text{const}$. Accordingly, the field dependences of the $\varepsilon_{rev}(H)$ and $\varepsilon_{irrev}(H)$ contributions to magnetostriction can be determined for the introduction of a field directed differently.

The field dependences $\varepsilon_{rev}(H)$ and $\varepsilon_{irrev}(H)$ for the $CoCl_2$ crystal obtained from the experimental data on the basis of the above considerations are shown in Fig. 4. The component reversible in the cycle of the introduction of crossed fields $\varepsilon_{rev}(H)$ does not have a hysteresis, its values are equal when the field is introduced and removed. Asymmetry of the field dependence becomes even more pronounced for $\varepsilon_{rev}(H)$. The irreversible component $\varepsilon_{irrev}(H)$ has the form of a hysteresis loop. The closing loop portion for the $\varepsilon_{irrev}(H)$ dependence in Fig. 4 (in the first quadrant) was reproduced proceeding from the above mentioned antisymmetric character of magnetostriction in low fields, because Fig. 1 does not contain experimental data in this region.

If the domains are "free" in the multidomain state and their magnetostriction is equal to that in the homogeneous state, an increase in the volume of domains with a favorable orientation in a field accompanied by a decrease in the volume of the other domains can be treated as domain wall displacements. For each magnetic field value, there is a certain balance between the volumes of domains with various **L** directions; that is, when equilibrium is attained, the action of magnetic forces on mobile walls in each field is balanced by forces that counteract changes in the initially equiprobable distribution of domains with different **L** orientations. If defects are responsible for the multidomain state, certain elastic stresses arise every time as the crystal attains equilibrium in a field. These stresses are related to mismatching between the elastic fields of defects and those of magnetostriction strain in domains.

4. THE MAGNETIC FIELD DEPENDENCE OF THE DISTRIBUTION DENSITY OF DOMAINS $p(\varphi)$

Let us consider two different domains that border each other; we will denote the **L** orientation angles in them by φ_1 and φ_2 (Fig. 3). Suppose that $H \neq 0$ and the

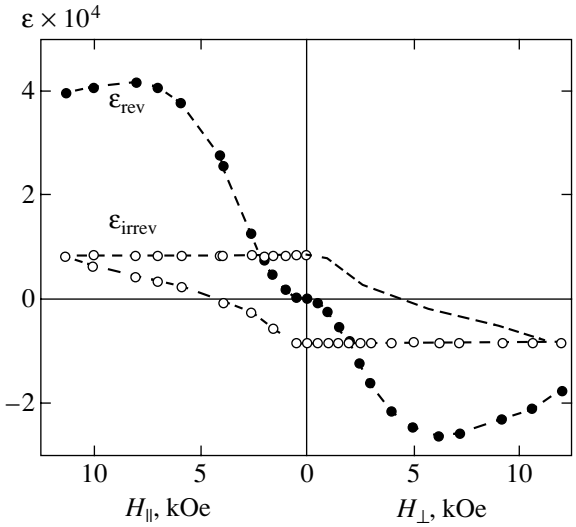


Fig. 4. Field dependences of reversible $\varepsilon_{rev}(H)$ and irreversible $\varepsilon_{irrev}(H)$ components of $CoCl_2$ crystal magnetostriction in crossed magnetic fields in magnetic field introduction–removal cycles.

field vector lies in the easy plane. The spins get canted in both domains under field action. The energy densities acquired by the domains are then

$$e_1 = -\frac{1}{2}\chi_e H^2 \cos^2\varphi_1, \quad e_2 = -\frac{1}{2}\chi_e H^2 \cos^2\varphi_2,$$

where χ_e is the magnetic susceptibility of the homogeneous state of the antiferromagnet for the easy-plane field orientation. The work done when a domain wall is displaced by Δd by the field is calculated as the product of the difference of the energy densities by the change in the volume,

$$\Delta A_{12} = \frac{1}{2}\chi_e H^2(\cos^2\varphi_1 - \cos^2\varphi_2) S\Delta d, \quad (9)$$

where S is the area of the moving wall. The same work value can be determined using the stress σ of the action of the domains on the wall,

$$\Delta A_{12} = \sigma S \Delta d. \quad (10)$$

A comparison of (8) and (9) yields the stress acting on the wall in a magnetic field in the form

$$\sigma = \frac{1}{2}\chi_e H^2(\cos^2\varphi_1 - \cos^2\varphi_2). \quad (11)$$

Upon the attainment of the equilibrium change in $p(\varphi)$ in the given field H, this stress should be balanced by counteraction forces for the reversible component of domain redistribution. Note that (11) does not include the contribution of the change in the magnetoelastic energy of domains in the multidomain state of $CoCl_2$ to

the work of domain wall displacement. The magnetoelastic energy is, however, much smaller than the energy of spins in a field, and this is why we ignore this contribution.

Let us perform averaging over φ_2 in (11) to determine the mean stress of magnetic field action (through differently canting spins) on the walls of the domains with the orientation φ_1. This gives

$$\overline{\sigma_\varphi} = \frac{1}{2}\chi_e H^2(\cos^2\varphi - \overline{\cos^2\varphi}), \quad (12)$$

where bars are used to denote the mean values over the orientations of the domains that surround the distinguished domain with the orientation φ. The number of neighboring domains is finite for each particular domain. However, expression (12) was obtained by calculations on average for the domains with the φ orientation. This is why angle indices are omitted in (12). Indeed, the mean value over all domain orientations should not depend on the index of φ. The mean values denoted by bars should then be calculated over all domain orientations.

The introduction of counteraction forces that ensure equilibrium of the multidomain structure at $H \neq 0$ is essentially thermodynamic. In reality, stress fields counteracting wall displacements can be inhomogeneous. For instance, the surface of the wall bends if the wall encounters a defect during movement. This causes the appearance of surface forces that impede wall displacements, and the larger the bend the greater the pinning force. There also exists a more general approach to describing domain wall displacements, when a wall is considered in a potential field, and the equilibrium wall position corresponds to minimum potential energy. The displacement of a wall from the equilibrium position causes the appearance of forces counteracting this displacement and equal to the potential energy gradient. The parameters of these quasi-elastic forces are determined by the strength of elastic interactions between the wall and defects. When considering wall movement, viscous friction forces are in addition introduced. As we are only interested in equilibrium in a field rather than the rate of its attainment, we disregard these forces.

The irreversible component of domain rearrangement is most probably related to the domains in which the orientation of **L** does not remain unchanged as a result of the action of equilibrating quasi-elastic forces. To distinguish between the reversible and irreversible rearrangement components, let us represent the domain distribution function in the form

$$p(\varphi) = (1-\delta)p_{\text{rev}}(\varphi) + \delta p_{\text{irrev}}(\varphi), \quad (13)$$

where δ is the volume fraction occupied by domains that orient their magnetic moments **M** along the field without causing the appearance of counteraction forces. We will not discuss the nature and form of $p_{\text{irrev}}(\varphi)$. As concerns $p_{\text{rev}}(\varphi)$, we can, taking into account (12) and putting $\varepsilon_r = 0$ in (7), write

$$dp_{\text{rev}}(\varphi) = \frac{8}{\pi\chi_e H_d^2} d\sigma_\varphi. \quad (14)$$

This proportionality has, however, been obtained and holds in low fields only, when $H \ll H_d$ and $p_{\text{rev}}(\varphi) \longrightarrow 1/\pi$. The obvious condition that an increase in the volume of certain domains is only possible at the expense of the volume of the other domains then does not hold. This condition is satisfied in high fields. Indeed, the $p(\varphi)$ density of domains with $\varphi = \pi/2$ is infinitesimal in high fields (that is, $p \longrightarrow 0$), whereas elastic stress applied to the walls of these domains is maximum in the corresponding field. Clearly, the absolute change in the distribution function of these domains dp in high fields, in spite of stress values that act on them, should tend to zero, whereas the relative change is large. Taking this into account, we in addition assume that the increment of the distribution density of domains has the form $dp(\varphi) \sim p(\varphi)d\sigma_\varphi$ rather than is proportional to a mere change in the stress value. As a criterion of the validity of this assumption, let us analyze the experimental data on the magnetic field dependence of magnetization and the field dependence of the reversible magnetostriction component in the cycle of the introduction–withdrawal of mutually orthogonal fields in the plane of the $CoCl_2$ crystal. The differential equation for $p_{\text{rev}}(\varphi)$ has the form

$$\frac{dp_{\text{rev}}(\varphi)}{p_{\text{rev}}(\varphi)} = \frac{8}{\chi_e H_d^2}\delta\sigma_\varphi, \quad (15)$$

which is in agreement with (7) and (14) at $H \ll H_d$. This equation limits an increase in the volume of the domains with the unfavorable orientation of **L** in high fields.

Equation (15) yields the angle φ distribution of domains in the form

$$p_{\text{rev}}(\varphi) = \frac{1}{I(H/H_0)}\exp\left(\frac{H^2}{2H_0^2}\cos^2\varphi\right), \quad (16)$$

where we use the notation

$$H_0 = \frac{H_d}{2\sqrt{2}}, \quad I\left(\frac{H}{H_0}\right) = \int_{-\pi/2}^{\pi/2}\exp\left(\frac{H^2}{2H_0^2}\cos^2\varphi\right)d\varphi.$$

The $p_{\text{rev}}(\varphi)$ dependences for $H_d = 5.7$ kOe and various field values $H = 2$, 5, 10, and 15 kOe are shown in Fig. 5. We see that the distribution becomes unidirectional with the predominant orientation of domains cor-

responding to $\varphi = 0$ ($\mathbf{L} \perp \mathbf{H}$) as H increases. Nevertheless, even at $H = 15$ kOe $\approx 2.5H_d$, the distribution remains fairly broad.

The equation for the distribution density of domains being obtained, let us consider the behavior of the magnetic properties of the multidomain state of antiferromagnets.

5. THE FIELD DEPENDENCES OF MAGNETOSTRICTION, MAGNETIZATION, AND MAGNETIC SUSCEPTIBILITY IN THE MULTIDOMAIN STATE

We will use the equation for the distribution density of domains [Eq. (16)] to analyze the field dependence of the reversible component of the relative crystal elongation under field introduction conditions over the entire interval of the fields of the transition from the multidomain to homogeneous state. The equation for the reversible magnetostriction of the multidomain state can be written in the approximation of the ideal multidomain state ($\varepsilon_{irrev} = 0$) in the form

$$\frac{\varepsilon_{\parallel}(H)}{\varepsilon^{(S)}} = \frac{1}{I(H/H_0)}$$

$$\times \int_{-\pi/2}^{\pi/2} \left[\left(1 - 2(1+\eta_{\parallel}) \frac{H^2 \cos^2\varphi}{H_{ff}^2} \right) \cos^2\varphi \right. \qquad (17)$$

$$\left. - \left(1 - 2(1+\eta_{\perp}) \frac{H^2 \cos^2\varphi}{H_{ff}^2} \right) \sin^2\varphi \right] \exp\left(\frac{H^2}{2H_0^2} \cos^2\varphi \right) d\varphi.$$

An equation similar to (17) describes magnetostriction perpendicular to the field being introduced.

It is easy to show that the expansion of (16) yields (7) in low fields $H/H_0 \ll 1$, and striction is then described by a quadratic field dependence similar to (4). Agreement with experiment is attained for $H_0 = H_d/2\sqrt{2}$.

The experimental $\varepsilon_{\parallel rev}(H)$ and $\varepsilon_{\perp rev}(H)$ values are compared with $\varepsilon_{\parallel}(H)$ calculated by (17) and $\varepsilon_{\perp}(H)$ calculated by the equation similar to (17) in Fig. 6. The $\varepsilon^{(S)}$, H_d, H_{ff}, and $\eta_{\parallel}$ parameters were taken to equal those determined from the experimental data on the reversible component of the magnetostriction of the $CoCl_2$ crystal. To satisfy the condition of asymmetric field dependences of longitudinal and transverse strictions described above, we used $\eta_{\perp} = 0.9$ in model calculations. The field dependences of the magnetostriction of the homogeneous state are given by solid lines in Fig. 6. These dependences were calculated by (3) using the same parameter values. The calculated longitudinal magnetostriction closely agrees with the experimental data over the entire interval of the fields of the transition to the homogeneous state. For transverse magnetostriction in the region of H_{min}, the calculated values are slightly exaggerated compared with the experimental $\varepsilon_{\perp rev}$ data. In reality, this discrepancy does not exceed the error of striction measurements. The conclusion should therefore be drawn that agreement between theory and experiment is quite satisfactory also for the transverse magnetostriction of the multidomain state. Closer agreement can be obtained by measuring transverse magnetostriction in higher fields.

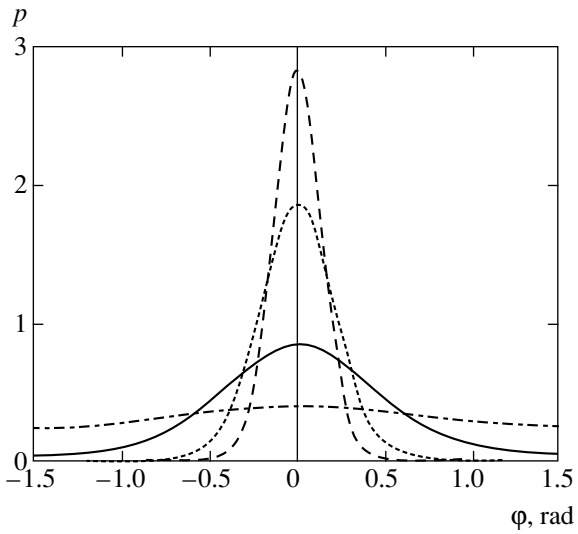


Fig. 5. Field dependences of the distribution density of domains at $H = 2$ kOe (dash-and-dot line), 5 kOe (solid line), 10 kOe (dotted line), and 15 kOe (dashed line).

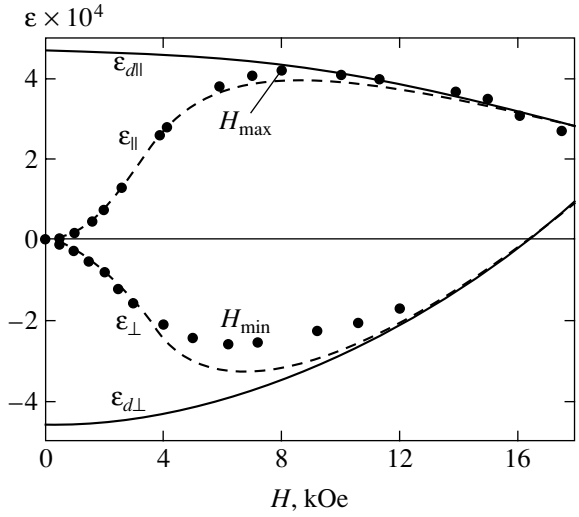


Fig. 6. Field dependences of the magnetostriction of the multidomain (dashed lines) and homogeneous (solid lines) states.

The distribution of domains [Eq. (15)] can be used to calculate the field dependence of the mean magnetization of the crystal in a magnetic field oriented in the

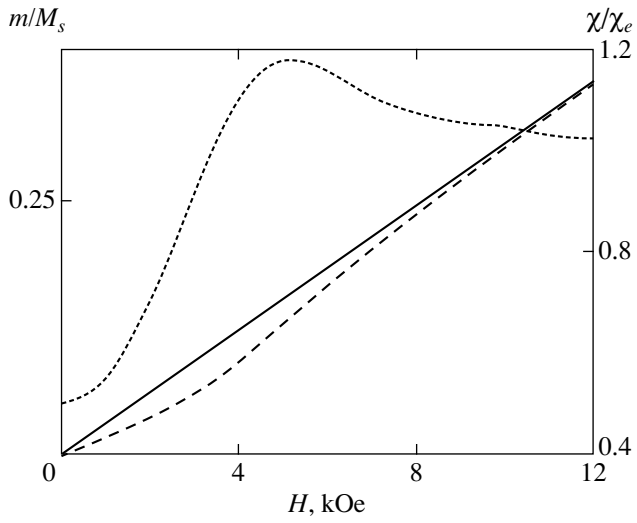


Fig. 7. Field dependences of the normalized magnetization $m(H)/M_S$ and magnetic susceptibility $\chi(H)/\chi_e$ of the multidomain state. The magnetization of the homogeneous state is shown by the solid line, the dashed line corresponds to the multidomain state, and the dotted line, to magnetic susceptibility.

easy plane. The equation for magnetization in the approximation of the ideal multidomain state ($\delta = 0$) can be written in the form

$$\frac{m(H)}{M_S} = \frac{1}{I(H/H_0)}\frac{H}{H_{ff}} \times \int_{-\pi/2}^{\pi/2} \cos^2\varphi \exp\left(\frac{H^2}{2H_0^2}\cos^2\varphi\right) d\varphi, \quad (18)$$

where M_S is the magnetization of the homogeneous state if the crystal is in field $H = H_{ff}$.

The $m(H)/M_S$ dependence calculated by (18) for the same parameter values as with mean striction is shown by the dashed line in Fig. 7. The same figure contains the $m(H)/M_S$ dependence for the homogeneous state of the crystal with $\mathbf{L} \perp \mathbf{H}$ in all fields (the solid thick line). The presence of the multidomain state and its rearrangement in a field result in a nonlinear field dependence of magnetization, which has a characteristic sag with respect to the linear dependence for the homogeneous state [20]. This nonlinearity was experimentally observed not only for the $CoCl_2$ crystal [21] but also for other crystals, in particular, $NiCl_2$ [19, 22]. This behavior of $m(H)$ was not explained in the cited works.

We measured $m(H)$ for $CoCl_2$ samples different from those used to study magnetostriction. The resulting dependences were closely similar to those shown in Fig. 7, but complete coincidence required selecting H_d somewhat different from that given above (H_d = 5.2 kOe). This seems quite natural considering that the multidomain state is related to the degree of structure imperfection, which is different in different crystals.

Note one more special feature of the field dependence of the magnetization of the multidomain state of easy-plane antiferromagnets. In low fields $H \ll H_d$, the $m_{\text{rev}}(H)$ dependence can be written as

$$m(H) = \frac{1}{2}\chi_e H\left(1 + \frac{H^2}{8H_0^2}\right). \quad (19)$$

It follows that the magnetic susceptibility of the reversible magnetization component should be two times lower than that for the homogeneous state at the onset of the rearrangement of the multidomain antiferromagnetic state. This is substantiated experimentally [19, 22]. At $\delta \neq 0$, the magnetic susceptibility changes as $H \longrightarrow 0$ by a factor slightly smaller than two; this change should then depend on the stage of the cycle of the introduction or removal of crossed magnetic fields.

The differentiation of dependence (19) with respect to field yields the field dependence of magnetic susceptibility $\chi(H)$. The field dependence of magnetic susceptibility normalized by the magnetic susceptibility of the homogeneous state $[\chi(H)/\chi_e]$ is shown by the dotted line in Fig. 7. In low fields ($H < H_d$), magnetic susceptibility increases as the square of field. At $H \sim H_d$, it has a maximum and tends to its value for the homogeneous state as field increases further. This maximum should be related to a comparatively sharp change in the distribution density of domains in fields $H \approx H_d$. A maximum was also observed experimentally for $NiCl_2$ [22], but its explanation in [22] was not well grounded.

6. THE INFLUENCE OF SPONTANEOUS ANISOTROPIC MAGNETOSTRICTION ON THE INTERACTION BETWEEN DEFECTS

It can be suggested that the rearrangement of the multidomain state of antiferromagnets is determined by the interaction of elastic fields of defects and elastic fields of the configuration of antiferromagnetic domains that experience magnetostriction. This interaction is responsible for equilibrium of the multidomain state at both $H = 0$ and $H \neq 0$ and for its reversibility. However, is this interaction capable of providing energy gain when the multidomain state is formed?

Clearly, it is exceedingly difficult to exhaustively describe the interaction of defects with each other and with domain elastic fields in a real crystal [23]. This problem will therefore be considered only qualitatively in the simplest case of a planar isotropic medium with defects. The interaction energy e between two neighboring defects (let i and j be their numbers) is a function of the $\mathbf{r}_{ij}$ vector connecting them; that is, $e = e(\mathbf{r}_{ij})$. In the approximation that we use, energy e only depends on the distance between the defects. We assume that the interaction energy of one of the defects under consider-

ation with the other defects is additive; that is, $E_i = \sum_j e(\mathbf{r}_{ij})$.

Let homogeneous spontaneous magnetostriction related to homogeneous antiferromagnetic ordering occur in this plane, and let the frame of reference be fixed at the ith defect. The deformation along the x axis is positive and equals ε_S, and that along the y axis has the same value but opposite sign, $-\varepsilon_S$. Clearly, both axes lie in the plane. After deformation, the equation for the interaction energy between the ith and other defects accurate to second-order terms has the form

$$E_i = \sum_j \left\{ e(r_{0ij}) + \left[\frac{\partial e}{\partial x_j} x_{0j} - \frac{\partial e}{\partial y_j} y_{0j} \right] \varepsilon_S + \frac{1}{2} \left[\frac{\partial^2 e}{\partial x_j^2} x_{0j}^2 - 2 \frac{\partial^2 e}{\partial x_j} x_{0j} y_{0j} + \frac{\partial^2 e}{\partial y_j^2} y_{0j}^2 \right] \varepsilon_S^2 \right\}. \quad (20)$$

Here, the derivatives are calculated at the position of the jth defect, whose coordinates prior to deformation are denoted by x_{0j} and y_{0j}. The distance between the ith and jth defects is r_{0ij}.

The total interaction energy between all defects can be written as the sum of three terms,

$$E = \sum_i E_i = \sum_{ij} e(r_{0ij}) + \sum_{ij} \left\{ \frac{\partial e}{\partial (x_i - x_j)} (x_{0i} - x_{0j}) - \frac{\partial e}{\partial (y_i - y_j)} (y_{0i} - y_{0j}) \right\} \varepsilon_S + \frac{1}{2} \sum_{ij} \left\{ \frac{\partial^2 e}{\partial (x_i - x_j)^2} (x_{0i} - x_{0j})^2 - 2 \frac{\partial^2 e}{\partial (x_i - x_j) \partial (y_i - y_j)} (x_{0i} - x_{0j})(y_{0i} - y_{0j}) + \frac{\partial^2 e}{\partial (y_i - y_j)^2} (y_{0i} - y_{0j})^2 \right\} \varepsilon_S^2. \quad (21)$$

Let us denote the first sum in (21) by E_0. It equals the interaction energy between the defects before the appearance of anisotropic magnetostriction. In the approximation that we use, according to which anisotropic deformation is isomorphic with respect to the local positions of defects, the second sum in (21) should be zero. The coefficient of ε_S^2 in the third sum will be denoted by k. The total interaction energy between the defects after anisotropic magnetostriction can now be written as the sum of two terms,

$$E = E_0 + k\varepsilon_S^2. \quad (22)$$

Growth defects can be considered almost equilibrium, and interaction energy between them should therefore be minimum. For this reason, we must assume that $k > 0$. It follows that homogeneous spontaneous anisotropic magnetostriction favorable for a perfect crystal increases the interaction energy between defects. The dependence of the interaction energy between defects on magnetostriction displacements is therefore responsible for the appearance of a quasi-elastic increasing force.

If defects cannot move over the crystal, then, as follows from (22), they impede the formation of the single-domain state of an antiferromagnet with anisotropic magnetostriction. The crystal then becomes divided into regions with homogeneous but differently directed magnetostrictions, the size of these domains being determined by the distances between defects. A variant of the geometrically correct structure of domains with defects that do not change their mutual arrangement as the multidomain structure is formed was analyzed in [13]. It was also shown in that work that elastic coordination of the striction of domains and defects resulted in an energy gain of the multidomain state additional to (22) for the domains that formed triads.

Note that the magnetoelastic mechanism of multidomain structure formation in antiferromagnets with a contribution to the free energy proportional to the square of the mean magnetostriction similar to the second term in (21) was also considered in [24, 25]. In these works, this term was related to domain elasticity self-action on one another similar to magnetodipole interaction of ferromagnetic domains. In our view, this approach requires additional justification.

7. CONCLUSIONS

We showed that the magnetostriction of the multidomain state of the $CoCl_2$ crystal was determined by the distribution of domains over the orientations of antiferromagnetic vectors in them, and the rearrangement of the multidomain state could be described by magnetic field-induced changes in this distribution.

The equilibrium multidomain state of the magnetoelastic nature can be formed in an antiferromagnet as a result of coordination of stresses that arise in the spontaneous striction of domains and fields of elastic stresses of defects. This interaction results in the appearance of quasi-elastic forces responsible for the equilibrium existence of the multidomain state. When the magnetic field changes, these forces counteract domain wall displacements. We obtained an equation for the field dependence of the equilibrium component of the distribution density of domains and used it to calculate the behavior of the reversible striction, magnetization, and magnetic susceptibility components. A comparison of the results of our calculations with the experimental data shows close agreement between

them over the whole region of fields of the transition to the single-domain state.

We showed that, in addition to the component of the distribution function of domains with respect to the directions of their antiferromagnetic vectors that was reversible in magnetic field introduction–removal cycles, there existed an irreversible component. The determination of the ratio between these two components in crystals with a controlled number of defects of various types is, we believe, an interesting problem for further studies.

REFERENCES

1. M. M. Farztdinov, *Physics of Magnetic Domains in Antiferromagnets and Ferrites* (Nauka, Moscow, 1981) [in Russian].
2. S. V. Vonsovskiĭ, *Magnetism* (Nauka, Moscow, 1971; Wiley, New York, 1974).
3. A. Hubert, *Theorie der Domanenwände in Geordneten Medien* (Springer, Berlin, 1974; Mir, Moscow, 1977).
4. L. Néel, Ann. Phys. (Paris) **3**, 137 (1948).
5. V. G. Bar'yakhtar, A. A. Galkin, and V. A. Popov, Zh. Éksp. Teor. Fiz. **62**, 2233 (1972) [Sov. Phys. JETP **35**, 1169 (1972)].
6. V. G. Bar'yakhtar, A. N. Bogdanov, and D. A. Yablonskiĭ, Usp. Fiz. Nauk **156**, 47 (1988) [Sov. Phys. Usp. **31**, 810 (1988)].
7. Y. Y. Li, Phys. Rev. **101**, 1450 (1956).
8. I. E. Dzyaloshinskiĭ, Pis'ma Zh. Éksp. Teor. Fiz. **25**, 110 (1977) [JETP Lett. **25**, 98 (1977)].
9. A. S. Kovalev and A. M. Kosevich, Fiz. Nizk. Temp. **3**, 259 (1977) [Sov. J. Low Temp. Phys. **3**, 125 (1977)].
10. E. M. Lifshitz, Zh. Éksp. Teor. Fiz. **42**, 1354 (1962) [Sov. Phys. JETP **15**, 939 (1962)].
11. V. M. Kalita, A. F. Lozenko, S. M. Ryabchenko, and P. A. Trotsenko, Ukr. Fiz. Zh. **43**, 1469 (1998).
12. M. K. Wilkinson, J. W. Cable, E. O. Wollan, and W. C. Koehler, Phys. Rev. **113**, 497 (1959).
13. V. M. Kalita and A. F. Lozenko, Fiz. Nizk. Temp. **27**, 489 (2001) [Low Temp. Phys. **27**, 358 (2001)].
14. M. E. Lines, Phys. Rev. **131**, 546 (1963).
15. J. W. Leech and A. J. Manuel, Proc. Phys. Soc. London, Sect. B **59**, 210 (1956).
16. Z. A. Kazeĭ, M. V. Levanidov, and V. I. Sokolov, Prib. Tekh. Éksp. **2**, 196 (1981).
17. A. F. Lozenko, P. E. Parkhomchuk, S. M. Ryabchenko, and P. A. Trotsenko, Fiz. Nizk. Temp. **14**, 941 (1988) [Sov. J. Low Temp. Phys. **14**, 517 (1988)].
18. V. M. Kalita, A. F. Lozenko, and S. M. Ryabchenko, Fiz. Nizk. Temp. **26**, 671 (2000) [Low Temp. Phys. **26**, 489 (2000)].
19. V. M. Kalita, A. F. Lozenko, S. M. Ryabchenko, *et al.*, Fiz. Tverd. Tela (St. Petersburg) **46**, 317 (2004) [Phys. Solid State **46**, 326 (2004)].
20. L. Néel, Izv. Akad. Nauk SSSR, Ser. Fiz. **21**, 890 (1957).
21. C. Starr, F. Bitter, and A. R. Kaufmann, Phys. Rev. **58**, 977 (1940).
22. V. M. Kalita, A. F. Lozenko, P. A. Trotsenko, and T. M. Yatkevich, Fiz. Nizk. Temp. **30**, 38 (2004) [Low Temp. Phys. **30**, 27 (2004)].
23. A. M. Kosevich, *Theory of Crystal Lattice* (Vishcha Shkola, Kharkov, 1988) [in Russian].
24. E. V. Gomonaĭ and V. M. Loktev, Fiz. Nizk. Temp. **25**, 699 (1999) [Low Temp. Phys. **25**, 520 (1999)].
25. H. Gomonay and V. Loktev, J. Phys. C **14**, 3959 (2002).

Translated by V. Sipachev

Journal of Experimental and Theoretical Physics, Vol. 99, No. 5, 2004, pp. 1065–1073.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 5, 2004, pp. 1221–1231.

SOLIDS
Electronic Properties

Magnetic Moment of Square SIS Josephson Arrays: Self-Organized Criticality

É. V. Matizen[a,*], S. M. Ishikaev[a], and V. A. Oboznov[b]

[a] *Nikolaev Institute of Inorganic Chemistry, Siberian Division, Russian Academy of Sciences, Novosibirsk, 630090 Russia*

[b] *Institute of Solid-State Physics, Russian Academy of Sciences, Chernogolovka, Moscow oblast, 142432 Russia*

**e-mail: matizen@che.nsk.su*

Received December 9, 2003

Abstract—The temperature- and magnetic-field dependences of the magnetic moment of square Josephson arrays with SIS-type junctions are studied experimentally. Two temperature regions are observed with different types of magnetization curves. Magnetic flux avalanches are detected in the low-temperature region. Statistical analysis of avalanche amplitudes A shows that their size distribution varies in accordance with the power law $P \propto A^n$ with crossover, when exponent n varies from $n = -0.7$ for small avalanches to $n = -6$ for large avalanches, while the frequency spectrum varies in accordance with the law $1/f^{\alpha}$. Such behavior is interpreted as a manifestation of self-organized criticality. *© 2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

In this paper, we report on the results of experimental investigation into magnetic properties of arrays with a Josephson junction of the superconductor–insulator–superconductor type (SIS arrays). Our aim was to observe peculiarities of magnetic field penetration in Josephson arrays (J arrays) by measuring their magnetic moment.

In the Introduction, the studies pertaining to our work are briefly reviewed. The characteristics of arrays and the description of the main features of a SQUID magnetometer specially designed by us are given in Section 2. The results of measurements of the magnetic moment of arrays and observation of self-organized criticality realized in SIS arrays are presented in Section 3. Section 4 contains our conclusions.

The interest in properties of regular two-dimensional J arrays has been growing in recent years. A large number of transport studies were undertaken in this field, but there are practically no experimental results on direct measurement of the magnetic moment. However, magnetic studies are undoubtedly of considerable importance. Such experiments may lead to the discovery of new phenomena since the complexity of the electrodynamics of J arrays does not allow all the factors to be considered in theoretical models. We are aware of only one publication [1] devoted to direct measurements of magnetic properties (susceptibility) of J arrays, which appeared after our first communication [2]. At the same time, a large number of theoretical works (see, for example, [3–7]) deal with the magnetic properties of J arrays and require verification of the matching of the theoretical and experimental results.

Experimental studies of magnetic properties of J arrays are also important from the practical point of view. Such arrays are currently considered as a prospective source of radio waves in the millimeter and submillimeter spectral ranges, which have not been explored extensively so far [8].

The problem concerning the existence of self-organized criticality in J arrays is of special interest. According to Bak *et al.* [9, 10], who developed this theory, self-organized criticality is a fairly wide-spread phenomenon in nature. It is observed in many complex systems formed by a large number of elements actively interacting with one another. Such discrete systems have a huge number of metastable states with an interesting dynamics. Upon monotonic variation of external conditions in a system that has attained instability, jumplike transitions occur between metastable states; such transitions have the form of avalanches with different sizes, which sustain the critical state of the system on average. It is important to note that these avalanches are independent of the intensity of external action or fluctuations; even an insignificant effect may lead to a huge avalanche (catastrophe). Another feature is that, in spite of random motion, the system is self-organized and acquires a certain constant (on average) parameter, e.g., the sandpile slope (sandpile problem). Thus the system itself maintains the critical state, which is the essence of self-organized criticality. In this case, there is no need to adjust any parameters in contrast, say, to the case of a liquid, when two of its parameters (temperature and density) have to be kept constant for investigating the critical point.

The size distribution of avalanches (probability density) is a power function with a negative nonintegral

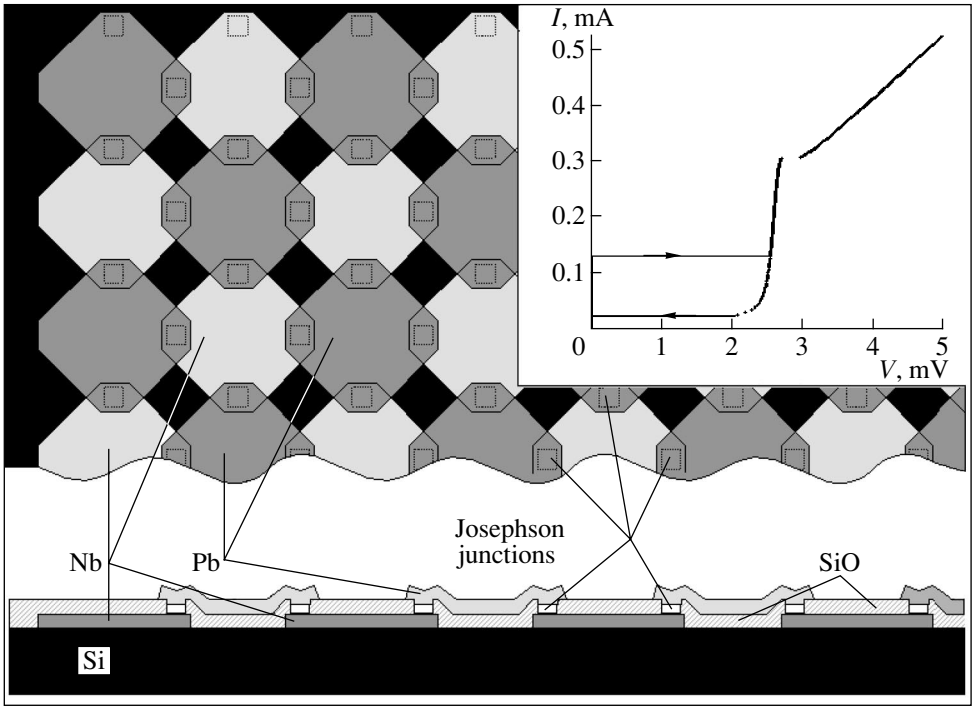


Fig. 1. J array with octagonal electrodes (SIS1 and SIS2). The inset shows the current-voltage characteristic of a single junction in the SIS2 array at 4.2 K.

exponent. This dependence is observed, for example, by the magnitudes of earthquakes (the Gutenberg–Richter law [11]). Fifteen years following pioneering studies of self-organized criticality [9, 10], a large number of theoretical models were constructed imitating various natural phenomena such as earthquakes [12, 13], crossing phase transitions [14], quark–hadron phase transitions [15], rain phenomena [16], spreading of forest fires [17, 18], economic crises [19], and evolution of populations in biology [20].

Experimental data on self-organized criticality have been obtained for a rather limited class of artificially prepared objects: in the study of the dynamics of sandpile growth [21], motion of a piece of sandpaper over a nylon carpet [22], film boiling of nitrogen on the surface of a high-temperature superconductor (HTSC) in the vicinity of the superconducting transition [23], and plastic flow of a loaded metal rod [24]. Ginzburg [25] was the first to indicate the possibility of self-organized criticality in a Josephson medium in 1994. Magnetic flux avalanches were probably observed in HTSC ceramics in experimental works [26, 27], but magnetic moment jumps were not studied in detail and the mechanism of their formation was not discussed in fact.

In a number of recent publications, the power dependence of the avalanche amplitude distribution in the case of self-organized criticality has fallen under doubt. A revision [28] and experiments [29] show that the distribution function is rather of an exponential type $P(x) \propto \exp(-(x/x_0)^\mu)$, where μ is a nonintegral exponent. In this case, we have a characteristic scale x_0 and the avalanche size distribution function is not uniform. This contradicts the "classical" self-organized criticality since this theory presumes gauge invariance while observing avalanches of all sizes. However, all real systems have a finite size and, hence, the maximal size of avalanches is also limited to a certain value. Thus, gauge invariance has natural limits of applicability.

2. SAMPLES AND EXPERIMENTAL TECHNIQUE

We studied square J arrays consisting of 100 × 100 cells with Nb–NbO_x–Pb Josephson tunnel junctions in the temperature range 2–10 K in magnetic fields up to 200 Oe. Two configurations of arrays with different shapes of Nb and Pb superconducting film islands (electrodes) were designed for our experiments. One electrode configuration was in the form of an octagon and the other was cross-shaped. Fragments of the arrays are shown in Figs. 1 and 2. The cross-shaped configuration was distinguished by a high inductance of a cell and had an area of a tunnel junction four times as large as that in the octagonal configuration; this allowed us to obtain high critical currents.

J arrays with octagonal electrodes containing 100 × 100 cells were made according to the same pattern in two copies with different critical currents (these arrays will be henceforth referred to as SIS1 and SIS2 arrays), while the arrays with cross-shaped electrodes (SISk

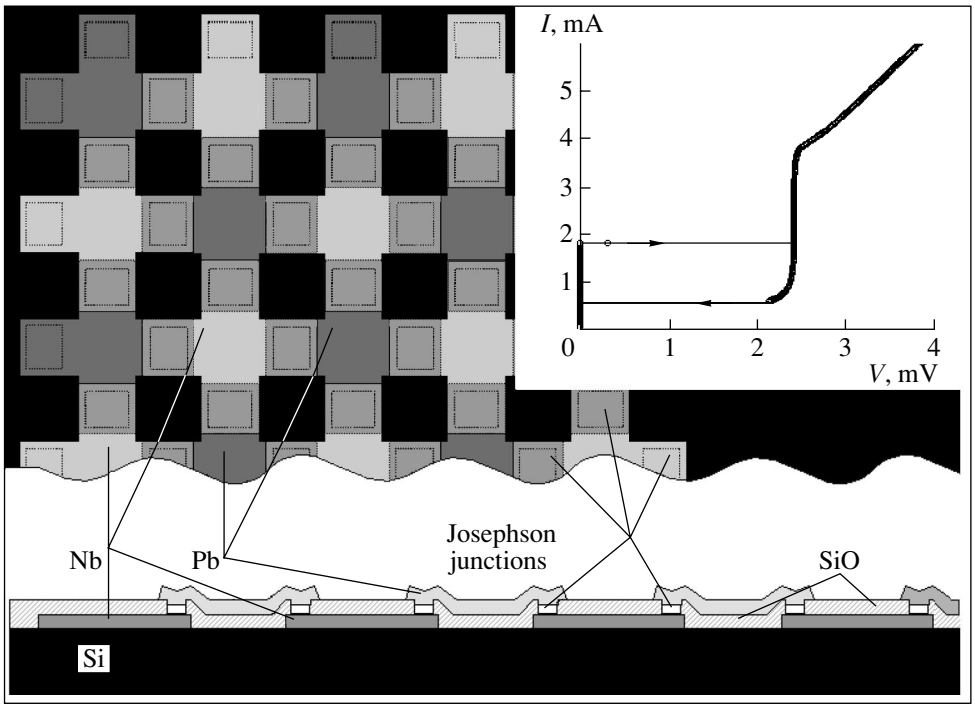


Fig. 2. J array with cross-shaped electrodes (SISk). The inset shows the current-voltage characteristic of a single junction at 4.2 K.

arrays containing 64×64 cells) had only one value of the critical current. The arrays were prepared using the thin-film technology described, for example, in [30]. The lower Nb layer was obtained using magnetron sputtering followed by photolithography and chemical etching. The insulating layer of silicon monoxide and the upper PB layer were obtained by vacuum evaporation; the structure of these layers was formed with the help of explosion photolithography. After the formation of windows (for the tunnel contacts) in the SiO layer, followed by ion cleaning of the surface, a NbO_x tunnel interlayer was formed by controlled oxidation of the niobium surface in an argon–oxygen mixture.

It should be noted that SIS arrays have a short service life (their parameters remained unchanged only for about two months). Comparative parameters of SIS1, SIS2, and SISk arrays are given in the table.

We studied magnetic properties of SIS arrays with the help of an original SQUID magnetometer developed at the Low-Temperature Physics Laboratory at the Institute of Inorganic Chemistry, Siberian Division, Russian Academy of Sciences. The magnetometer had a number of considerable advantages in the design of detecting coils of the flux transformer, in the form of compensation of their astaticism, and in the construction of the solenoid. The detecting coils of the flux transformer were prepared in the form of a symmetric second-order gradiometer [31, 32]. However, in contrast to the classical scheme, the central coil was divided into two equal parts separated by a certain dis-

Table

Junction type	SIS1	SIS2	SIS3
Junction structure	Nb–NbO_x–Pb	Nb–NbO_x–Pb	Nb–NbO_x–Pb
Array size	100×100	100×100	from 64×64
Cell size, μm^2	20×20	20×20	20×20
Junction area, μm^2	≈7	≈7	≈25
Critical current at 4.2 K, μA	≈80	≈150	≈1800
Normal resistance, Ω	10	20	≈0.7
Inductance of a cell, H	$\approx 2.5 \times 10^{-12}$	$\approx 2.5 \times 10^{-12}$	$\approx 10^{-11}$
Capacitance of junction, nF	0.01	0.01	0.03

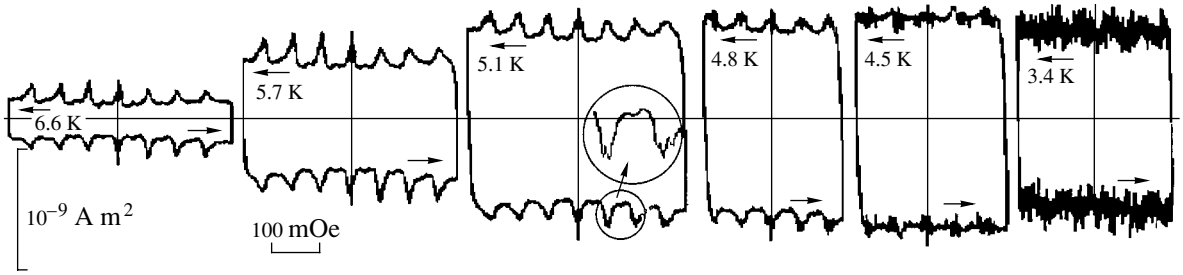


Fig. 3. Magnetization curves for the SIS2 array containing 100×100 cells at various temperatures.

tance so that these coils had the form of Helmholtz coils. This ensured a number of advantages (in particular, reduced microphone noise and a substantially weaker dependence of the signal on the position of the sample), which enabled us to study magnetic properties in the direct magnetization mode and to obtain reproducible results.

The solenoid consisted of two superconducting parts: an external short-circuited part and an internal open part. In the short-circuited solenoid, a certain value of the field was frozen, while the open solenoid served for continuous field variation in certain limits. Astaticism of thoroughly prepared detecting coils was approximately 3×10^{-4}; a small coil consisting of several turns of copper wire wound on the same frame as in the flux transformer and coupled inductively with it was introduced for additional compensation. This coil was connected in series with the open solenoid. The number of turns in this coil (six in our case) was selected so as to compensate astaticism of the system of detecting coils to the maximal possible extent. During operation, a current was passed through the additional coil, which was not equal to the current in the solenoid, but was proportional to it with a certain coefficient that could be varied in certain limits. Thus, the total slope of magnetization curves could be varied by adding a value proportional to the field to the sample signal, which almost completely compensated the contribution from screening currents of intrinsic superconducting Nb and Pb film electrodes. As a result, the magnetic moment being measured for the most part contained only the contribution from the currents flowing in the Josephson array. It should be noted that, without such compensation, the weakly manifested structure of the signal could not be singled out in the course of subsequent processing against the background of a large total slope of the magnetization curve.

To reduce drifts and noise, liquid helium containing the flux transformer, the solenoid, and the superconducting magnetic screen was transformed to the superfluid state by evacuating vapor. For the same purpose, measurements were mostly performed at night. The temperature was measured using a Cu + 0.1%Fe–Cu + 0.1%Ge thermocouple with a sensitivity of about 10 μV/K at helium temperatures; the reference points in this case were the superconducting transitions in niobium and lead as well as the point of transition of helium to the superfluid state, which was easily fixed from a sharp decrease in low-frequency noise in the recording system.

3. MAGNETIC PROPERTIES OF SIS ARRAYS

Figure 3 shows the magnetic moment hysteresis loops for SIS2 array at different temperatures. Analogous curves were also obtained for the other array, SIS1 [2]. It can be seen from Fig. 3 that two temperature regions with different types of magnetization curves were obtained.

Our experiments show that in the first (high-temperature) region from 5.3 K up to the superconducting transition temperature of lead (T_c = 7.2 K), the hysteresis loops for SIS2 array are continuous reproducible curves. All the curves in this region exhibit a clearly manifested structure in the form of magnetic moment peaks of the same height with a temperature-independent field period of about 59 mOe. If we take into account the suppression of the mean field in the region of cells of the array as a result of the screening action of superconducting film electrodes, the period of the peaks matches the value $\Delta H = \Phi_0/a^2 \approx 52$ mOe, where Φ_0 is the magnetic flux quantum and a is the period of the array.

The hysteresis loops reflect a specific critical state formed in a regular Josephson structure. Periodic peaks of the magnetic moment correspond to an increase in the critical current of fluxon pinning (depinning current) for integral frustrations, when the same integral number of flux quanta (fluxons) corresponds to each cell and, hence, the fluxon distribution in the array is the most regular and stablest. The stability of the vortex system is manifested in the increase in the depinning current and, accordingly, in the total magnetic moment.

In addition to large-amplitude peaks, small humps located exactly between the peaks (a sort of second harmonic) can be clearly seen. These humps correspond to a change in the flux in the array on the average by a fluxon for each two cells. This is apparently a consequence of a rather stable distribution of flux quanta in J arrays, which are added in staggered order [33, 34].

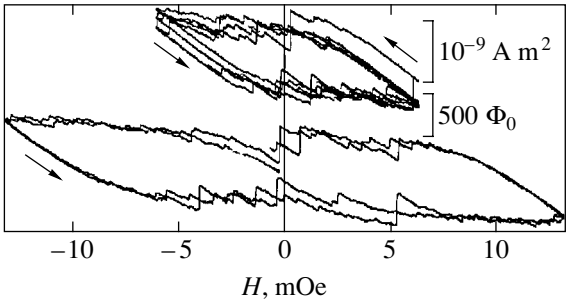


Fig. 4. Hysteresis loops for the SIS1 array containing 100 × 100 cells at T = 2.15 K in fields up to ±15 mOe.

In all probability, higher order harmonics associated with periodic formation of fluxon superlattices with a still larger period also exist in J arrays; however, such harmonics cannot be distinguished in the experiment against the noise background.

In the second (low-temperature) region from 5.3 to 2 K, the hysteresis loops for SIS1 array acquire magnetic moment jumps resembling noise; in contrast to temperature fluctuations, the amplitude of these jumps increases with decreasing temperature. First, such jumps appear at the tops of peaks and then propagate to the sides, forming compact periodic groups with the same period as on high-temperature curves. With a further decrease in temperature, the groups of jumps merge together.

For illustration, Fig. 4 shows the magnetization curves for SIS1 array upon a change in the external field within ±15 mOe at 2.15 K. The upper curve contains four complete consecutive cycles while the lower curve contains two cycles. All curves have segments on which the monotonic variation of the magnetic moment is interrupted by sharp spontaneous drops followed again by the monotonic dependence until the next jump. It can be clearly seen that these jumps occur at random values of the field and their amplitudes exhibit a substantial spread. Note the presence of monotonic and quite reproducible segments of 5–6 mOe on which a transition to another branch of the loop after the field reversal takes place.

Characteristic sharp shifts, which occur over a very short time in the same direction only at the tops of the peaks (owing to which these shifts can be reliably singled out from noise) can be reliably traced from direct observation of the recorder pen, which recorded the signals simultaneously. In this way, we could fix the emergence of jumps for the other array (SIS2), now below 5.7 K.

The study of arrays of the other configuration (SISk) revealed that the general form of hysteresis loops does not differ in principle from those for SIS1 and SIS2 arrays. Figure 5 shows the hysteresis loops for a SISk array containing 64 × 64 cells at different temperatures.

The formation of flux avalanches in this array begins at relatively high temperatures (lower than the superconducting transition temperature for lead by less than one degree) in view of a substantially higher critical current of junctions and a larger inductance of the cell of this array. At the same time, the temperature below which the type of the magnetization curve changes (crossover) and magnetic moment jumps appear in arrays SIS1 and SIS2 is approximately equal to 6 K.

For interpreting crossover, it is interesting to consider the temperature dependence of depinning current I_{dep} in SIS1 array (Fig. 6). The fluxon depinning current was estimated from the half-width of magnetization hysteresis loops on the basis of the simplified assumption that currents flow in the array along concentric square paths; the width of the loop is obviously proportional to the depinning current. In the case of square current lines, simple calculation leads to the following expression for the magnetic moment of the array:

$$M = I_{dep} N^3 a^2/6,$$

where I_{dep} is the depinning current, which is assumed to be constant over the entire array and N is the number of cells in the array. We can expect that real current lines differ from squares; these lines are as if rounded at the corners and the actual areas of current contours are smaller than the predicted values. Consequently, the estimate based on the simplest model is slightly lowered. To obtain the estimate from above, we can consider current contours in the form of circles (in this

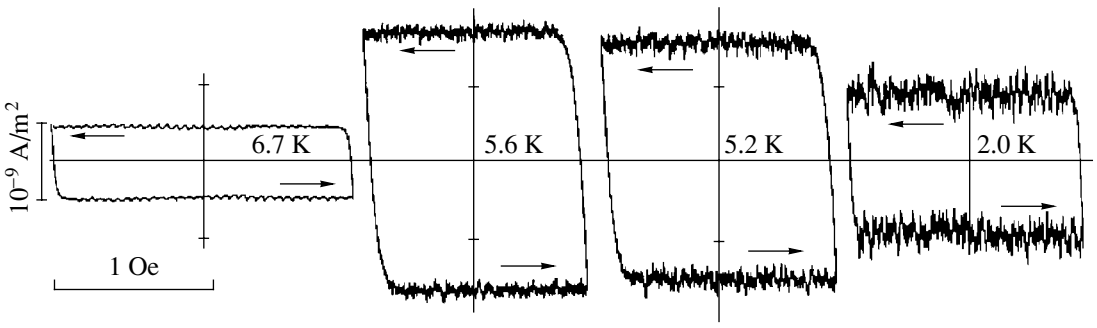


Fig. 5. Hysteresis loops for the SISk array containing 64 × 64 cells at various temperatures.

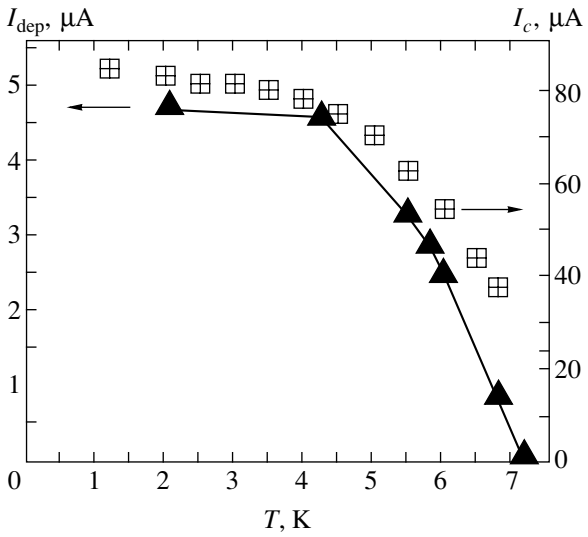


Fig. 6. Temperature dependence of the depinning current in the SIS1 array. Triangles correspond to estimates of current from the magnetic moment of the array, while squares denote the data obtained from direct measurements of the critical current in a single junction.

case, we assume that the current does not flow in the corners of the array). This estimate gives a value that is higher approximately by 20%. In Fig. 6, the mean value of the depinning current I_{dep} of the array under investigation at various temperatures is compared with the directly measured critical current I_c of a single Josephson junction. It can be seen that the two temperature dependences are in good agreement for an appropriate ratio of the scales. In accordance with the theory [34], depinning currents and the critical current must differ approximately by an order of magnitude: $I_{dep}/I_c = 1/10$. Our ratio is approximately equal to 1/15, which can be explained both by the error in determining the depinning current from the magnetic moment on the basis of the simple model and by a slight decrease in the critical current as a result of degradation of junctions during several months.

To obtain a qualitative description of electrodynamic properties of J arrays considered here, we will follow the theoretical publications [3, 4, 25, 35] in which the existence of two different regimes of magnetic flux flow in a J array or two regions differing in their magnetic properties. The boundary between the regions is determined, analogously to [3, 4], by parameter $k = \lambda/a$, where λ is the magnetic field penetration depth in the array. It should be noted that the reciprocal of parameter k corresponds to the Ginzburg parameter $V = 1/k$ [25, 35]. Parameter k is a function of temperature since penetration depth λ varies with temperature as [36, 37]

$$\lambda(T) = \frac{\Phi_0}{2\pi\mu_0 j_c(T)}, \qquad (1)$$

where $j_c(T)$ is the critical current density in Josephson junctions, which increases with decreasing temperature. Thus, by decreasing temperature, it is possible to pass from one regime to the other in experiments.

Substituting the experimentally measured temperature dependence $I_c(T)$ into formula (1) (Fig. 6) shows that $\lambda = a$ for $T \approx 6$ K. This value corresponds to the experimentally measured crossover temperature (5.8 K) below which the shape of the magnetization curves changes and magnetic moment jumps appear.

At high temperatures for which $LI_c \ll \Phi_0$, where L is the inductance of a cell, an individual cell cannot confine a flux quantum and each fluxon is distributed over several cells. This corresponds to the condition $k \gg 1$ (weak pinning); in this case, fluxons penetrate the J array in the form of hypervortices extending over a large number of cells. The interaction between fluxons at weak pinning leads to their deep penetration into the array with an almost uniform distribution. The field profile on the array in this case is formed by peaks at the centers of hypervortices, which are distributed almost regularly over the area of the array (see, for example, 4, 38]. For $k \gg 1$, when a fluxon is spread over a large number of cells, the dynamics of Josephson vortices can be described in the continuous limit, when states with minimal energy exist in the system. Such a theoretical model is confirmed experimentally since the curves obtained in Fig. 6 match the curves calculated for large values of the Josephson penetration depth even in detail (see Fig. 14 in [3]).

In the opposite case, when the critical current is quite large, the condition $LI_c \gg \Phi_0$ is satisfied, each cell can confine a magnetic flux much stronger than one quantum, and each cell may contain only an integral number of fluxons. The fluxon dynamics in this regime can be described as the motion of discrete quasiparticles, which are localized within a cell and possessing a certain effective mass. This corresponds to the case $k \ll 1$ (strong pinning state). As the external magnetic field increases in contours of cells (with initial zero flux), the screening current also increases together with the magnetic moment of a cell. When the current attains its critical value, a fluxon enters the cell; its magnetic moment decreases jumpwise, the magnetic field penetrates into the J array discretely and almost synchronously over nearly quadratic contours. In this case, the system of fluxons is in metastable states far from equilibrium (absolute minimum) and the profile of the increasing field forms a rectangular well with steps from contour to contour; i.e., the profile resembles the Bean distribution of the field in a type II bulk superconductor.

It was proposed in [25] that a J array for $k \ll 1$ may acquire self-organized criticality, because for large value of critical currents through the Josephson junctions, the ensemble of fluxons in it can be treated as a discrete interactive system. For large values of the J array, the entire array as a whole possesses a large num-

ber of metastable states, which is a necessary condition for observation of self-organized criticality. In a slowly varying magnetic field, when screening currents approach the critical value, the system of fluxons attains an unstable state from which is passes to one of numerous metastable states under the action of spontaneous perturbations. Such a form of existence near the critical state is typical of self-organized criticality. The final state occupied by the system as a result of each such jump is naturally determined by the configuration of metastable states of the system as well as the dynamics of the collective motion of fluxons in an array and a large number of random perturbing factors (e.g., thermal fluctuations).

Figure 7 shows a fragment of the magnetization curve for a SISk array, which contains segments with a monotonic variation of the magnetic moment interrupted by sharp drops. Dependences of this type are characteristic of self-organized criticality studies on entirely different objects [21, 22, 24, 29]. For SIS2 and SISk arrays, we managed to record sufficiently large bodies of data for obtaining reliable statistical characteristics of avalanches.

We believe that the magnetic moment jumps observed in our experiments (see Figs. 4 and 7) confirm the existence of self-organized criticality in a J array. The histogram of the number of jumps as a function of their amplitudes demonstrates a power dependence $P \propto A^n$, which is a "calling card" of self-organized criticality. The exponent for the SIS1 array was found to be $n = -1.9$ [2]; however, it should be noted that the data were comparatively scarce; for this reason, the spread of points was quite large in the region of rare events (large amplitudes). It is interesting to note that computer simulation was used in [35] to obtain close values of $n = -1.75$ and $N = -1.80$ for structures consisting of 256 and 128 cells for a 1D Josephson array (one-dimensional multiple-contact SQUID). However, self-organized criticality appeared in [35] only in calculations with discrete introduction of external currents to random junctions of the system and did not appear when the external field varied continuously as in this study.

The curve describing the amplitude distribution of jumps for the SIS2 array has a different form: two clearly manifested regions of power dependence of the probability density of the emergence of avalanches on their amplitude are observed on the curve with exponent $n = -0.7$ for small avalanches and about $n = -6$ for large avalanches with a sharp crossover (Fig. 8). Some authors (starting from [11]) noted a similar crossover in the avalanche distribution for other objects or even used an exponential dependence for describing such a distribution [29].

We did not observe crossover [2] on the SIS1 array, probably due to the fact that the number of avalanches was too small for a detailed statistics. Crossover is observed in the range of large avalanches, whose probability is low; for this reason, for a relatively small

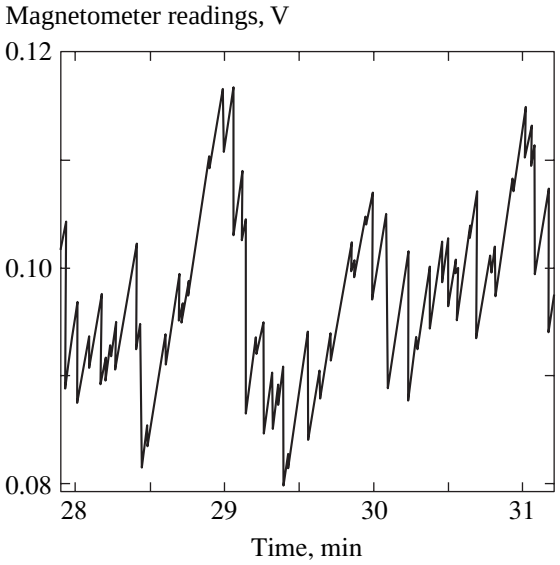


Fig. 7. A fragment of the magnetization curve of the SISk array at 4.1 K; magnetic moment jumps corresponding to magnetic flux avalanches are clearly seen.

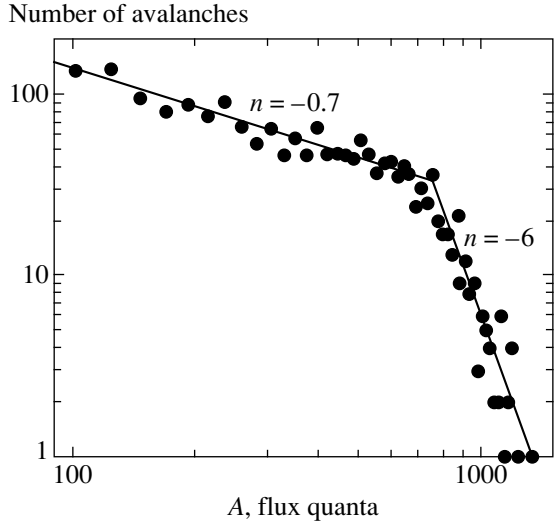


Fig. 8. Distribution of magnetic moment jumps (magnetic flux avalanches) with respect to amplitude in SIS2 grid at $T = 4.1$ K.

number of avalanches (slightly larger than 1000), the spread of results in this region was quite large and crossover could not be fixed reliably [2].

Figure 9 shows the Fourier spectrum of avalanches into which the magnetization curves were decomposed in sweeping the magnetic field with a constant rate with jumps (avalanches) appearing on the curves. It can be seen from Fig. 9 that the spectrum for high frequencies behaves as $1/f^{\alpha}$ in the region of at least 1.5 decades.

Our studies remain unique where the behavior of the magnetic moment is studied during continuous magnetization of regular J arrays and self-organized criticality

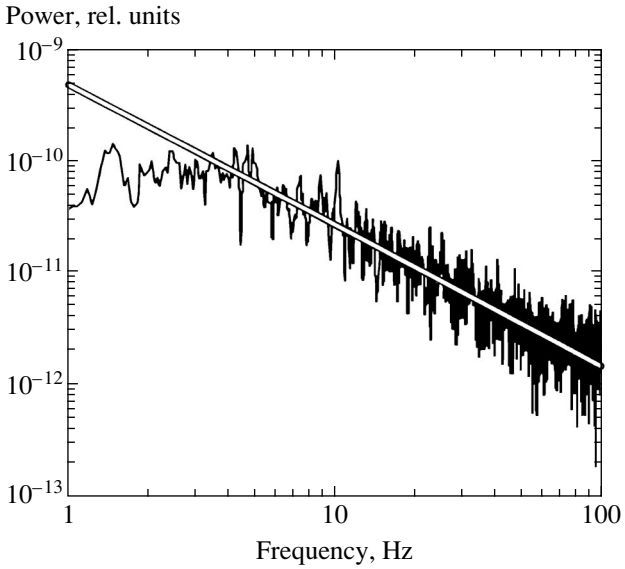


Fig. 9. Fourier spectrum for fragments on the magnetization curve, on which avalanches appear.

is observed. It should be noted that this phenomenon was also observed recently in polycrystalline niobium films [39] as well as in a granulated Nb film [29] almost simultaneously with our first publication [2]. According to Altshuler *et al.* [39], they observed a power dependence of the distribution upon a change in amplitudes within two orders of magnitude, although the curve plotted on the log–log scale is convex rather than a straight line. Behnia *et al.* [29] assert that they observed an exponential dependence of the amplitude distribution function and therefore doubt that self-organized criticality exists in this structure. Indeed, many computational models (see, for example, [40]) lead to an analogous distribution, which appears on the log–log scale as a convex curve with "heaping" at large amplitudes.

After our first experiments, Ginzburg and Savitskaya [40] confirmed in their calculations that avalanches in arrays can also be observed during magnetization of a J array and not only when current pulses are supplied to individual random junctions, as was proposed in [35]. In their opinion, the reason for self-organized criticality during magnetization is not the spread in critical currents I_c in different junctions of the array under investigation, but a weak spatial aperiodicity of the array of about 5%; in our case, this corresponds to a random deviation of 1 μm and lies within the error of our technology.

4. CONCLUSIONS

Our experiments aimed at studying magnetic properties of regular square Josephson SIS arrays revealed a number of new effects. Above all, these effects include magnetic flux avalanches obeying the regularities of self-organized criticality.

As regards the recent dispute over the avalanche distribution law, it should be noted that the size distribution, which is governed by a power law according to the results of our studies, apparently exhibits crossover, which is probably due to a finite size of the array; for this reason, the probability of observing large avalanches is low. In [29] (Fig. 3), crossover can also be clearly seen in the avalanche distribution, but the authors of [29] describe it by an "expanded" exponential.

Our spectrum of magnetic flux avalanches also has a power form within almost two frequency decades with exponent $n = -1.3$.

Josephson arrays have a number of advantages as a model object for experimental study of self-organized criticality. In this model, internal and external conditions can easily be varied to study their effect on the characteristics of self-organized criticality. Our experimental model makes it possible to calculate the magnetic dynamics of a J array on the basis of well-developed concepts and to verify the correctness of the results.

It should be noted in conclusion that we did not observe self-organized criticality in SNS arrays and in HTSC ceramics with intergranular contacts close in their properties to SNS. The absence of avalanches can be explained by dissipation of energy required for evolution of avalanches in the conducting interlayer of junctions, which does not take place in SIS junctions. The analysis of reasons for the emergence of self-organized criticality requires additional theoretical and experimental studies of regular J arrays of various types (SIS and SNS) and configurations with various areas of junctions and critical current densities.

ACKNOWLEDGMENTS

We are grateful to V.V. Ryazanov, S.L. Ginzburg, and N.E. Savitskaya for fruitful discussion of experimental results.

This study was financed by the Russian Foundation for Basic Research (project no. 02-02-16564) and state contract no. 40.012.1.1.1356.

REFERENCES

1. W. Maluf, Jr., G. M. Cecato, P. Barbara, *et al.*, J. Magn. Magn. Mater. **226–230**, 290 (2001).
2. S. M. Ishikaev, É. V. Matizen, V. V. Ryazanov, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **72**, 39 (2000) [JETP Lett. **72**, 26 (2000)].
3. D. Dominguez and J. V. Jose, Phys. Rev. B **53**, 11692 (1996).
4. D.-X. Chen, A. Sanches, and A. Hernando, Phys. Rev. B **50**, 10342 (1994); D.-X. Chen, J. J. Morreno, and A. Hernando, Phys. Rev. B **53**, 6579 (1996).
5. D. Reinel, W. Dieterich, A. Majhofer, and T. Wolf, Physica C (Amsterdam) **245**, 193 (1995).

6. V. V. Bryskin, A. V. Gol'tsev, and S. N. Dorogovtsev, Pis'ma Zh. Éksp. Teor. Fiz. **51**, 53 (1990) [JETP Lett. **51**, 63 (1990)].
7. H. S. J. van der Zant, F. C. Fritshy, T. R. Orlando, and J. E. Mooij, Phys. Rev. B **47**, 295 (1993).
8. P. Barbara, A. B. Cawthorne, S. V. Shitov, and C. J. Lobb, Phys. Rev. Lett. **82**, 1963 (1999).
9. P. Bak, C. Tang, and K. Wisenfeld, Phys. Rev. Lett. **59**, 381 (1987).
10. P. Bak, K. Chen, and M. Creutz, Nature **342**, 780 (1989).
11. B. Gutenberg and C. F. Richter, Ann. Geophys. **9**, 1 (1956).
12. Z. Olami, H. J. S. Feder, and K. Christensen, Phys. Rev. Lett. **68**, 1244 (1992).
13. J. M. Carlson and J. S. Langer, Phys. Rev. Lett. **62**, 2632 (1989).
14. V. P. Koverda and V. P. Skripov, in *Metastable States and Phase Transitions* (Inst. Teplofiz. Ural. Otd. Ross. Akad. Nauk, Yekaterinburg, 2000), p. 45 [in Russian].
15. R. C. Hwa and J. Pan, Nucl. Phys. A **590**, 601c (1995).
16. R. F. S. Andrade, S. T. R. Pinho, S. C. Fraga, and A. P. M. Tanajura, Physica A (Amsterdam) **314**, 405 (2002).
17. B. Drossel, Phys. Rev. Lett. **76**, 936 (1996).
18. B. Drossel and F. Schwabl, Phys. Rev. Lett. **69**, 1629 (1992).
19. G. Iori and S. Jafarey, Physica A (Amsterdam) **299**, 205 (2001).
20. P. Bak and K. Sneppen, Phys. Rev. Lett. **71**, 4083 (1993).
21. G. A. Held, D. H. Solina, D. T. Keane, *et al.*, Phys. Rev. Lett. **65**, 1120 (1990).
22. H. J. S. Feder and J. Feder, Phys. Rev. Lett. **66**, 2669 (1991).
23. V. P. Koverda, V. N. Skokov, and V. P. Skripov, Pis'ma Zh. Éksp. Teor. Fiz. **63**, 739 (1996) [JETP Lett. **63**, 775 (1996)]; V. N. Skokov, A. V. Reshetnikov, V. P. Koverda, and A. V. Koverda, Physica A (Amsterdam) **372**, 1 (2001).
24. M. A. Lebyodkin, Y. Brechet, Y. Estrin, and L. P. Kubin, Phys. Rev. Lett. **74**, 4758 (1995); M. A. Lebedkin and L. R. Dunin-Barkovskiĭ, Zh. Éksp. Teor. Fiz. **113**, 1816 (1998) [JETP **86**, 993 (1998)].
25. S. L. Ginzburg, Zh. Éksp. Teor. Fiz. **106**, 607 (1994) [JETP **79**, 334 (1994)].
26. C. E. Gough, Int. J. Mod. Phys. B **1**, 891 (1987).
27. C. E. Gough, Physica C (Amsterdam) **153**–**155**, 1569 (1988).
28. E. Laherrere and D. Sornette, Eur. Phys. J. B **2**, 525 (1998).
29. K. Behnia, C. Capan, D. Mailly, and B. Etienne, Phys. Rev. B **61**, R3815 (2000).
30. V. A. Oboznov and A. V. Ustinov, Phys. Lett. A **139**, 481 (1989).
31. S. M. Ishikaev and E. V. Matizen, in *Proceedings of Joint Symposium of the SB RAS and the CNEAS TU on High-Temperature Superconductivity: New Materials and Properties* (Tohoku Univ., Japan, 1999), p. 65.
32. S. M. Ishikaev, Prib. Tekh. Éksp., No. 3, 145 (2002) [Instrum. Exp. Tech. **45**, 426 (2002)].
33. J. R. Philips, H. S. J. van der Zant, J. White, and T. P. Orlando, Phys. Rev. B **47**, 5219 (1993).
34. T. E. Trias, J. R. Philips, H. S. J. van der Zant, and T. P. Orlando, IEEE Trans. Appl. Supercond. **5**, 2707 (1995).
35. S. L. Ginzburg and N. E. Savitskaya, Pis'ma Zh. Éksp. Teor. Fiz. **68**, 688 (1998) [JETP Lett. **68**, 719 (1998)].
36. S. J. Lobb, D. W. Abraham, and M. Tinkham, Phys. Rev. B **27**, 150 (1983).
37. A. Barone and G. Paterno, *Physics and Applications of the Josephson Effect* (Wiley, New York, 1982; Mir, Moscow, 1984).
38. V. Marconi and D. Dominguez, Phys. Rev. B **63**, 174509 (2001).
39. E. Altshuler, T. H. Johansen, Y. Paltiel, *et al.*, cond-mat/0208266.
40. S. L. Ginzburg and N. E. Savitskaya, Pis'ma Zh. Éksp. Teor. Fiz. **73**, 163 (2001) [JETP Lett. **73**, 145 (2001)].

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 99, No. 5, 2004, pp. 1074–1089.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 5, 2004, pp. 1232–1248.

SOLIDS
Electronic Properties

Resonance Tunneling in Superconducting Junctions with Different Order Parameter Symmetries

D. V. Goncharov, I. A. Devyatov*, and M. Yu. Kupriyanov**
Institute of Nuclear Physics, Moscow State University, Vorob'evy gory, Moscow, 119992 Russia
**e-mail: idev@pn.sinp.msu.ru*
***e-mail: mkupr@pn.sinp.msu.ru*
Received March 24, 2004

Abstract—Resonance tunneling in superconducting junctions with electrode order parameters of *s* and *d* symmetry was studied. The Green function formalism was used to derive an equation for the resonance current in junctions of arbitrary dimensionality and order parameter symmetry. A universal equation for resonance supercurrent was obtained for junctions of arbitrary dimensionality with isotropic electrode order parameters. A numerical analysis of resonance current transport in junctions of various types was performed for the two-dimensional model. © *2004 MAIK "Nauka/Interperiodica"*.

1. INTRODUCTION

In recent years, much interest has been shown in Josephson junctions based on high-T_c superconducting materials. Experimental works in this area revealed the existence of a large number of localized states in high-T_c superconductors [1]. It was shown [2] that the transfer of the normal current component in such structures is resonance in character and proceeds through localized states. For this reason, theories that only take into account tunneling of quasi-particles through the "weak coupling" region [3–7] are insufficient for calculating the transport properties of the junctions.

Theoretical studies of resonance tunneling in NIN structures were performed using the tunnel model [8] and the three-dimensional model of junctions [9, 10]. Resonance current transport in junctions in which one of or both electrodes were superconductors with order parameters of *s* symmetry was also considered [11–18]. Resonance current transport in superconducting junctions with *d* pairing in the electrodes was theoretically studied using the one-dimensional model in [19]. It was shown in [20] that, at low voltages, resonance tunneling in NID structures suppressed junction conduction singularities arising because of the presence of bound zero energy states in superconducting electrodes [21]. At the same time, a consistent theory of the resonance transport of supercurrent in DID structures has not been developed thus far.

In addition, an analysis of the experimental data [2] showed that the transport of the normal current component was resonance in character in some high-T_c superconducting junctions, whereas the transport of supercurrent was determined by direct tunneling without the participation of localized states. No consistent explanation of this effect has been suggested thus far.

The purpose of this work was to develop a theory of resonance transport of supercurrent in high-T_c superconducting junctions. Our preliminary results were reported in [22].

2. A JUNCTION MODEL

We assume that the tunnel barrier $V(\mathbf{r})$ in the junction under consideration is the sum of two potentials (see Fig. 1)

$$V(\mathbf{r}) = V_{\text{rect}} + V_{\text{imp}}, \tag{1}$$

where the first term models a rectangular barrier of height V and thickness d,

$$V_{\text{rect}}(x) = V\theta(x(d-x)), \tag{2}$$

and the second term describes a localized state in the

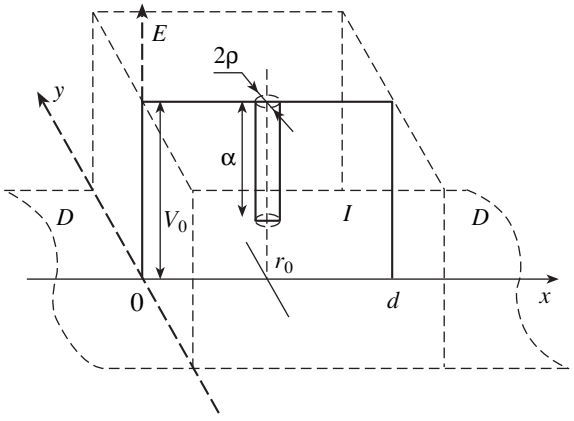


Fig. 1. Junction model.

1063-7761/04/9905-1074$26.00 © 2004 MAIK "Nauka/Interperiodica"

interlayer material,

$$V_{\text{imp}}(\mathbf{r}-\mathbf{r}_0) = \begin{cases} -\alpha, & |\mathbf{r}-\mathbf{r}_0| \le \rho, \\ 0, & |\mathbf{r}-\mathbf{r}_0| > \rho, \end{cases} \quad (3)$$

which is situated at some point $\mathbf{r}_0 = (x_0; y_0)$ and has radius $\rho \ll |\mathbf{k}|^{-1}$ ($\mathbf{k}$ is the Fermi momentum of quasi-particles in the electrodes). We restrict our consideration to the limit of weak coupling and assume vector $\mathbf{k}$ to be fixed on the Fermi surface ($|\mathbf{k}| = k \approx k_F$). Potential (3) disturbs spatial homogeneity of the structure and results in the nonconservation during tunneling of the component of the momentum of quasi-particles parallel to the barrier. The barrier thickness is taken to be fairly large,

$$\lambda_0 d \gg 1, \quad (4)$$

where $\lambda_0 = \sqrt{2mV_0}$ is the momentum of quasi-particles in the interlayer ($V_0 = V - \mu$), m is the mass of the electron, and μ is the Fermi energy. In addition, we assume that the conditions

$$\lambda_0(d - x_0) \gg 1, \quad \lambda_0 x_0 \gg 1 \quad (5)$$

are met for the localized states under consideration. Meeting (5) is necessary for the effective localization of the wave function of a quasi-particle on a defect. Junction transport properties will also be calculated on the assumption that the current that passes through the junction does not drive the superconducting electrodes from the state of thermodynamic equilibrium. This condition is automatically satisfied in structures with *s*- and *d*-type superconducting electrodes separated by a broad potential barrier of low transparency, even when the barrier contains localized states spaced at distances that far exceed their effective transverse radius $l_\perp = d/\sqrt{\lambda_0 d}$. It is assumed that the density of localized states in the interlayer is low and their mutual influence is insignificant.

3. THE GREEN FUNCTION OF THE PROBLEM

The Green function of the problem is found from the Gor'kov equations [23] modified for the case of an anisotropic contact [24],

$$\mathbf{G}_\omega(\mathbf{r}, \mathbf{r}') = \begin{pmatrix} G_{11}(\mathbf{r}, \mathbf{r}') & G_{12}(\mathbf{r}, \mathbf{r}') \\ G_{21}(\mathbf{r}, \mathbf{r}') & G_{22}(\mathbf{r}, \mathbf{r}') \end{pmatrix},$$

$$\left\{ -\frac{1}{2}\int d\mathbf{r}_1 (\Delta(\mathbf{r}, \mathbf{r}_1)\sigma_+ + \Delta^*(\mathbf{r}_1, \mathbf{r})\sigma_-) \right. \quad (6)$$

$$\left. + \delta(\mathbf{r}-\mathbf{r}_1)(i\omega\hat{I} - \sigma_z\hat{h}) \right\} \mathbf{G}_\omega(\mathbf{r}_1, \mathbf{r}') = \delta(\mathbf{r}-\mathbf{r}')\hat{I},$$

where the operator $\hat{h}$ is given by

$$\hat{h} = -\nabla^2/2m + V(\mathbf{r}) - \mu.$$

In (6), $\Delta(x, x_1)$ is the order parameter of the system, which depends on the coordinates; T is the temperature; and $\omega = (2n+1)\pi T$ denotes the Matsubara frequencies. In addition, (6) contains the second-order unit matrix $\hat{I}$ and the Pauli matrices σ_x, σ_y, and σ_z ($\sigma_\pm = \sigma_x \pm i\sigma_y$). Equation (6) can conveniently be rewritten in the integral form

$$\mathbf{G}_\omega(\mathbf{r}, \mathbf{r}') = \mathbf{G}^0_\omega(\mathbf{r}, \mathbf{r}') + \int d\mathbf{r}_1 \mathbf{G}^0_\omega(\mathbf{r}, \mathbf{r}_1)\sigma_z V_{\text{imp}}(\mathbf{r}_1 - \mathbf{r}_0)\mathbf{G}_\omega(\mathbf{r}_1, \mathbf{r}'). \quad (7)$$

The Green function $\mathbf{G}^0_\omega$ of the unperturbed problem is found from (6) after replacing the operator $\hat{h}$ with the operator of the spatially homogeneous problem without localized states. The $\mathbf{G}^0_\omega$ Green function is then obtained as the sum of plane waves with coefficients that follow from the condition of continuity of the functions themselves and their derivatives at the boundaries of the structure. (Details of calculating $\mathbf{G}^0_\omega$ are given in the Appendix.)

Localized states have the atomic size scale. The Green functions $G_\omega(r, r')$ change along interatomic distances; that is, they are slowly varying functions of the coordinate on the scale of $V_{\text{imp}}(\mathbf{r})$ function changes. This circumstance allows them to be removed from the integrand in (7),

$$\mathbf{G}_\omega(\mathbf{r}, \mathbf{r}') \approx \mathbf{G}^0_\omega(\mathbf{r}, \mathbf{r}') + \{\int d\mathbf{r}_1 \mathbf{G}^0_\omega(\mathbf{r}, \mathbf{r}_1)V_{\text{imp}}(\mathbf{r}_1 - \mathbf{r}_0)\}\sigma_z \mathbf{G}_\omega(\mathbf{r}_0, \mathbf{r}'), \quad (8)$$

to arrive at a matrix equation for determining the local $G_\omega(r_0, r')$ function of the form

$$\mathbf{G}_\omega(\mathbf{r}_0, \mathbf{r}') = \mathbf{G}^0_\omega(\mathbf{r}_0, \mathbf{r}') + V(\mathbf{r}_0)\sigma_z \mathbf{G}_\omega(\mathbf{r}_0, \mathbf{r}'), \quad (9)$$

where

$$\mathbf{V}(\mathbf{r}_0) = \begin{pmatrix} V_{11}(\mathbf{r}_0) & V_{12}(\mathbf{r}_0) \\ V_{21}(\mathbf{r}_0) & V_{22}(\mathbf{r}_0) \end{pmatrix} = \int d\mathbf{r}_1 \mathbf{G}^0_\omega(\mathbf{r}_0, \mathbf{r}_1)V_{\text{imp}}(\mathbf{r}_1 - \mathbf{r}_0). \quad (10)$$

Equation (9) yields the Green function value at the point where the localized state is situated,

$$\mathbf{G}_\omega(\mathbf{r}_0, \mathbf{r}') = (1 + \det[\mathbf{V}(\mathbf{r}_0)]\sigma_z \mathbf{V}^{-1}(\mathbf{r}_0))\frac{\mathbf{G}^0_\omega(\mathbf{r}_0, \mathbf{r}')}{\Phi}. \quad (11)$$

Here, the denominator Φ is determined through the components of matrix (10) as

$$\Phi = (1 + V_{22}(\mathbf{r}_0))(1 - V_{11}(\mathbf{r}_0)) + V_{12}(\mathbf{r}_0)V_{21}(\mathbf{r}_0).$$

Like $\mathbf{G}^0_\omega(\mathbf{r}, \mathbf{r}_1)$, the $\mathbf{G}_\omega(\mathbf{r}, \mathbf{r}')$ function in (8) is a slowly varying function of $\mathbf{r}$ on the scale of $V_{\text{imp}}(r)$ variations and can also be removed from the integrand. We can therefore write the solution to the integral equation (8) in terms of the unperturbed Green function of problem (11) and the localized state potential,

$$\begin{aligned}\mathbf{G}_\omega(\mathbf{r}, \mathbf{r}') &= \mathbf{G}^0_\omega(\mathbf{r}, \mathbf{r}') \\ &+ \{\int d\mathbf{r}_1 V_{\text{imp}}(\mathbf{r}_1 - \mathbf{r}_0)\}\mathbf{G}^0_\omega(\mathbf{r}, \mathbf{r}_0)\sigma_z\mathbf{G}_\omega(\mathbf{r}_0, \mathbf{r}').\end{aligned} \tag{12}$$

4. THE TRANSPORT PROPERTIES OF THE JUNCTION

The transport properties of the junction can be calculated using the general equation for the current,

$$I(\varphi) = \frac{eT}{2im}\lim_{x' \to x}\left(\frac{\partial}{\partial x'} - \frac{\partial}{\partial x}\right)\sum_{\omega_n}\text{Tr}\{\mathbf{G}_\omega(\mathbf{r}, \mathbf{r}')\}\Big|_{x=0}, \tag{13}$$

where $\varphi = \varphi_L - \varphi_R$ is the difference of the macroscopic phases of the order parameters of the left and right superconductors and e is the charge of the electron.

In the two-dimensional junction model, the trace of the $\mathbf{G}_\omega(\mathbf{r}, \mathbf{r}')$ matrix in (13) is determined from (12); it can be represented in the form

$$\begin{aligned}\text{Tr}\{\mathbf{G}_\omega(\mathbf{r}, \mathbf{r}')\} &= G^0_{11}(\mathbf{r}, \mathbf{r}') + G^0_{22}(\mathbf{r}, \mathbf{r}') \\ &+ \Lambda(\mathbf{r}_0)\sum_{j,l=1}^{2}(-1)^{l+1}G^0_{jl}(\mathbf{r}, \mathbf{r}_0)\Phi G_{jl}(\mathbf{r}_0, \mathbf{r}'),\end{aligned} \tag{14}$$

$$\Lambda(\mathbf{r}_0) = \frac{1}{\Phi}\int d\mathbf{r}_1 V_{\text{imp}}(\mathbf{r}_1 - \mathbf{r}_0).$$

Here, the parameter $\Lambda(\mathbf{r}_0)$ is the modulus of the amplitude of resonance scattering for electronic and hole excitations. Equation (14) can be partitioned into two components responsible for different transport channels of quasi-particles through the junction. The first component, $G^0_{11}(\mathbf{r}, \mathbf{r}') + G^0_{22}(\mathbf{r}, \mathbf{r}')$, describes potential scattering of quasi-particles by barrier (2), and the second component, resonance scattering by localized state (3).

The tunnel current of Josephson junctions of various configurations, including junctions with high-T_c superconducting electrodes, has well been studied theoretically [3–5]. The d symmetry of the order parameter of the high-T_c superconductor is responsible for several nontrivial effects, which are not observed for junctions with isotropic superconductors. These are the appearance of bound electron–hole zero energy states at the insulator–d-type superconductor boundary (as a result, zero bias anomalies appear in the current–voltage characteristics at low voltages [21]) and several effects related to the suppression of the order parameter at the insulator–superconductor boundary [6, 7].

In our problem, the potential component of the current through the junction is determined by the well-known equation

$$\begin{aligned}I_{\text{pot}}(\varphi) &= \frac{eT}{2im}\lim_{x' \to x}\left(\frac{\partial}{\partial x'} - \frac{\partial}{\partial x}\right) \\ &\times \sum_{\omega_n}(G^0_{11}(\mathbf{r}, \mathbf{r}') + G^0_{22}(\mathbf{r}, \mathbf{r}'))\Big|_{x=0}.\end{aligned} \tag{15}$$

If the order parameters of the left and right superconductors are denoted by the superscripts "L" and "R" and both have d symmetry, then we can, ignoring order parameter suppression close to the junction boundaries, write

$$\Delta_\pm(x, \theta) = \begin{cases}\Delta^{\text{L}}(T)\cos(2\theta \mp 2\alpha_L)e^{i\varphi_L}, & x < 0, \\ 0, \quad 0 < x < d, \\ \Delta^{\text{R}}(T)\cos(2\theta \mp 2\alpha_R)e^{i\varphi_R}, & x > d.\end{cases} \tag{16}$$

Here, $\Delta_\pm(x, \theta)$ is the anisotropic order parameter, which depends on the angle θ of the propagation of quasi-particles with respect to the x axis. Equation (15) allows us to obtain the following equation for the potential current:

$$I_{\text{pot}}(\varphi) = \frac{eTk_F}{i\pi}\sum_{\omega_n}\int_{-\pi/2}^{\pi/2} d\theta\cos\theta \cdot |t|^2\frac{(\Gamma_3 - \Gamma_4)}{Z}. \tag{17}$$

The notation in (17) is as follows:

$$\begin{aligned}Z &= -\Gamma_1\Gamma_2 - |t|^2(\Gamma_3\Gamma_4 - \Gamma_1\Gamma_2), \\ \Gamma_1 &= 1 + \beta^{\text{L}}_-\beta^{\text{L}*}_+, \quad \Gamma_2 = 1 + \beta^{\text{R}}_+\beta^{\text{R}*}_-, \\ \Gamma_3 &= 1 + \beta^{\text{L}}_-\beta^{\text{R}*}_-, \quad \Gamma_4 = 1 + \beta^{\text{R}}_+\beta^{\text{L}*}_+, \\ \beta^{\text{L, R}}_\pm &= \frac{-i(\Delta^{\text{L, R}}_\pm)^*}{\omega + \Omega^{\text{L, R}}_\pm}, \\ \Omega^{\text{L, R}}_\pm &= \sqrt{\omega^2 + |\Delta^{\text{L, R}}_\pm|^2},\end{aligned} \tag{18}$$

and $|t|^2$ is the tunnel transparency of the barrier. Equation (17) reduces to the result obtained in [3], and our numerical calculations for the SID and DID junction

models repeat the phase dependences of the current and the temperature dependences of the critical current obtained in that work.

The equation for the supercurrent of direct tunneling [Eq. (17)] was derived ignoring the difference in the rates of the decay of the wave functions of electrons and holes in the junction (ignoring dephasing [25]). With superconducting electrodes, this approach is justified if $\mu \gg |\Delta^L|, |\Delta^R|$. Dephasing in the barrier is of no significance if

$$\frac{\omega}{V_0} \lesssim \exp(-\lambda_0 d). \tag{19}$$

Condition (19) also allows us to ignore dephasing in the tunneling of electrons and holes to localized states and backward in resonance current calculations [14]. Resonance currents must, however, be calculated taking into account the energy difference between electrons and holes in the preexponential terms in the equations for the Green functions. The necessary accuracy of the expansion in powers of ω/V_0 is then $\exp(-2\lambda_0 d)$ (see Appendix).

The equation for resonance current transport through the junction is obtained from (13) by substituting the remaining terms present under the sum sign in (14),

$$I_{\text{res}}(\varphi) = \frac{eT\Lambda(\mathbf{r}_0)}{2im}\lim_{x'\to x}\left(\frac{\partial}{\partial x'} - \frac{\partial}{\partial x}\right) + \sum_{\omega_n}\sum_{j,l=1}^{2}(-1)^{l+1}G^0_{jl}(\mathbf{r}, \mathbf{r}_0)\Phi G_{jl}(\mathbf{r}_0, \mathbf{r}')\Big|_{x=0}. \tag{20}$$

5. PARTICULAR CASES OF RESONANCE SCATTERING

5.1. General Equation for Resonance Current in the Two-Dimensional DID Junction

Of the greatest interest in the high-T_c superconducting structures is two-dimensional transport in the *ab* plane. In the two-dimensional junction model, the potential parameters of the localized state are selected from the condition that potential well (3) should contain at least one energy level. This procedure is described in detail in [20]; here, we only give final equations. In our problem, the energy parameters of localized states are written in terms of model constants ρ and α as

$$I_1^{-1}(\lambda_0\rho)\left(\frac{\lambda_0}{2\rho\alpha m} - \eta\right)\Big|_{\lambda_0\rho\to 0} \approx \frac{1}{m\alpha\rho^2} - \frac{2\eta}{\lambda_0\rho} = \frac{E_0}{2V_0},$$

where the η parameter is

$$\eta = \frac{1}{\lambda_0\rho} - K_1(\lambda_0\rho)\Big|_{\lambda_0\rho\to 0} \approx -\frac{\lambda_0\rho}{2}\left(\ln\left(\frac{\lambda_0\rho}{2}\right) + \gamma_E - \frac{1}{2}\right).$$

Here, $\gamma_E \approx 0.577$ is the Euler constant, I_1 is the modified Bessel function of the first kind, K_1 is the modified Bessel function of the second kind (the Macdonald function), and E_0 is the energy of the resonance level of the localized state. It then follows from (20) that

$$I^{2D}_{\text{res}} = -2eT\sum_{\omega}\frac{S_1\sin\varphi}{S_0 - S_1\cos\varphi}, \tag{21}$$

where

$$S_1 = 2\gamma_0^2 J_L J_R,$$

$$S_0 = -\omega^2 - E_R^2 - \frac{\Gamma^2}{4} - \omega(2(J_1 + J_2)\gamma_0 - \Gamma) + \gamma_0^2(\Gamma\gamma_0^{-1}(J_1 + J_2) - (J_1 + J_2)^2 - J_L^2 - J_R^2),$$

$$\Gamma_{01} = V_0|t|\exp(\lambda_0(d - 2x_0))\sqrt{\frac{\pi}{\lambda_0 x_0}}, \tag{22}$$

$$\Gamma_{02} = V_0|t|\exp(-\lambda_0(d - 2x_0))\sqrt{\frac{\pi}{\lambda_0(d - x_0)}},$$

$$\Gamma_0^2 = \Gamma_{01}\Gamma_{02}, \quad \Gamma = \Gamma_{01} + \Gamma_{02}, \quad \gamma_0 = 2V_0|t|,$$

$$E_R \approx E_0 - \frac{\lambda_0^2 + \kappa^2}{4\lambda k_F}\Gamma.$$

The Γ_0 parameter is the width of the electronic level state for the localized state situated in the middle of the barrier and E_R is the effective resonance level energy that takes into account the shift caused by a finite width of the barrier.

The other parameters in (21) are determined as follows:

$$\begin{pmatrix} J_1 \\ J_L \end{pmatrix} = \left\langle \frac{\exp(\lambda_0 d - 2\lambda(k_y)x_0)}{\lambda(k_y)\Gamma_1}\begin{pmatrix} 1 \\ \hat{\beta}_\pm^{\text{L}} \end{pmatrix}\right\rangle_{k_y},$$
$$\begin{pmatrix} J_2 \\ J_R \end{pmatrix} = \left\langle \frac{\exp(\lambda_0 d - 2\lambda(k_y)(d - x_0))}{\lambda(k_y)\Gamma_2}\begin{pmatrix} 1 \\ \hat{\beta}_\pm^{\text{R}} \end{pmatrix}\right\rangle_{k_y}. \tag{23}$$

Here, the Fermi momentum of quasi-particles in the barrier is obtained ignoring dephasing (that is, $\lambda(k_y) = \sqrt{\lambda_0^2 + k_y^2}$ and the angle brackets $\langle\ldots\rangle$ denote averaging over the transverse momentum k_y,

$$\langle f(k_y)\rangle_{k_y} = \frac{k_F}{2}\int_{-\pi/2}^{\pi/2} d\theta\cos\theta\cdot f(k_F\sin(\theta)).$$

The $\hat{\beta}_\pm^{\text{L, R}}$ parameters in the integrands in (23) are real and phase-independent components of the $\beta_\pm^{\text{L, R}}$ vari-

ables from (18); they are given by

$$\hat{\beta}_{\pm}^{\mathrm{L,R}} = \frac{\Delta^{\mathrm{L,R}}(T)\cos(2\theta \mp 2\alpha_{L,R})}{\omega + \Omega_{\pm}^{\mathrm{L,R}}}. \quad (24)$$

When writing (23) and deriving (21), we proceeded from the equalities

$$J_L(\hat{\beta}_{+}^{\mathrm{L}}) = J_L(\hat{\beta}_{-}^{\mathrm{L}}), \quad J_R(\hat{\beta}_{+}^{\mathrm{R}}) = J_R(\hat{\beta}_{-}^{\mathrm{R}}).$$

These equalities follow from the symmetry properties of the anisotropic order parameter. They are valid because the integration ranges correspond to the period of the $\hat{\beta}_{\pm}^{\mathrm{L,R}}(\theta)$ functions equal to π. The equalities hold for arbitrary orientation angles $\alpha_{L,R}$.

Equation (21) describes the general case of resonance current transport in two-dimensional structures with an arbitrary symmetry of the order parameter in the electrodes. This is the main result of the present work. The particular form of the $\hat{\beta}_{\pm}^{\mathrm{L,R}}$ parameters determined by (24) is valid for the model with a step dependence of the modulus of the order parameter on spatial coordinates. This result is, however, easy to generalize to a more general case if the $\hat{\beta}_{\pm}^{\mathrm{L,R}}$ parameters are understood as Andreev reflection coefficients of electrons having the corresponding trajectories [20], which can be obtained by numerically solving the problem of the neighborhood effect at the boundary between the superconducting electrode and the dielectric interlayer [7].

5.2. SIS Junctions of Different Dimensionalities

In the limiting case of isotropic electrode order parameters, (21) takes the simplest form. Integrals (23) are calculated analytically, and (21) gives the resonance supercurrent in the SIS junction in the form

$$I_{\mathrm{res}} = eT\Delta^2\sin\varphi\Gamma_0^2\sum_{\omega}\left[\Gamma_0^2\Delta^2\sin^2\frac{\varphi}{2} - (\Delta^2+\omega^2)\left(\omega^2 + E_R^2 + \frac{\Gamma^2}{4}\right) - \omega^2\Gamma\sqrt{\Delta^2+\omega^2}\right]^{-1}. \quad (25)$$

The analytic continuation $\omega \longrightarrow -iE$ in (25) gives the dispersion equation

$$\Gamma_0^2\Delta^2\sin^2\frac{\varphi}{2} + (\Delta^2 - E^2)\left(E^2 - E_R^2 - \frac{\Gamma^2}{4}\right) + E^2\Gamma\sqrt{\Delta^2 - E^2} = 0, \quad (26)$$

which has been studied for one-dimensional models of a short junction $d \ll \xi_0$ (ξ_0 is the coherence length) [13] and generalized to arbitrary ratios between the barrier width and coherence length [14].

The form of (25) and (26) is insensitive to problem dimensions. Different junction dimensions only change the form of the energy parameters of localized states. In the two-dimensional problem, the width of the resonance level and its energy are determined by (22), and, in the one-dimensional geometry, the properties of localized states are determined by parameters similar to those obtained in [13],

$$\Gamma_{01}^{1D} = 2V_0|t|\exp(\lambda_0(d-2x_0)),$$
$$\Gamma_{02}^{1D} = 2V_0|t|\exp(-\lambda_0(d-2x_0)).$$

The characteristic resonance level width for a three-dimensional SIS junction is given by the equations

$$\Gamma_{01}^{3D} = V_0|t|\frac{\exp(\lambda_0(d-2x_0))}{\lambda_0 x_0},$$
$$\Gamma_{02}^{3D} = V_0|t|\frac{\exp(-\lambda_0(d-2x_0))}{\lambda_0(d-x_0)}.$$

It follows that the Beenakker–van Houten dispersion equation [Eq. (26)] obtained in 1991 for the one-dimensional model is of a more general applicability and remains valid for junctions with wide barriers and problems of large dimensions.

Equation (25) allows the summation of the Matsubara frequencies for "broad" ($\Delta/\Gamma_0 \ll 1$) and "narrow" ($\Delta/\Gamma_0 \gg 1$) resonances to be performed analytically. Studies of supercurrent in these two limits lead to the same conclusions as drawn in [14], this time for junctions with two- and three-dimensional geometries. We, however, obtained resonance transparency close to one and a nonsinusoidal phase dependence of supercurrent only at the resonance level energy close to zero ($E_R \sim 0$) and when the localized state was situated close to the barrier center ($x_0 \sim d/2$). A shift of the localized state resonance energy from zero or a displacement of the localized state from the barrier center decreases resonance transparency and supercurrent and restores the sinusoidal phase dependence of supercurrent $I_{\mathrm{res}}(\varphi)$. Averaging over energy and coordinates of localized states makes supercurrent in two- and three-dimensional SIS junctions proportional to $\sin\varphi$ in both limiting cases.

5.3. Resonance Scattering in a Two-Dimensional SID Junction

Resonance current transport in junctions with high-T_c superconducting electrodes will be studied on the assumption that the *ab* plane of the high-T_c superconducting crystal is perpendicular to the surface of the

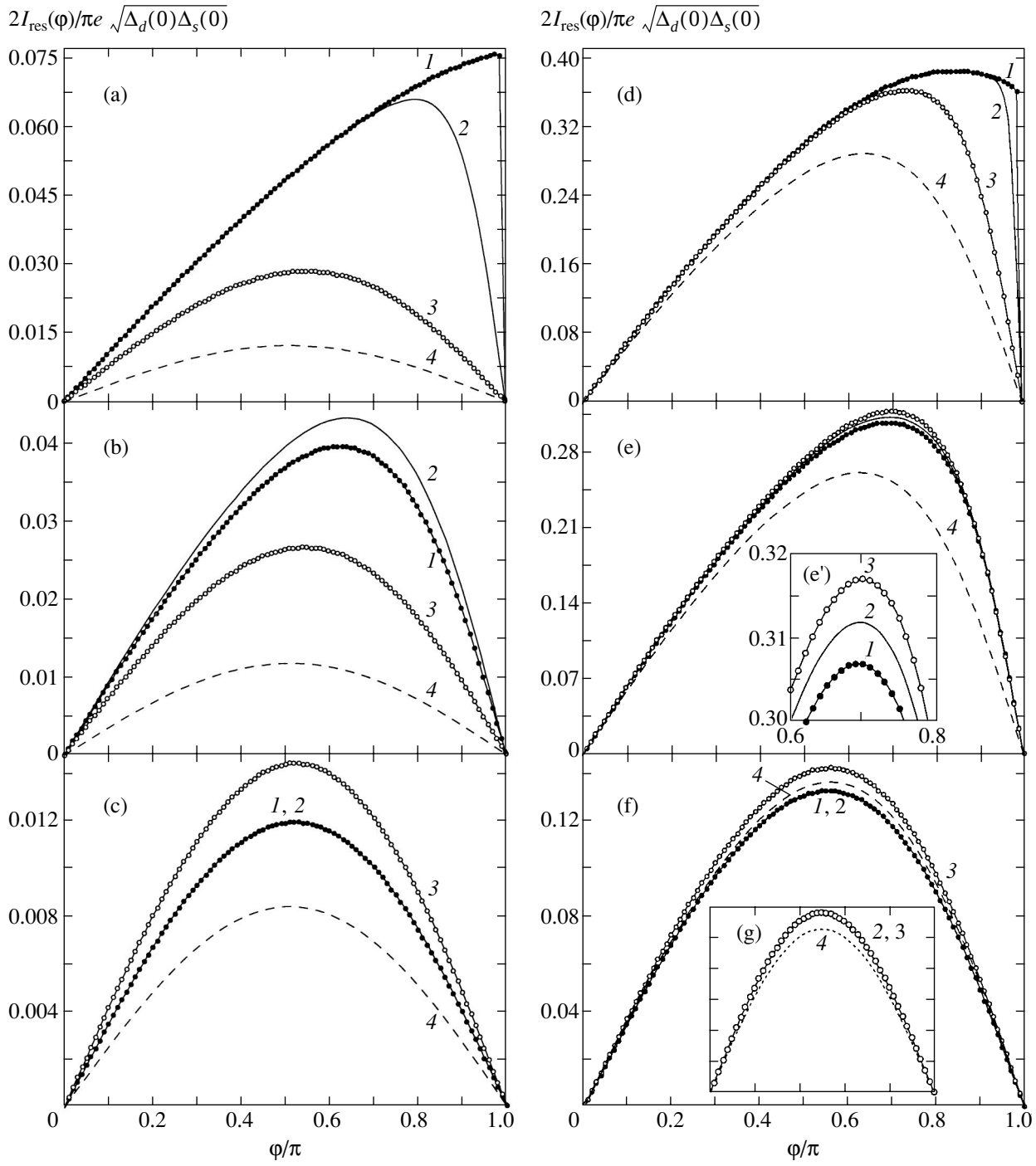


Fig. 2. Phase dependences of supercurrent in the SID junction for $\gamma_0/\Delta_d(0)$ = (a, b, c) 0.1 and (d, e, f) 10 at $\lambda d = 6$, $k/\lambda = 2$, $E_R = 0$, and $x_0 = d/2$; α_R = (a, d) 0, (b, e) $\pi/12$, and (c, f) $\pi/6$; and T/T_s = (*1*) 0, (*2*) 0.05, (*3*) 0.3, and (*4*) 0.6. Inset e' shows the region of maximum supercurrent corresponding to curves *1*, *2*, and *3* in Fig. 2g. Phase dependences of nonresonance supercurrent are plotted in Fig. 2g for the same parameters in arbitrary units.

junction and the order parameter is determined by (16). Electronic and hole excitations at the boundary of an anisotropic superconductor are influenced by different order parameters if $\alpha \neq 0$. If the sign of the order parameter changes after the reflection of a quasi-particle from the surface of the high-T_c superconductor, bound Andreev states appear on the surface of the superconductor with zero energy with respect to the Fermi level (zero energy states or mid gap states) [21]. The region of the appearance of zero energy states depends on the angle of high-T_c superconductor crystal lattice orientation. For instance, for the SID junction, the order

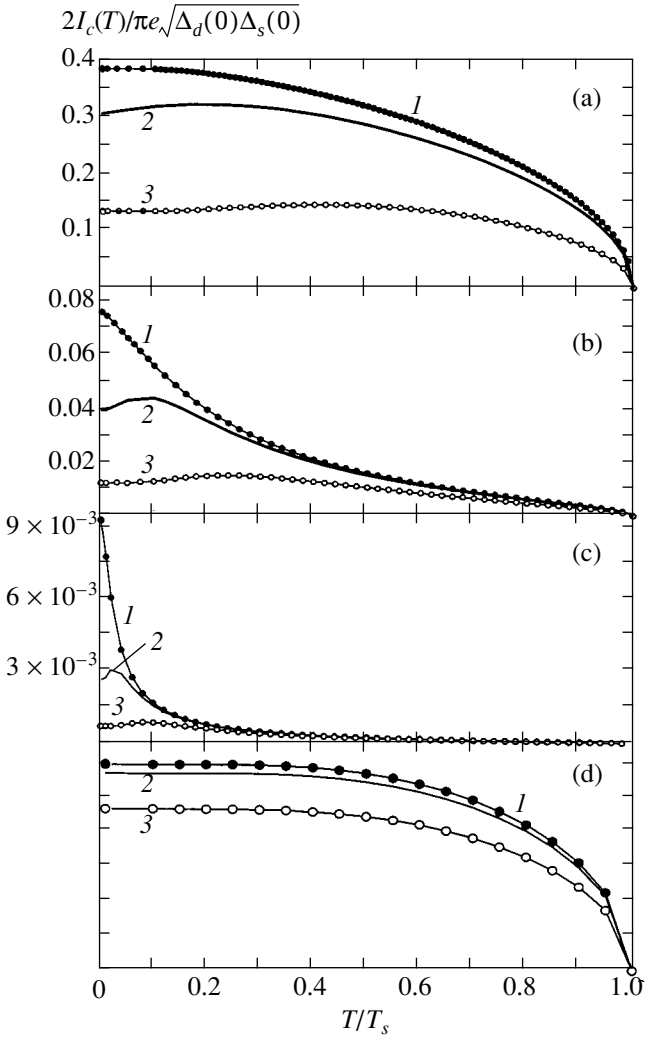


Fig. 3. Temperature dependences of supercurrent $I_c(T)$ in the SID junction for $\gamma_0/\Delta_d(0)$ = (a) 10, (b) 0.1, and (c) 0.01 at $\lambda d = 6$, $k/\lambda = 2$, $E_R = 0$, $x_0 = d/2$, and α_R = (*1*) 0, (*2*) $\pi/12$, and (*3*) $\pi/6$. Temperature dependences of nonresonance supercurrent are plotted in Fig. 3d in arbitrary units for the same parameters.

parameter has the form

$$\Delta(x, T) = \begin{cases} \Delta_s(T)e^{i\varphi_L}, & x < 0 \\ 0, & 0 < x < d \\ \Delta_d(T)\cos(2\theta \mp 2\alpha_R)e^{i\varphi_R}, & x > d. \end{cases} \quad (27)$$

Andreev states then appear on the trajectories satisfying the condition

$$\theta \in (\pm\pi/4 - \alpha_R; \pm\pi/4 + \alpha_R).$$

The question of how the appearance of zero energy states influences resonance current is of interest.

The phase dependences of supercurrent and the temperature dependences of critical current are shown in Figs. 2 and 3. The dependences were obtained on the assumption that the order parameters $\Delta_s(T)$ and $\Delta_d(T)$ satisfied the Bardeen–Cooper–Schrieffer theory equations. Numerical calculations were performed for the ratio between the critical temperatures of the superconductors $T_s/T_d = 1/9$. At $\alpha_R = 0$, the behavior of the $I_{res}(\varphi)$ curves in "narrow" and "broad" resonances is similar to that for junctions with isotropic electrode order parameters; namely, we observe phase dependences close to $\sin(\varphi/2)$ at low temperatures and resonance transparency equal to one. As in Section 5.2, shift of the resonance energy of the localized state from zero or a displacement of the localized state from the middle of the barrier decreases resonance transparency, causes a drop in resonance supercurrent, and restores the sinusoidal dependence of $I_{res}(\varphi)$. After averaging over localized state energies and coordinates, resonance current becomes proportional to $\sin\varphi$ in both limiting cases.

Figures 2a and 2d are visual illustrations of deviations of the $I(\varphi)$ dependence from $\sin\varphi$ at low temperatures. An increase in the temperature or the orientation angle α_R makes the curves closer to sinusoidal (see Figs. 2b, 2c, 2e, 2f). For comparison, the phase dependences of tunnel supercurrent at the same junction parameters and $\alpha_R = 0$ are shown in Fig. 2g; these dependences follow from the equation for potential supercurrent [Eq. (17)] and the results obtained in [3]. We see that the phase dependences of tunnel supercurrent are sinusoidal at low temperatures under long junction conditions.

The temperature dependences of the critical current for the SID junction at various superconductor crystal orientations are shown in Fig. 3. Figures 3a, 3b, and 3c demonstrate a sharper decrease in I_c as the temperature increases caused by resonance narrowing. At the same time, the theory described in [3] predicts a smooth decrease in the critical tunnel current as the temperature increases (Fig. 3d). The large slopes of the temperature dependences of resonance current are explained by strong resonance level "smearing" at $T > \Gamma_0$; accordingly, the narrower the resonance, that is, the smaller the resonance level width Γ_0, the lower the temperature of the onset of critical current drop as the temperature increases. A visual illustration of such a behavior of $I_c(T)$ is given by Fig. 3.

We see from Figs. 2 and 3 that an increase in the angle of high-T_c superconductor crystal orientation causes a decrease in supercurrent. This process is shown in detail in Fig. 4, where the dependences of critical current on the angle α_R are plotted. According to this figure, "isothermal" curves decrease as the α_R angle increases, and, at $\alpha_R = \pi/4$, the resonance current is identically equal to zero, which is formally related to the equality to zero of the J_R integral in (21) (at $\alpha_R = \pi/4$, the integrand becomes antisymmetric with respect

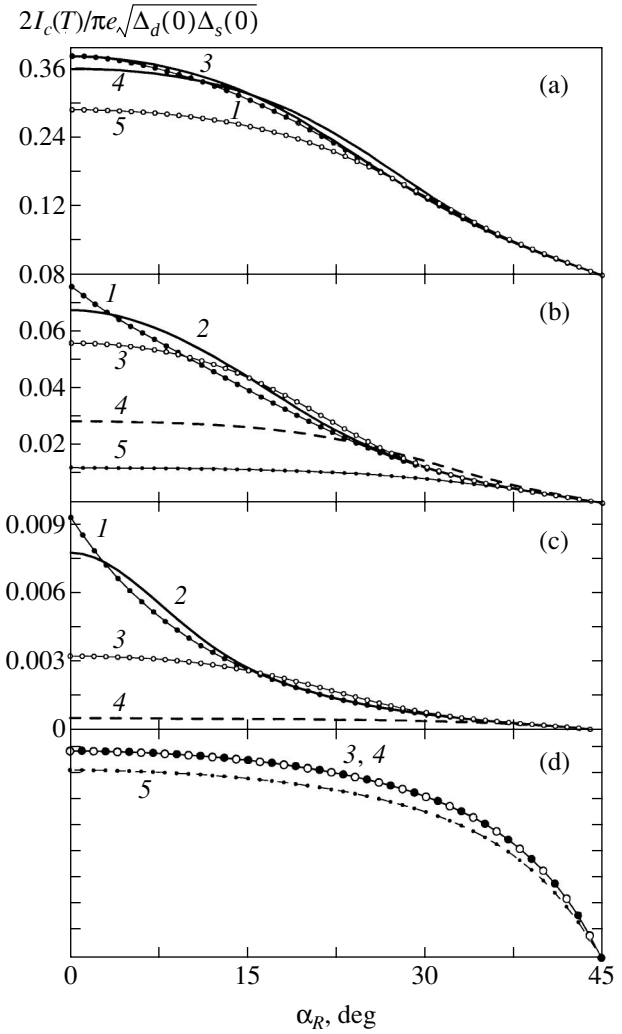


Fig. 4. Supercurrent $I_c(\alpha_R)$ in the SID junction for $\gamma_0/\Delta_d(0)$ = (a) 10, (b) 0.1, and (c) 0.01 at $\lambda d = 6$, $k/\lambda = 2$, $E_R = 0$, $x_0 = d/2$ and T/T_s = (*1*) 0, (*2*) 0.01, (*3*) 0.05, (*4*) 0.3, and (*5*) 0.6. Angular dependences of nonresonance supercurrent are plotted in Fig. 4d in arbitrary units for the same parameters.

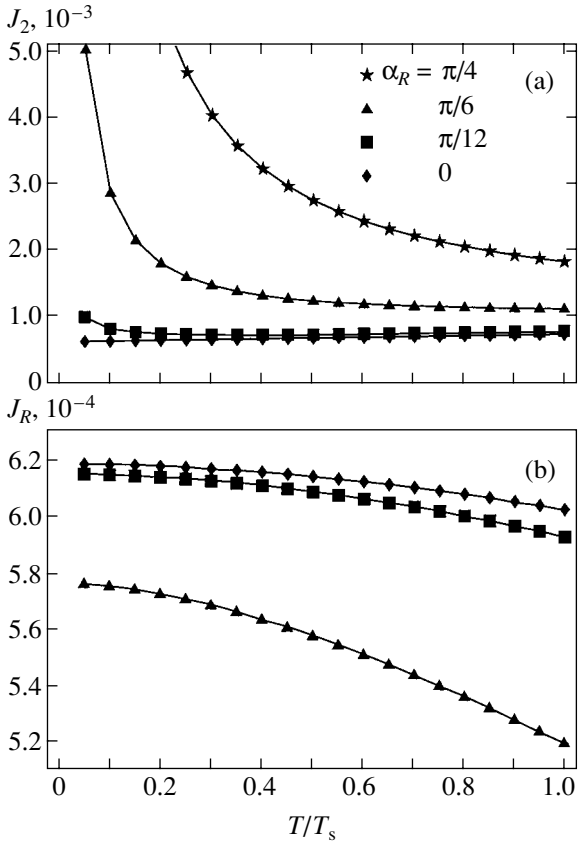


Fig. 5. Temperature dependences of the integrals (a) J_2 and (b) J_R in the SID transition for various α_R angles. The plots were constructed for the Matsubara frequency $\omega = \pi T$ and for a narrow resonance with $\gamma_0/\Delta_d(0) = 0.1$. The other parameter values: $\lambda d = 6$, $k/\lambda = 2$, $E_R = 0$, and $x_0 = d/2$.

to θ, and J_R therefore vanishes under averaging over angles). Physically, the absence of resonance current at $\alpha_R = \pi/4$ is easily explained from symmetry considerations. When quasi-particles are rescattered by localized states, the probabilities of the trajectories toward the high-T_c superconductor with the angles $\pm\theta$ with respect to the normal are equal. At the same time, quasiparticles interacting with the high-T_c superconducting electrodes are influenced by order parameters of different signs. As a result, the contributions to the resonance current of the directions $\pm\theta$ are equal in magnitude and opposite in sign. For this reason, averaging over all trajectories gives zero supercurrent value. A similar situation arises with potential tunneling [3]. In SID junctions at $\alpha_R = \pi/4$, the coefficient of $\sin\varphi$ in the equation for supercurrent vanishes and the contribution of terms proportional to $\sin 2\varphi$ therefore becomes predominant. This is, in particular, the reason for changes in the periodicity of the phase dependence of the potential supercurrent in such junctions [26, 27]. In the equation for resonance supercurrent [Eq. (21)], the terms proportional to $\sin 2\varphi$, like the other (higher) terms of the trigonometric series $\sin n\varphi$, synchronously vanish. For this reason, the resonance current vanishes at $\alpha_R = \pi/4$.

An analysis of the plots shown in Figs. 4a and 4b shows that resonance narrowing results in a rapid decrease in the critical current as the angle of high-T_c superconductor orientation increases. A comparison of the $I_c(\alpha_R)$ curves calculated for resonance and direct tunneling (see Fig. 4d) shows that potential supercurrent decreases more monotonically as α_R increases.

An interesting feature of the temperature dependences of resonance supercurrent shown in Fig. 3 is their anomalous decrease at low temperatures and nonzero orientation angles of high-T_c superconductors. Formally, such a behavior of the temperature curves is explained as follows: whereas the other integrals in (21) show a fairly monotonic behavior (see Fig. 5b, where

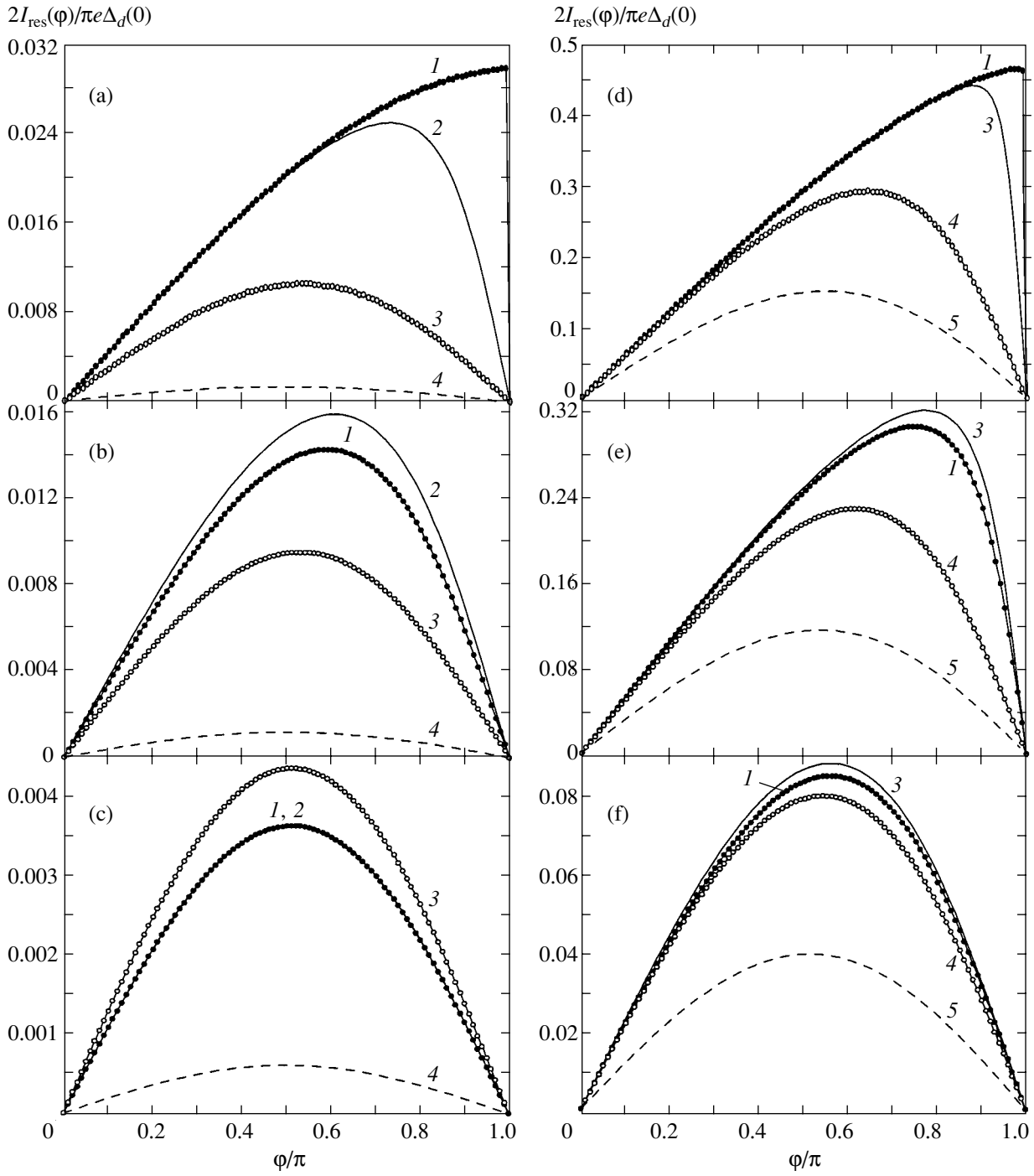


Fig. 6. Phase dependences of supercurrent $I_{res}(\varphi)$ in the DID junction for $\gamma_0/\Delta_d(0)$ = (a, b, c) 0.1 and (d, e, f) 10 and $\alpha_L = \alpha_R$ at $\lambda d = 6$, $k/\lambda = 2$, $E_R = 0$, $x_0 = d/2$ and α_L = (a, d) 0, (b, e) $\pi/12$, and (c, f) $\pi/6$, and T/T_d = (*1*) 0, (*2*) 0.01, (*3*) 0.05, (*4*) 0.3, and (*5*) 0.6.

the $J_R(T)$ dependence is plotted by way of example for various orientation angles of high-T_c superconductors), the J_2 integral rapidly increases as the temperature approaches zero (see Fig. 5a). For this reason, the denominator in (21) is minimum at nonzero temperature.

The J_2 and J_R integrals describe interactions between two resonance tunneling processes, namely, through localized states and through bound Andreev states that appear on the surface of high-T_c superconductors at $\alpha_R \neq 0$. It was shown in [20] that the interference of these processes did not necessarily cause supercurrent amplification. Moreover, at zero voltage at the junction, the interference current component related to rescattering of quasi-particles in the Andreev channel of tunneling at localized states can make a negative

contribution to the supercurrent. This is the effect of complex interference of two resonance processes that may be responsible for a sharp decrease in the angle dependence of the critical resonance current compared with the potential current and the deficiency of the critical current at zero orientation angles of high-T_c superconductors noticeable in the region of low temperatures (Figs. 3a, 3b, 3c).

5.4. RESONANCE CURRENT IN A TWO-DIMENSIONAL DID JUNCTION

In the general case of anisotropic order parameters, not all integrals (23) can be calculated analytically. For symmetrical junctions with high-T_c superconducting edges, analytic equations can only be obtained from (21) for a narrow resonance at $\alpha_L = \alpha_R = 0$ and $k/\lambda \gg 1$. Equations similar to (25) and (26) with $\Delta = \Delta_d(T)$ are then obtained for supercurrent and the energy spectrum of resonance junction states with equal order parameters in the left and right electrodes. In the general case of arbitrary k/λ ratios and orientation angles of high-T_c superconducting crystals, resonance junction conduction can only be analyzed numerically.

Equation (21) for current can, however, be simplified for junction configurations of practical interest with orientation angles of the high-T_c superconducting electrodes of $\alpha_L = \pm\alpha_R$. If the condition

$$\Delta^{\mathrm{L}}(T) = \Delta^{\mathrm{R}}(T) = \Delta_d(T)$$

is met, we have $J_1 = J_2$ and $J_L = J_R$ for integrals (23). It follows that junctions with the orientation angles $|\alpha_L| = |\alpha_R|$ have equal resonance properties. The resonance transport of current is sharply different in this respect from the direct tunneling current. It follows from [3] and Eq. (17) obtained in this work that the direct supercurrent behaves differently in junctions with the configurations $\alpha_L = \pm\alpha_R$.

Applying (21) to symmetrical junctions yields

$$I_{\mathrm{res}}^{2D\ DID}\Big|_{|\alpha_L| = |\alpha_R|} = -4eT\sum_{\omega}\frac{\gamma_0^2 J_d^2 \sin\varphi}{S_d},$$

$$J_d = J_L = J_R, \quad J = J_1 = J_2,$$

$$S_d = -\omega^2 - E_R^2 - \frac{\Gamma^2}{4} - \omega(4J\gamma_0 - \Gamma) \tag{28}$$

$$+ 2\gamma_0^2\left(\Gamma\gamma_0^{-1}J - 2J^2 - J_d^2\cos^2\frac{\varphi}{2}\right).$$

Equation (28) will be analyzed numerically on the assumption that the *ab* planes of the *d*-type superconducting crystals to the left and right of the insulating interlayer are perpendicular to the surface of the junction, and the order parameter of the problem is given by (16).

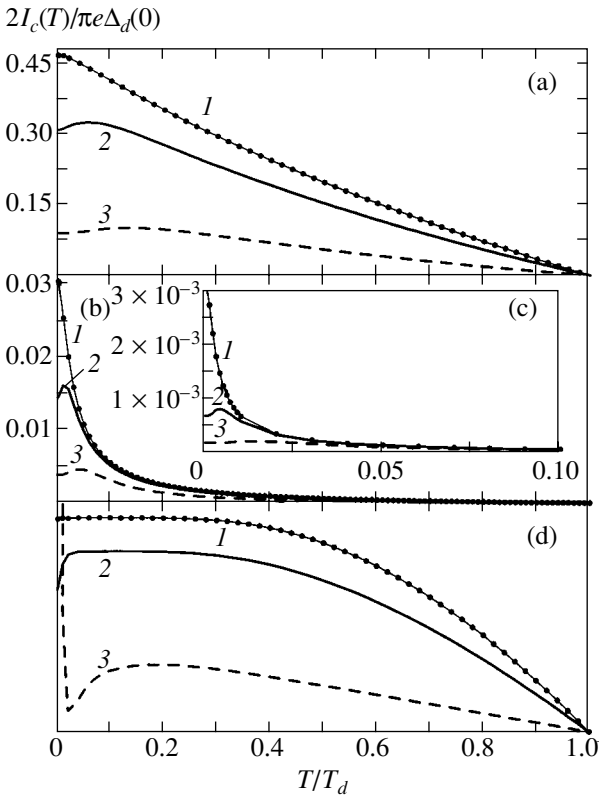


Fig. 7. Temperature T/T_d dependences of critical current $I_c(T)$ in the DID junction for $\gamma_0/\Delta_d(0)$ = (a) 0.1 and (b) 10 and $\alpha_L = \alpha_R$ at $\lambda d = 6$, $k/\lambda = 2$, $E_R = 0$, $x_0 = d/2$, and α_L = (*1*) 0, (*2*) $\pi/12$, and (*3*) $\pi/6$.

The phase dependences of the resonance supercurrent obtained using (28) are shown in Fig. 6. If both orientation angles α_R and α_L are nonzero, the behavior of the phase curves is similar to that of the curves for SID junctions, as follows from Figs. 6b, 6c, 6e, and 6f. We can only note that supercurrent decreases more rapidly as the orientation angles of the high-T_c superconductors increase in magnitude. In addition, if $\alpha_{L,\,R} \neq 0$, the phase curves of DID junctions more rapidly approach sinusoidal dependences because of the presence of two regions with anisotropic order parameters instead of one such region in SID junctions. As with SIS junctions, a displacement of the localized state from the center of the barrier or a deviation of the resonance energy from the Fermi level causes a sharp decrease in supercurrent and the disappearance of the deviations of phase characteristics from the sinusoidal dependence. As a result, resonance current averaging over the resonance energy E_R and the coordinate x_0 of the localized state gives a sinusoidal phase dependence of the supercurrent.

Similarity of the resonance tunneling processes in SID and DID junctions is also observed when we com-

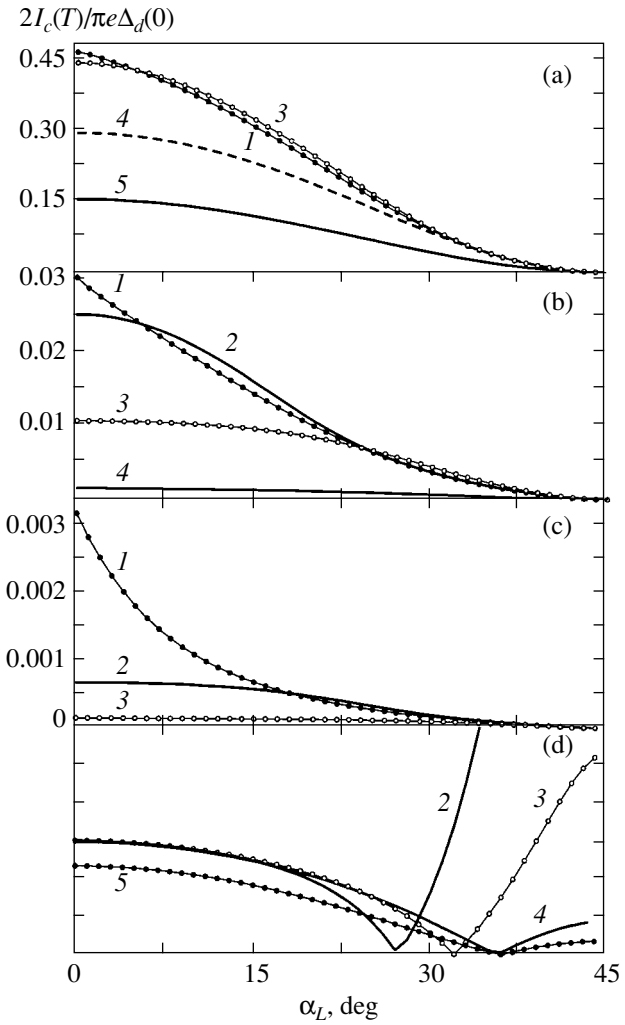


Fig. 8. Dependences $I_c(\alpha_L)$ of critical current in the DID junction for $\gamma_0/\Delta_d(0)$ = (a) 10, (b) 0.1, and (c) 0.01 and $\alpha_L = \alpha_R$ at $k/\lambda = 2$, $\lambda d = 6$, $E_R = 0$, $x_0 = d/2$, and T/T_s = (*1*) 0, (*2*) 0.01, (*3*) 0.05, (*4*) 0.3, and (*5*) 0.6. Angle dependences of nonresonance supercurrent are plotted in Fig. 8d in arbitrary units for the same parameters.

pare the temperature dependences of the critical current. A rapid decrease in supercurrent as the temperature increases or resonance narrows in DID junctions is shown in Figs. 7a and 7b. Note also that the temperature curves at $\alpha_{L,R} \neq 0$ exhibit anomalous behavior at low temperatures; similar anomalies have been given attention above, when resonance supercurrent in SID junctions was analyzed. At large orientation angles of high-T_c superconducting crystals, we do not observe a sharp increase in the critical current in resonance current plots such as is characteristic of direct tunneling (Fig. 7d, curve *3*). The situation with the angle dependences of resonance current in DID junctions is similar. Interestingly, according to [3], the orientation angle dependence of direct supercurrent should be very far from monotonic (see Fig. 8d). At the same time, the dependences of resonance supercurrent shown in Figs. 8a–8c are monotonic, which lends support to the conclusions made in describing the angle dependences of the critical current in SID junctions.

To summarize, our study of the resonance supercurrent in junctions in which both electrodes have *d*-symmetry order parameters shows that the supercurrent value is determined by the moduli of the orientation angles of superconductor crystals rather that the angles themselves. For this reason, structures with symmetric ($\alpha_L = \alpha_R$) and antisymmetric ($\alpha_L = -\alpha_R$) electrode configurations that we often come across in theoretical and experimental studies are equivalent. Indeed, in resonance tunneling, a quasi-particle falls into a localized state potential well and "forgets" about the initial direction of its motion. Subsequent particle rescattering in the directions $\pm\theta$ with respect to the *x* axis is equiprobable. As a result, physical resonance tunneling processes in structures with $\alpha_L = \pm\alpha_R$ are spatially symmetrical and do not differ from each other. In reality, the existence of a localized state between two high-T_c superconducting electrodes with pronounced anisotropic properties is responsible for isotropy of not only resonance tunneling. It was shown in [20] that the presence of a localized state along the trajectory of the appearance of bound zero energy states results in rescattering of particles in other directions, which weakens the zero bias anomaly effect.

6. CONCLUSIONS

We have presented calculations of resonance supercurrent transport in equilibrium Josephson junctions of different configurations, determined the supercurrent using the formalism of Green functions, and obtained an equation for the Green function in a nonhomogeneous potential barrier with a localized state. For the two-dimensional junction model, we derived an analytic equation for resonance supercurrent valid for an arbitrary symmetry of electrode order parameters. For the SIS junction, we generalized the Beenakker–van Houten dispersion equation obtained for the one-dimensional model [13] to junctions of arbitrary dimensions. We numerically studied the dependences of the resonance supercurrent on the macroscopic phase, temperature, and high-T_c superconductor crystal orientation angles with respect to the direction of current propagation. A comparison of these results with those for tunnel current transport through the potential barrier reveals a steeper resonance current drop as the temperature or the orientation angle of the high-T_c crystals deviates from zero. For narrow resonances usually observed in experiments, finite temperatures and nonzero orientation angles substantially decrease the resonance supercurrent. Our results lead us to conclude that, in long high-T_c superconducting junctions and in the presence of localized states in the interlayer, resonance supercurrent is indeed suppressed compared

with direct tunneling through the barrier, in conformity with [2].

Analysis of the results obtained in this work has substantiated the conclusions drawn in [20]; namely, the interaction of two resonance processes of tunneling through localized states and bound Andreev levels (zero energy states) can weaken supercurrent and the effects related to anisotropy of high-T_c superconductors.

The approach developed above ignores the suppression of the order parameter close to the boundaries of high-T_c superconducting structures. It follows that we did not study the contribution to supercurrent transport caused by resonance tunneling to Andreev levels with nonzero energies localized near the boundary. Exactly including this contribution requires numerical calculations. Nevertheless, the analysis performed in this work leads us to conclude that taking into account these effects would not cause serious qualitative changes.

ACKNOWLEDGMENTS

This work was financially supported by the Ministry of Education and Technology of the Russian Federation.

APPENDIX

Let the source of the Green function lie within the barrier. The unperturbed Green function can conveniently be found by writing (6) for the Fourier transform of $\mathbf{G}_\omega^0(\mathbf{r}, \mathbf{r}')$. In the two-dimensional model, the Fourier transform of the two-component Green function $\mathbf{G}_\omega^0(\mathbf{r}, \mathbf{r}')$ is given by

$$\mathbf{G}_\omega^0(\mathbf{r}, \mathbf{r}') = \frac{1}{2\pi}\int dk_y \exp(-ik_y(y-y'))\mathbf{G}_\omega^0(k_y, x, x').$$

If the barrier potential $V(\mathbf{r})$ is homogeneous, that is, if the presence of the localized state is ignored, the Green function $\mathbf{G}_\omega^0(\mathbf{r}, \mathbf{r}')$ in the Gor'kov equations can be partitioned in such a way that one of its components be a slowly varying function. Indeed, the Green function experiences oscillations at a length of about k_F^{-1}, whereas the order parameter changes over much longer distances, on the order of the coherence length [24]. We can therefore redefine the Green function and the order parameter accordingly,

$$\begin{aligned}\mathbf{G}_\omega^0(k_y, x, x') &= \exp(\pm ik_x x)\bar{\mathbf{G}}_\omega^0(k_y, x, x'),\\ \Delta(x, \theta) &= \int dx_1 \exp(-i\mathbf{k}\cdot\mathbf{r}_1)\Delta(x, x_1),\end{aligned} \tag{29}$$

where $\bar{\mathbf{G}}_\omega^0$ is the slowly varying component of the unperturbed Green function of the problem and

$$\exp(i\theta) = \frac{k_x}{|\mathbf{k}|} + \frac{ik_y}{|\mathbf{k}|}.$$

This allows us to linearize system (6) and write it in the form of the matrix equation

$$\left\{i\omega\hat{I} - \sigma_z\hat{h}_0 - \frac{(\Delta(x, \theta)\sigma_+ + \Delta^*(x, \theta)\sigma_-)}{2}\right\}$$

$$\times \bar{\mathbf{G}}_\omega^0(k_y, x, x') = \delta(x-x')\hat{I}, \tag{30}$$

$$\hat{h}_0 = \mp\frac{ik_x}{m}\frac{\partial}{\partial x} + V_{\text{rect}} - \mu - \frac{k_y^2}{2m}.$$

The anisotropic order parameter $\Delta(x, \theta)$ depends on the angle θ of the propagation of quasi-particles with respect to the x axis. In the operator $\hat{h}_0$, we omitted the $\partial^2/\partial x^2$ terms, which can be ignored if the condition $\Delta/\mu \ll 1$ is satisfied.

Representing the x components of the Green functions in the form of plane waves, we can write

$$\bar{\mathbf{G}}_\omega^0(k_y, x, x') = \exp(\xi x)\bar{\mathbf{G}}_\omega^0(k_y). \tag{31}$$

Substituting (31) into homogeneous equation (30) yields the final matrix equation that can be used to determine the Green functions in a superconductor,

$$\left\{i\omega\hat{I} - \sigma_z h_\xi - \frac{(\Delta(x, \theta)\sigma_+ + \Delta^*(x, \theta)\sigma_-)}{2}\right\}$$

$$\times \bar{\mathbf{G}}_\omega^0(k_y) = 0, \tag{32}$$

$$h_\xi = \mp\frac{ik_x\xi}{m} + V_{\text{rect}} - \mu - \frac{k_y^2}{2m}.$$

System (32) separates into two pairs of equations for determining the $G_{11,21}$ and $G_{12,22}$ functions. The condition of the equality of the determinant to zero is used to find the ξ parameter,

$$\xi_\pm^{(\pm)} = \pm\frac{m^2}{k_x^2}\sqrt{|\Delta_\pm|^2 + \omega^2}, \tag{33}$$

where the notation $\Delta_\pm = \Delta(x, \theta_\pm)$ is used to stress that quasi-particles in anisotropic superconductors are influenced by different order parameters depending on the direction of their motion; there is a simple relation between the $\theta_\pm$ angles, namely, $\theta_+ = \theta$ and $\theta_- = \pi - \theta$. The sign in (33) is selected depending on whether the

quasi-particle is an electron or a hole. In addition, the relation between the Green function components,

$$\bar{G}^0_{21}(k_y)^{(\pm)}_{\pm} = \frac{i\Delta^*_{\pm}}{-\omega \pm \frac{k_x}{m}\xi^{(\pm)}_{\pm}}\bar{G}^0_{11}(k_y)^{(\pm)}_{\pm}, \quad (34)$$

follows from (32). The selection of different signs in (33) and (34) allows us to obtain four independent solutions for the Green function components $G^0_{11,21}$ in a superconductor, namely,

$$\begin{pmatrix} G^0_{11} \\ G^0_{21} \end{pmatrix}^{(\pm)}_{\pm} = \exp[(\pm ik_x + \xi^{(\pm)}_{\pm})x]\begin{pmatrix} \bar{G}^0_{11}(k_y) \\ \bar{G}^0_{21}(k_y) \end{pmatrix}^{(\pm)}_{\pm}. \quad (35)$$

The solution for the unperturbed Green functions in the barrier is found from (6) taking into account that the order parameter in a normal material is zero,

$$\{i\omega\hat{I} - \sigma_z\hat{h}_n\}\mathbf{G}^0_{\omega}(k_y, x, x') = \delta(x-x')\hat{I},$$
$$\hat{h}_n = \frac{\nabla^2}{2m} + V_{\text{rect}} - \mu - \frac{k_y^2}{2m}. \quad (36)$$

Four independent general solutions can be obtained from (36) for the Green functions $G^0_{11,21}$. The solutions responsible for the transport of electron-like excitations in the barrier are described by plane waves,

$$\begin{pmatrix} G^0_{11} \\ G^0_{21} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}\exp(\pm\lambda x). \quad (37)$$

The solution for holes in the barrier is

$$\begin{pmatrix} G^0_{11} \\ G^0_{21} \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}\exp(\pm\lambda^* x), \quad (38)$$

where $\lambda = \sqrt{\lambda_0^2 + k_y^2 - 2mi\omega}$ characterizes the two-dimensional momentum of electrons in the barrier. The particular solution to (37) for the $G^0_{11,21}$ components is written in the form

$$-\begin{pmatrix} 1 \\ 0 \end{pmatrix}\frac{m}{\lambda}\exp(-\lambda|x-x'|). \quad (39)$$

It follows that the $G^0_{11,21}$ Green functions of the problem in various junction regions are representable in the form of the superposition of the corresponding independent solutions (35), (37)–(39) provided that the functions are finite at infinity,

$$\begin{pmatrix} G^0_{11} \\ G^0_{21} \end{pmatrix} = \begin{cases} \begin{pmatrix} -\beta^{L*}_{+} \\ 1 \end{pmatrix}A_1\exp(\kappa^L_+ x) + \begin{pmatrix} 1 \\ \beta^L_- \end{pmatrix}A_3\exp(\kappa^{L*}_- x), & x<0, \\ \begin{pmatrix} 1 \\ 0 \end{pmatrix}\left(B_1e^{\lambda x} + B_2e^{-\lambda x} - \frac{m}{\lambda}e^{-\lambda|x-x'|}\right) + \begin{pmatrix} 0 \\ 1 \end{pmatrix}(B_3e^{\lambda^* x} + B_4e^{-\lambda^* x}), & 0<x<d, \\ \begin{pmatrix} 1 \\ \beta^R_+ \end{pmatrix}A_2\exp(-\kappa^{R*}_+ x) + \begin{pmatrix} -\beta^{R*}_- \\ 1 \end{pmatrix}A_4\exp(-\kappa^R_- x), & x>d. \end{cases} \quad (40)$$

The two other Green function components $G^0_{22,12}$ are found from (32) and (36) in a similar way,

$$\begin{pmatrix} G^0_{22} \\ G^0_{12} \end{pmatrix} = \begin{cases} \begin{pmatrix} 1 \\ -\beta^{L*}_{+} \end{pmatrix}C_1\exp(\kappa^L_+ x) + \begin{pmatrix} \beta^L_- \\ 1 \end{pmatrix}C_4\exp(\kappa^{L*}_- x), & x<0, \\ \begin{pmatrix} 1 \\ 0 \end{pmatrix}\left(D_1e^{\lambda^* x} + D_2e^{-\lambda^* x} + \frac{m}{\lambda^*}e^{-\lambda^*|x-x'|}\right) + \begin{pmatrix} 0 \\ 1 \end{pmatrix}(D_3e^{\lambda x} + D_4e^{-\lambda x}), & 0<x<d, \\ \begin{pmatrix} \beta^R_+ \\ 1 \end{pmatrix}C_2\exp(-\kappa^{R*}_+ x) + \begin{pmatrix} 1 \\ -\beta^{R*}_- \end{pmatrix}C_3\exp(-\kappa^R_- x), & x>d. \end{cases} \quad (41)$$

The A_i, B_i, C_i, and D_i coefficients (i = 1, 2, 3, 4) of the exponential functions are found from the condition of continuity of the Green functions and their derivatives at the structure boundaries. Equations (40) and (41) are written using notation (18). We also introduced the Fermi momentum of quasi-particles in the superconducting electrodes in these equations,

$$\kappa_{\pm}^{\mathrm{L,R}} = ik_x + \frac{m}{k_x}\Omega_{\pm}^{\mathrm{L,R}}.$$

On the assumption that the Fermi level energy $\mu \gg |\Delta^L|$, $|\Delta^R|$ and the contribution, to the current, of terms with high Matsubara frequencies is negligibly small, we can ignore dephasing in the superconducting electrodes [25]; that is, the difference between the Fermi momenta of electrons and holes can be considered insignificant, and the approximate definitions $\kappa_{\pm}^{\mathrm{L,R}} \approx ik_x = \kappa$ can be used. In addition, we can ignore dephasing in all the exponents and in the equations where the Fermi momentum of interlayer electrons λ is in some way compared with κ. From these considerations, the equations for the coefficients of (40) and (41) as functions of the x' variable can be written in the form

$$\begin{pmatrix} B_1 \\ B_2 \end{pmatrix} = \frac{1}{2\lambda}[A_3(\lambda \mp \kappa) - \beta_+^{\mathrm{L}*}A_1(\lambda \pm \kappa)] + \begin{pmatrix} 1 \\ 0 \end{pmatrix}\frac{m}{\lambda}e^{-\lambda x'},$$

$$\begin{pmatrix} B_3 \\ B_4 \end{pmatrix} = \frac{1}{2\lambda^*}[A_1(\lambda \pm \kappa) + \beta_-^{\mathrm{L}}A_3(\lambda \mp \kappa)],$$

$$A_1 = -\frac{m}{\kappa Z}[d_e^{\rightarrow *}(x')\beta_-^{\mathrm{L}}\Gamma_2 - d_e^{\leftarrow}(x')t_h(\beta_-^{\mathrm{L}} - \beta_+^{\mathrm{R}})],$$

$$A_2 = -\frac{m}{\kappa Z}[d_e^{\rightarrow *}(x')t_e(\Gamma_3 - \Gamma_1) + d_e^{\leftarrow}(x')\Gamma_1],$$

$$A_3 = -\frac{m}{\kappa Z}[d_e^{\rightarrow}(x')\Gamma_2 - d_e^{\leftarrow *}(x')t_e(\Gamma_2 - \Gamma_4)],$$

$$A_4 = -\frac{m}{\kappa Z}[d_e^{\rightarrow}(x')t_h(\beta_-^{\mathrm{L}} - \beta_+^{\mathrm{R}}) + d_e^{\leftarrow *}(x')\beta_+^{\mathrm{R}}\Gamma_1],$$

$$\begin{pmatrix} D_1 \\ D_2 \end{pmatrix} = \frac{1}{2\lambda^*}[C_1(\lambda \pm \kappa) + \beta_-^{\mathrm{L}}C_4(\lambda \mp \kappa)] - \begin{pmatrix} 1 \\ 0 \end{pmatrix}\frac{m}{\lambda^*}e^{-\lambda x'},$$

(42)

$$\begin{pmatrix} D_3 \\ D_4 \end{pmatrix} = \frac{1}{2\lambda}[C_4(\lambda \pm \kappa) - \beta_+^{\mathrm{L}*}C_1(\lambda \mp \kappa)],$$

$$C_1 = -\frac{m}{\kappa Z}[d_e^{\leftarrow}(x')t_h(\Gamma_3 - \Gamma_2) + d_e^{\rightarrow *}(x')\Gamma_2],$$

$$C_2 = \frac{m}{\kappa Z}[d_e^{\rightarrow *}(x')t_e(\beta_+^{\mathrm{L}*} - \beta_-^{\mathrm{R}*}) + d_e^{\leftarrow}(x')\beta_-^{\mathrm{R}*}\Gamma_1],$$

$$C_3 = -\frac{m}{\kappa Z}[d_e^{\rightarrow}(x')t_h(\Gamma_4 - \Gamma_1) + d_e^{\leftarrow *}(x')\Gamma_1],$$

$$C_4 = \frac{m}{\kappa Z}[d_e^{\rightarrow}(x')\beta_+^{\mathrm{L}*}\Gamma_2 + d_e^{\leftarrow *}(x')t_e(\beta_-^{\mathrm{R}*} - \beta_+^{\mathrm{L}*})].$$

The Z and $\Gamma_{1,2,3,4}$ parameters are defined by (18). Some of the parameters of (42) have a simple physical meaning. For instance, t_e is the transparency coefficient of the barrier for the electron, $t_h = t_e^*$ is the transparency coefficient of the barrier for the hole, $d_e^{\rightarrow}(x')$ and $d_e^{\rightarrow *}(x')$ are the probabilities of electron and hole tunneling from the right superconductor to the localized state region, and $d_e^{\leftarrow}(x')$ and $d_e^{\leftarrow *}(x')$ are the probabilities of reaching the localized state by the quasi-particles that move from the left superconductor. There are standard equations for these probabilities,

$$t_e = \frac{4\kappa\lambda e^{-\kappa d}}{(\lambda+\kappa)^2 e^{-\lambda d} - (\lambda-\kappa)^2 e^{\lambda d}},$$

$$d_e^{\rightarrow}(x') = \frac{2\kappa[(\lambda-\kappa)e^{\lambda(d-x')} + (\lambda+\kappa)e^{-\lambda(d-x')}]}{(\lambda+\kappa)^2 e^{-\lambda d} - (\lambda-\kappa)^2 e^{\lambda d}},$$

$$d_e^{\leftarrow}(x') = \frac{2\kappa e^{-\kappa d}[(\lambda-\kappa)e^{\lambda x'} + (\lambda+\kappa)e^{-\lambda x'}]}{(\lambda+\kappa)^2 e^{-\lambda d} - (\lambda-\kappa)^2 e^{\lambda d}}.$$

If the dephasing effect is ignored, the probabilities of reaching the localized state are related by the simple equations

$$d_e^{\rightarrow}(x')r_h - d_e^{\rightarrow *}(x') = -t_h d_e^{\leftarrow}(x'),$$

$$d_e^{\leftarrow}(x')r_h - d_e^{\leftarrow *}(x') = -t_h d_e^{\rightarrow}(x'),$$

where $r_{e,h}$ are the reflection coefficients of electrons and holes from the V_{rect} potential.

Conditions (4) and (5) introduced in the construction of our model help us to simplify calculations somewhat. As the localized state that effectively participates in resonance current transport is situated approximately in the middle of the barrier, $x_0 \sim d/2$, we can assume that the exponential functions $\exp(-\lambda d)$ and $\exp(-2\lambda x_0)$ have the same order of smallness. In addition, we assume that the barrier height is fairly large and condi-

tion (19) is therefore satisfied. The Green functions with an accuracy to terms on the order of $\exp(-2\lambda d)$ inclusive are then given by the equations

$$\begin{aligned}
G_{11}^0(x, x_0) \approx -\frac{m}{\lambda}\Bigg[& \exp(-\lambda|x - x_0|) \\
& - |t|^2 \exp(\lambda(x - x_0))\frac{\Gamma_3\Gamma_4 - \Gamma_1\Gamma_2}{\Gamma_1\Gamma_2} \\
& + \exp(-\lambda(x + x_0))\left(\frac{\lambda_0 - \kappa}{\lambda_0 + \kappa} + \frac{4\kappa\lambda_0}{\lambda_0^2 - \kappa^2}\frac{1}{\Gamma_1}\right) \\
& + \exp(-\lambda(2d - x - x_0))\left(\frac{\lambda_0 - \kappa}{\lambda_0 + \kappa} + \frac{4\kappa\lambda_0}{\lambda_0^2 - \kappa^2}\frac{1}{\Gamma_2}\right) \\
& + \exp(-\lambda(2d - x + x_0)) \\
& \times\left(2\left(\frac{\lambda_0 - \kappa}{\lambda_0 + \kappa}\right)^2 + \frac{8\kappa\lambda_0(\lambda_0^2 + \kappa^2)}{(\lambda_0^2 - \kappa^2)^2}\frac{(\Gamma_1 + \Gamma_2)}{\Gamma_1\Gamma_2}\right)\Bigg],
\end{aligned} \tag{43}$$

$$\begin{aligned}
G_{21}^0(x, x_0) \approx -\frac{m}{\lambda^*}\Bigg[& \exp(-\lambda(x + x_0))\frac{4\kappa\lambda_0}{\lambda_0^2 - \kappa^2}\frac{\beta_-^{\mathrm{L}}}{\Gamma_1} \\
& + \exp(-\lambda(2d - x - x_0))\frac{4\kappa\lambda_0}{\lambda_0^2 - \kappa^2}\frac{\beta_+^{\mathrm{R}}}{\Gamma_2} \\
& + \exp(-\lambda(2d - x + x_0)) \\
& \times\frac{8\kappa\lambda_0(\lambda_0^2 + \kappa^2)}{(\lambda_0^2 - \kappa^2)^2}\frac{(\beta_+^{\mathrm{R}}\Gamma_1 + \beta_-^{\mathrm{L}}\Gamma_2)}{\Gamma_1\Gamma_2}\Bigg],
\end{aligned} \tag{44}$$

$$\begin{aligned}
G_{12}^0(x, x_0) \approx \frac{m}{\lambda}\Bigg[& \exp(-\lambda(x + x_0))\frac{4\kappa\lambda_0}{\lambda_0^2 - \kappa^2}\frac{\beta_+^{\mathrm{L}*}}{\Gamma_1} \\
& + \exp(-\lambda(2d - x - x_0))\frac{4\kappa\lambda_0}{\lambda_0^2 - \kappa^2}\frac{\beta_-^{\mathrm{R}*}}{\Gamma_2} \\
& + \exp(-\lambda(2d - x + x_0)) \\
& \times\frac{8\kappa\lambda_0(\lambda_0^2 + \kappa^2)}{(\lambda_0^2 - \kappa^2)^2}\frac{(\beta_-^{\mathrm{R}*}\Gamma_1 + \beta_+^{\mathrm{L}*}\Gamma_2)}{\Gamma_1\Gamma_2}\Bigg],
\end{aligned} \tag{45}$$

$$\begin{aligned}
G_{22}^0(x, x_0) \approx \frac{m}{\lambda^*}\Bigg[& \exp(-\lambda|x - x_0|) \\
& - |t|^2 \exp(\lambda(x - x_0))\left(\frac{\Gamma_3\Gamma_4 - \Gamma_1\Gamma_2}{\Gamma_1\Gamma_2}\right) \\
& + \exp(-\lambda(x + x_0))\left(\frac{\lambda_0 + \kappa}{\lambda_0 - \kappa} - \frac{4\kappa\lambda_0}{\lambda_0^2 - \kappa^2}\frac{1}{\Gamma_1}\right) \\
& + \exp(-\lambda(2d - x - x_0))\left(\frac{\lambda_0 + \kappa}{\lambda_0 - \kappa} - \frac{4\kappa\lambda_0}{\lambda_0^2 - \kappa^2}\frac{1}{\Gamma_2}\right) \\
& + \exp(-\lambda(2d - x + x_0)) \\
& \times\left(2\left(\frac{\lambda_0 + \kappa}{\lambda_0 - \kappa}\right)^2 - \frac{8\kappa\lambda_0(\lambda_0^2 + \kappa^2)}{(\lambda_0^2 - \kappa^2)^2}\frac{(\Gamma_1 + \Gamma_2)}{\Gamma_1\Gamma_2}\right)\Bigg].
\end{aligned} \tag{46}$$

These equations are valid if the condition $\Gamma_1\Gamma_2 \neq 0$ is satisfied in the denominator Z of coefficients (42). We considered the situation with $\Gamma_1\Gamma_2 = 0$ separately; it was found that the contribution of such states to the resonance current was zero in equilibrium DID junctions.

Beenakker and van Houten [13] included dephasing in the derivation of the dispersion equation that related energy to the macroscopic phase and localized state parameters. Obtaining equally accurate results with Green functions (43)–(46) only requires taking dephasing into account in the Fermi momenta of interlayer quasi-particles that appear in the m/λ and m/λ^* multipliers. Since it is assumed that $\omega/V_0 \ll 1$, the Fermi momenta of electrons in a normal material can approximately be described by the equation

$$\begin{aligned}
\frac{1}{\lambda} &= \frac{1}{\sqrt{\lambda_0^2 + k_y^2 - 2mi\omega}} \\
&\approx \frac{1}{\lambda(k_y)}\left(1 + \frac{mi\omega}{\lambda^2(k_y)} + \frac{3}{2}\left(\frac{mi\omega}{\lambda^2(k_y)}\right)^2\right),
\end{aligned} \tag{47}$$

where

$$\lambda(k_y) = \sqrt{\lambda_0^2 + k_y^2}.$$

Accordingly, for holes, the equation complex conjugate to (47) can be used.

REFERENCES

1. M. Yu. Kupriyanov and J. S. Tsai, IEEE Trans. Appl. Supercond. **5**, 2531 (1995).
2. J. Yoshida, IEICE Trans. Electron. **83C**, 49 (2000).
3. Y. Tanaka and S. Kashiwaya, Phys. Rev. B **56**, 892 (1997).
4. Y. Tanaka and S. Kashiwaya, Phys. Rev. Lett. **74**, 3451 (1995).
5. Y. Tanaka and S. Kashiwaya, Phys. Rev. B **53**, R11957 (1996).
6. Yu. S. Barash, A. A. Svidzinsky, and H. Burkhardt, Phys. Rev. B **55**, 15282 (1997).

7. A. A. Golubov and M. Yu. Kupriyanov, Pis'ma Zh. Éksp. Teor. Fiz. **69**, 242 (1999) [JETP Lett. **69**, 262 (1999)].
8. A. I. Larkin and K. A. Matveev, Zh. Éksp. Teor. Fiz. **93**, 1030 (1987) [Sov. Phys. JETP **66**, 580 (1987)].
9. H. Knauer, J. Rihter, and P. Siedel, Phys. Status Solidi A **44**, 303 (1977).
10. I. M. Lifshits and V. Ya. Kirpichenkov, Zh. Éksp. Teor. Fiz. **77**, 989 (1979) [Sov. Phys. JETP **50**, 499 (1979)].
11. I. A. Devyatov and M. Yu. Kupriyanov, Pis'ma Zh. Éksp. Teor. Fiz. **39**, 187 (1994) [JETP Lett. **59**, 200 (1994)].
12. I. L. Aleiner, H. Clarke, and L. I. Glazman, Phys. Rev. B **53**, R7630 (1996).
13. C. W. J. Beenakker and H. van Houten, Phys. Rev. Lett. **66**, 3056 (1991).
14. I. A. Devyatov and M. Yu. Kupriyanov, Zh. Éksp. Teor. Fiz. **112**, 342 (1997) [JETP **85**, 189 (1997)].
15. L. G. Aslamazov and M. V. Fistul', Zh. Éksp. Teor. Fiz. **83**, 1170 (1982) [Sov. Phys. JETP **56**, 666 (1982)].
16. M. V. Fistul' and A. V. Tartakovskiĭ, Zh. Éksp. Teor. Fiz. **94** (9), 353 (1988) [Sov. Phys. JETP **67**, 1935 (1988)].
17. L. I. Glazman and K. A. Matveev, Pis'ma Zh. Éksp. Teor. Fiz. **49**, 570 (1989) [JETP Lett. **49**, 659 (1989)].
18. A. Golub, Phys. Rev. B **52**, 7458 (1995).
19. G. Johansson, E. N. Bratus, V. S. Shumeiko, *et al.*, Phys. Rev. B **60**, 1382 (1999).
20. I. A. Devyatov, D. V. Goncharov, and M. Yu. Kupriyanov, Zh. Éksp. Teor. Fiz. **119**, 749 (2001) [JETP **92**, 652 (2001)].
21. C.-R. Hu, Phys. Rev. Lett. **72**, 1526 (1994).
22. D. V. Goncharov, I. A. Devyatov, and M. Yu. Kupriyanov, Pis'ma Zh. Éksp. Teor. Fiz. **78**, 1126 (2003) [JETP Lett. **78**, 631 (2003)].
23. A. A. Abrikosov, L. P. Gor'kov, and I. E. Dzyaloshinskiĭ, *Methods of Quantum Field Theory in Statistical Physics* (Fizmatgiz, Moscow, 1962; Prentice Hall, Englewood Cliffs, N.J., 1963).
24. Chr. Bruder, Phys. Rev. B **41**, 4017 (1990).
25. G. Wendin and V. S. Shumeiko, Phys. Rev. B **53**, R6006 (1996).
26. S. Kashiwaya and Y. Tanaka, Rep. Prog. Phys. **63**, 1641 (2000).
27. A. Huck, A. van Otterlo, and M. Sigrist, Phys. Rev. B **56**, 14163 (1997).

Translated by V. Sipachev

Journal of Experimental and Theoretical Physics, Vol. 99, No. 5, 2004, pp. 1090–1107.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 5, 2004, pp. 1249–1266.

SOLIDS
Electronic Properties

Symmetry-Breaking Solutions of the Ginzburg–Landau Equation¶

Yu. N. Ovchinnikov[a,b] and I. M. Sigal[c,d]

[a] *Landau Institute, Moscow, Russia*
[b] *Max-Planck-Institute for Physics of Complex Systems, Dresden, Germany*
[c] *University of Toronto, Canada*
[d] *University of Notre Dame, USA*
e-mail: ovc@itp.ac.ru
Received March 1, 2004

Abstract—We consider the question of the existence of nonradial solutions of the Ginzburg–Landau equation. We present results indicating that such solutions exist. We seek such solutions as saddle points of the renormalized Ginzburg–Landau free-energy functional. There are two main points in our analysis: searching for solutions that have certain point symmetries and characterizing saddle-point solutions in terms of critical points of certain intervortex energy function. The latter critical points correspond to forceless vortex configurations.

1. INTRODUCTION

The Ginzburg–Landau equation describes, among other things, macroscopic stationary states of superfluids, Bose–Einstein condensation, and solitary waves in plasmas. In recent years, it has become a subject of active mathematical research (see monographs [1–3] and reviews [4–7] for some of the recent references). This equation is simple to write,

$$-\Delta\psi + (|\psi|^2 - 1)\psi = 0, \tag{1.1}$$

where (in the case of the entire plane $\mathbb{R}^2$) ψ: $\mathbb{R}^2 \longrightarrow \mathbb{C}$, with the boundary condition

$$|\psi| \longrightarrow 1 \text{ as } |x| \longrightarrow \infty, \tag{1.2}$$

but not easy to analyze. In fact, so far only radially symmetric solutions, i.e., solutions of the form $\psi_n(x) = f_n(r)e^{in\theta}$, where r and θ are polar coordinates for $x \in \mathbb{R}^2$, are known for (1.1) and (1.2) (see [8–17]). Solutions ψ_n are called *n vortices*. We note that $n = \deg\psi_n$, where $\deg\psi$, the degree (or vorticity) of ψ (satisfying (1.2)) is the total index (winding number) at ∞ of ψ considered as a vector field on $\mathbb{R}^2$, i.e.,

$$\deg\psi := \frac{1}{2\pi}\int\limits_{|x|=R} d(\arg\psi)$$

for sufficiently large R.

The existence and properties of the vortex solutions were established only recently. The known facts are as follows.

(i) Existence and uniqueness (modulo symmetry transformations and in a class of radially symmetric functions) [10–13].

(ii) Stability for $|n| \le 1$ and instability for $|n| > 1$ ([13], earlier results on stability for the disc are due to [15–17]).

(iii) Uniqueness of $\psi_{\pm 1}$ (again, modulo symmetry transformation) in a class of functions ψ with $\deg\psi = \pm 1$ and $\int(|\psi|^2 - 1)^2 < \infty$ [16].

Therefore, the next question is: Are there nonradially symmetric solutions?

In this paper, we present results indicating that such solutions exist. There are two key ingredients in our analysis. First, we characterize nonradially symmetric solutions as critical points of the intervortex energy function described below (see also [18]). Second, we seek solutions having certain point symmetries. The latter fact reduces the number of free parameters describing such solutions to one (the size of the corresponding polygon of vortices).

Solutions breaking the rotational symmetry were found to exist in the case of the Ginzburg–Landau equation in the ball $B_R = \{x \in \mathbb{R}^2 \mid |x| \le R\}$ with the boundary condition $\psi|_{\partial B_R} = e^{in\theta}$ and $|n| \ge 2$ (see [1, 2], Theorem IX.1). However, in the case of the ball, there is an external mechanism leading to the symmetry breaking: the boundary condition. It repels vortices, forcing their confinement. On the other hand, the energy is lowered by breaking up multiple vortices into (+1)- (or (–1)-) vortices and merging vortices of opposite signs. Thus, for R that are not very small, the lowest energy is reached by a configuration of $|n|$ vortices of vorticities

¶ This article was submitted by authors in English.

±1 depending on the sign of n, which, obviously, is not rotationally symmetric.

This paper is organized as follows. In Sections 2 and 3, we review some material in [13]: the variational formulation of the problem and some specific properties of vortex solutions. In Section 4, we define the intervortex energy and discuss its properties. In particular, we discuss the correlation term in (the upper bound on) the expansion of the intervortex energy for large intervortex separations and a definition of G-symmetric vortex energies, where G is a subgroup of the symmetry group of (1.1).

In Section 5, we consider point symmetries (C_{Nv}), present one of our main results, Theorem 5.1, on the existence of critical points for C_{Nv}-symmetric intervortex energies, and derive some general relations for those energies. In Section 6, we prove Theorem 5.1 and discuss some other cases.

Finally, in our five appendices all the hard analytic and numerical work is concentrated. In these appendices, we compute various asymptotic expansions beyond the leading order. We feel that these appendices are of interest on their own because they address rather subtle computational issues.

2. RENORMALIZED GINZBURG–LANDAU ENERGY

It is a straightforward observation that Eq. (1.1) is the equation for critical points of the functional

$$\mathscr{E}(\psi) = \frac{1}{2}\int\left(|\nabla\psi|^2 + \frac{1}{2}(|\psi|^2 - 1)^2\right). \tag{2.1}$$

Indeed, if we define the variational derivative $\partial_\psi\mathscr{E}(\psi)$ of $\mathscr{E}$ by

$$\mathrm{Re}\int\xi\partial_\psi\mathscr{E}(\psi) = \frac{\partial}{\partial\lambda}\mathscr{E}(\psi_\lambda)\Big|_{\lambda=0} \tag{2.2}$$

for any path ψ_λ such that $\psi_0 = \psi$ and $\frac{\partial}{\partial\lambda}\psi_\lambda\Big|_{\lambda=0} = \xi$, then the left-hand side of Eq. (1.1) is equal to $\overline{\partial_\psi\mathscr{E}(\psi)} = \partial_{\bar\psi}\mathscr{E}(\psi)$ for $\mathscr{E}(\psi)$ given by (2.1).

Equation (2.1) is the celebrated Ginzburg–Landau (free) energy. However, there is a problem with it in our context. It is shown in [13] that, if ψ is an arbitrary C^1-vector field on $\mathbb{R}^2$ such that $|\psi| \longrightarrow 1$ as $|x| \longrightarrow \infty$ uniformly in $\hat{x} = x/|x|$ and $\deg\psi \neq 0$, then $\mathscr{E}(\psi) = \infty$.

We renormalize the Ginzburg–Landau energy functional as follows (see [13]). Let $\chi(x)$ be a smooth positive function on $\mathbb{R}^2$ vanishing at the origin and converging to one at infinity. We define

$$\mathscr{E}_{\mathrm{ren}}(\psi) = \frac{1}{2}\int\left(|\nabla\psi|^2 - \frac{(\deg\psi)^2}{r^2}\chi + F(|\psi|^2)\right)d^2x, \tag{2.3}$$

where

$$F(u) = \frac{1}{2}(u-1)^2. \tag{2.4}$$

Properties of the renormalized energy functional $\mathscr{E}_{\mathrm{ren}}(\psi)$ are investigated in [13].

In this paper, we take

$$\chi(x) = \begin{cases} 1 & \text{for } |x| \geq R + R^{-1}, \\ 0 & \text{for } |x| \leq R \end{cases} \tag{2.5}$$

for R very large compared to all length scales appearing below.

3. VORTICES

It is shown in [10–13] that, for any n, Eq. (1.1) has a solution, unique modulo symmetry transformations, of the form

$$\psi_n(x) = f_n(r)e^{in\theta}, \tag{3.1}$$

where f_n, with $1 > f_n \geq 0$, monotonically increase from $f_n(0) = 0$ to 1 as r increases to ∞. For $n = 0$, $f_n(r) = 1$. For $|n| > 0$, $f_n(r)$ does not admit an explicit expression. These are the n vortices mentioned in the Introduction. Of course, each solution ψ_n generates a one-parameter (for $n = 0$) or a three-parameter (for $|n| > 0$) family of solutions of (1.1). The latter are obtained by applying symmetry transformations to ψ_n.

The function $f_n(r)$ in (3.1) satisfies the ordinary differential equation

$$-\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial f_n}{\partial r}\right) + \frac{n^2}{r^2}f_n - (1 - f_n^2)f_n = 0. \tag{3.2}$$

The (self) energy of the n vortex is given by $E_{n,R} := \mathscr{E}_{\mathrm{ren}}(\psi_n)$. To compute $E_{n,R}$, we use the fact that, if ψ is a solution of (1.1), then, due to the formula $\int|\nabla\psi|^2 = -\int\bar\psi\Delta\psi$ of integration by parts, we have

$$\mathscr{E}_{\mathrm{ren}}(\psi) = \frac{1}{2}\int\left(1 - |\psi|^2 - \frac{1}{2}(1-|\psi|^2)^2 - \frac{(\deg\psi)^2}{r^2}\chi\right). \tag{3.3}$$

Using this formula for $\psi = \psi_n$ and using the asymptotic expression (which can be easily derived from (3.2); see [19, 20] for the general case)

$$f_n(r) = 1 - \frac{n^2}{2r^2} + O\left(\frac{1}{r^4}\right) \tag{3.4}$$

for $r \gg 1$, we obtain

$$E_{n,R} = \pi n^2 \ln\left(\frac{R}{|n|}\right) + c(|n|) + O\left(\frac{1}{R^2}\right). \tag{3.5}$$

The constant $c(n)$ can be computed numerically (which is not quite trivial; see Appendix 1), which yields

$$\begin{aligned} c(1) &= 0.376\pi, \quad c(2) = 0.535\pi, \\ c(3) &= 0.577\pi, \quad c(5) = 0.615\pi. \end{aligned} \tag{3.6}$$

The asymptotic form of $c(n)$ for $n \gg 1$ is found analytically in Appendix 2.

4. INTERVORTEX ENERGY

In this section, we introduce and discuss a key concept of the intervortex energy (see also [4, 18]). We begin with some definitions.

By a vortex configuration $\underline{c}$, we understand a pair $(\underline{a}, \underline{n})$, where $\underline{a} = (a_1, \ldots, a_K)$, $a_j \in \mathbb{R}^2$, and $\underline{n} = (n_1, \ldots, n_K)$, $n_j \in \mathbb{Z}$, for some $K \geq 1$ (positions of the vortex centers and their vorticities). We consider once-differentiable functions ψ: $\mathbb{R}^2 \longrightarrow \mathbb{C}$ satisfying $|\psi| \longrightarrow 1$ as $|x| \longrightarrow \infty$. We say that the vortex configuration of ψ is $\underline{c} = (\underline{a}, \underline{n})$, $\mathrm{conf}\psi = \underline{c}$, if ψ has zeros (only) at $a_1, \ldots, a_K$ with the respective local indices $n_1, \ldots, n_K$; i.e.,

$$\int_{\gamma_j} d(\arg\psi) = 2\pi n_j \tag{4.1}$$

for any contour γ_j containing a_j, but not the other zeros of ψ, and for $j = 1, \ldots, K$. (Strictly speaking, we have to specify the phase factor, or rotation angle, for each vortex; but these play no role in our considerations and are not displayed or mentioned in what follows.) We now define

$$E_R(\underline{c}) = \inf\{\mathscr{E}_{\mathrm{ren}}(\psi) \mid \mathrm{conf}\psi = \underline{c}\}. \tag{4.2}$$

We expect that $E_R(\underline{c}) > -\infty$. An argument supporting this statement is presented in [18]. Of course, for bounded domains, this inequality is trivial. We call $E_R(\underline{c})$ the energy of the vortex configuration $\underline{c}$. It plays a central role in our analysis. We also note that $E(\underline{c})$ serves as a Hamiltonian for the vortex dynamics in the adiabatic approximation (see [21]).

In what follows, we keep the vortex indices $\underline{n}$ fixed and write $E_R(\underline{a})$ for $E_R(\underline{c})$. It is clear intuitively that a minimizer in (4.2) exists if and only if $\nabla E_R(\underline{a}) = 0$ (the force acting on the vortex centers is zero). However, to establish this fact is not so easy.

Theorem 4.1. *If there is a minimizer for variational problem* (4.2), then *this minimizer satisfies Ginzburg–Landau equation* (1.1).

Proof. Let ψ be a minimizer for (4.2). Because we have

$$\begin{aligned} 0 &= \frac{\partial}{\partial\lambda}\mathscr{E}_{\mathrm{ren}}(\psi + \lambda\xi)\Big|_{\lambda = 0} \\ &= \mathrm{Re}\int \bar{\xi}(-\Delta\psi + (|\psi|^2 - 1)\psi) \end{aligned}$$

for any differentiable function ξ: $\mathbb{R}^2 \longrightarrow \mathbb{C}$ vanishing together with its gradient sufficiently fast at ∞ and vanishing at the points $a_1, \ldots, a_m$, we conclude that ψ satisfies (1.1) for $x \neq a_1, \ldots, a_m$. On the other hand, because $\psi \in H_1^{\mathrm{loc}}(\mathbb{R}^2)$, we have $-\Delta\psi + (|\psi|^2 - 1)\psi \in H_{-1}^{\mathrm{loc}}(\mathbb{R}^2)$. Hence, $-\Delta\psi + (|\psi|^2 - 1)\psi = 0$ on $\mathbb{R}^2$.

Arguments and results in [18] (see, in particular, Theorem 3.2) justify the following conjecture.

Conjecture 4.2. $\nabla E_R(\underline{a}_0) = 0$ for some $\underline{a}_0$ (with $\underline{n}$ fixed) if and only if there is a minimizer for problem (4.2) at the configuration $\underline{a}_0$ and, consequently, due to Theorem 4.1, if and only if Ginzburg–Landau equation (1.1) has a solution with the configuration $\underline{a}_0$.

The goal of this paper is to find forceless vortex configurations, i.e., configurations $\underline{c}$ such that

$$\nabla E_R(\underline{a}) = 0. \tag{4.3}$$

For this, we study the intervortex energy $E_R(\underline{a})$ for very small and very large intervortex separations.

Let

$$d_{\underline{a}} = \min_{i \neq j}|a_i - a_j| \quad \text{for} \quad \underline{a} = (a_1, \ldots, a_K).$$

For $d_{\underline{a}}$ large, we prove in Section 7 the upper bound

$$E_R(\underline{a}) \leq E_R^{(0)} - A(\underline{a}) + O(d_a^{-8/3}) + O(R^{-2}), \tag{4.4}$$

where

$$E_R^{(0)} = \sum_{i=1}^{K} E_{n_i,R} + H\left(\frac{\underline{a}}{R}\right),$$

and $A(\underline{c})$ is a homogeneous function of degree –2, provided that $\underline{a}$ satisfies $\nabla H(\underline{a}) = 0$. We recall that $E_{n,R} = \mathscr{E}_{\mathrm{ren}}(\psi_n)$ is the self-energy of the n vortex (see (3.5)) and $H(\underline{a})$ is the energy of the vortex pair interactions,

$$H(\underline{a}) = -\pi\sum_{i \neq j} n_i n_j \ln|a_{ij}|, \tag{4.5}$$

with $a_{ij} = a_i - a_j$.

The correlation term $A(\underline{a})$is of importance for us here. We have an explicit expression for it (see Eqs. (A.3.4) and (A.3.5)) and compute it explicitly in the cases of interest. We conjecture that $A(\underline{a}) > 0$ always.

We observe that the upper bound (4.4) with the remainder $O(d_{\underline{a}}^{-1})$ instead of $-A(\underline{a}) + O(d_{\underline{a}}^{-8/3})$ is obtained by choosing the Hartree-type function

$$\psi^{(0)}(x) = \prod_{i=1}^{K} \psi_{n_j}(x - a_i)$$

describing "independent" vortices. For asymptotically forceless configurations, i.e., ones with $\nabla H(\underline{a}) = 0$, this estimate can be somewhat improved, but in order to move even to the remainder estimate $O(d_{\underline{a}}^{-2} \ln d_{\underline{a}})$ in the latter case, one has to refine this function and include the leading correlations.

Remark 4.3. As $d_{\underline{a}} \longrightarrow \infty$, the important asymptotic expression

$$E_R(\underline{a}) = \sum_{i=1}^{K} E_{n_i, R} + H\left(\frac{\underline{a}}{R}\right) + \mathrm{Rem} \qquad (4.6)$$

was proved in [18] with $\mathrm{Rem} = O(d_{\underline{a}}^{-2} \ln d_{\underline{a}})$ in general and $\mathrm{Rem} = O(d_{\underline{a}}^{-2})$ if $\nabla H(\underline{a}) = 0$.

As mentioned in the Introduction, our second idea is to consider solutions of (1.1) that are invariant under point group transformations. Consequently, we introduce intervortex energy functions invariant under such groups. We consider a subgroup G of the total symmetry group

$$G_{\mathrm{sym}} = O(2) \times T(2) \times U(1)$$

(where $T(n)$ is the group of translations of $\mathbb{R}^n$) of Ginzburg–Landau equation (1). For a G-invariant vortex configuration $\underline{c} = (\underline{a}, \underline{n})$ (i.e., invariant under the spatial part of G), we define the G-invariant vortex interaction energy $E_{R,G}(\underline{a})$ as

$$E_{R,G}(\underline{a}) = \inf\{\mathscr{E}_{\mathrm{ren}}(\psi) \mid \mathrm{conf}\,\psi = \bar{c},\ \psi \ \text{is}\ G\text{-invariant}\}$$

(as before, we fix $\underline{n}$ and omit it from the relation).

Theorem 4.1 and Conjecture 4.2 obviously extend to the G-symmetric situation. In particular, we have the following conjecture:

If $\underline{a}_0$ is a critical point of $E_{R,G}(\underline{a})$ (i.e., $\nabla E_{R,G}(\underline{a}_0) = 0$), then Eq. (1.1) has a G-invariant solution.

Our goal in what follows is to find critical points of the G-invariant intervortex energy $E_{R,G}(\underline{a})$ for appropriate groups G, namely, point groups C_{Nv} (see the next section).

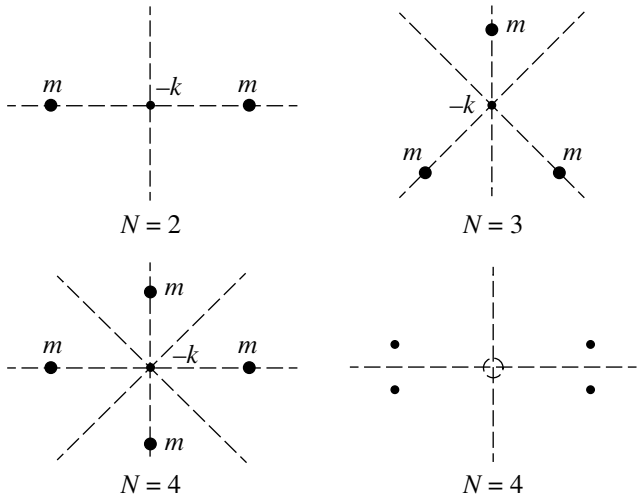


Fig. 1. Symmetric configurations and their reflection lines.

5. POINT SYMMETRIES

We seek solutions of Eq. (1.1) having symmetry groups C_{Nv}. These groups consist of rotations around the origin by angles given by integer multiples of $2\pi/N$ and reflection(s) in one (and therefore N) line(s) passing through the origin. Such solutions are determined by fixing vortex configurations that have the desired symmetry group. We consider vortex configurations consisting of N m vortices uniformly spaced on a circle of radius a and a single $(-k)$ vortex at the center of the circle, which is placed at the origin. Several such configurations and their symmetry lines are shown in Fig. 1. Such configurations have the symmetry group C_{Nv}. The symmetry group C_{Nv} determines such a configuration uniquely up to the vortex values m and k and the size a.

As noted at the end of the previous section, we rely on the argument that C_{Nv}-symmetric solutions are in one-to-one correspondence with critical points of the C_{Nv}-symmetric intervortex energy

$$E_R(\underline{c}) \equiv E_{R,C_{Nv}}(\underline{c})$$

(hereinafter, we consider only C_{Nv}-symmetric intervortex energies, and often omit the subscript C_{Nv}). Our goal is to find critical points of $E_R(\underline{c})$. One of the central results in this paper is the following theorem.

Theorem 5.1. *There exist critical points of $E_{R,C_{Nv}}(\underline{c})$ among the configurations $\underline{c}$ described above for the parameter values*

$$(N, m, k) = (2, 2, 1) \ \text{and}\ (4, 2, 3).$$

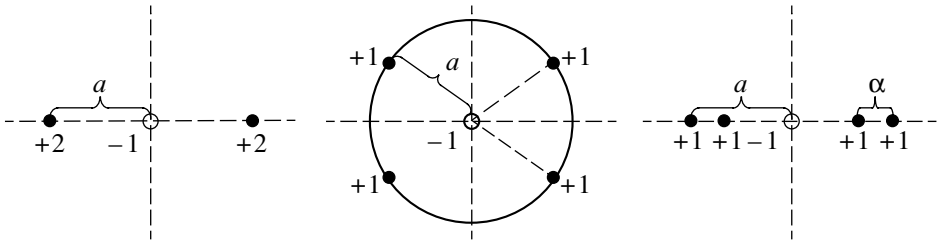


Fig. 2.

This theorem is proved in Section 6. In the rest of this section, we establish general properties of the energy $E_{R, C_{Nv}}(\underline{c})$ and find a necessary condition on the parameters N, m, and k.

We observe that, if $\underline{c}$ is a configuration described above, then

$$\begin{aligned} \nabla_{a_j} E_R(\underline{a}) &= \hat{a}_j \partial_{|a_j|} E_R(\underline{a}) \\ \text{and } \nabla_{a_j} H(\underline{a}) &= \hat{a}_j \partial_{|a_j|} H(\underline{a}) \quad \forall j, \end{aligned} \tag{5.1}$$

where $\hat{a} = a/|a|$ (again, we do not display the parameters $\underline{n}$). In this case, it therefore suffices to investigate the energy $E_R(\underline{a})$ as a function of one variable, the scale parameter a.

We note that, if $m \geq 2$, then there is a continuum of configurations, labeled by a parameter $\alpha > 0$, with the same symmetry group C_{Nv} as a given configuration, which have the given configuration as the limit as $\alpha \longrightarrow 0$. For instance, for $m = 2$, each m-vortex can be split into a pair of 1-vortices with all pairs lying either on the circle or on the lines joining their parent m-vortices to the origin at equal distance α to those m-vortices (see Fig. 2).

By symmetry, the energy of the resulting configurations has a critical point at $\alpha = 0$. A simple analysis of the breakup of a 2-vortex shows that this critical point is a local maximum. Indeed, e.g., for $m = 2$, it was shown in [13] that the linearization of Eq. (1.1) (the Hessian of the energy functional) around the 2-vortex solutions $\psi_2 = f_2(r)e^{2i\theta}$ has exactly one negative mode (an eigenfunction corresponding to a negative eigenvalue) of the form $\xi = e^{4i\varphi}\xi_4(r) + \xi_0(r)$, where $\xi_k(r)$ are some real functions. Then, the function $\psi_2 + \lambda\xi$ for sufficiently small $|\lambda|$ lowers the energy of ψ_2. On the other hand, this function has two simple zeros (i.e., of vorticities +1) in a vicinity of $x = 0$. Indeed, in the complex notation $z = x_1 + ix_2 \longleftrightarrow x = (x_1, x_2)$, $\psi_2(z) = bz^2 + O(z^3)$ and $\xi(z) = c + O(z)$ for some positive numbers b and c in the neighborhood of $z = 0$. Hence, $\psi_2(z) + \lambda\xi(z) = bz^2 + \lambda c + O(z^3) + O(\lambda z)$, which therefore has two simple zeros $z_\pm = \pm\sqrt{\frac{\lambda c}{b}} + O(\lambda^{3/4})$ in the neighborhood of $z = 0$. This shows in particular that splitting of a 2-vortex lowers the energy.

Proposition 5.2. *Let a configuration $\underline{c}_0$, as described above, be asymptotically forceless, i.e., $\nabla H(\underline{a}_0) = 0$. Then,*

$$k = \frac{1}{2}(N-1)m. \tag{5.2}$$

Proof. By virtue of (4.1), the equation $\nabla H(\underline{a}_0) = 0$ for the configuration described is equivalent to the equation

$$\frac{\partial}{\partial a} H(\underline{a}_0) = 0. \tag{5.3}$$

Because

$$H(\underline{a}) = H\left(\frac{\underline{a}}{a}\right) - \pi \sum_{i \neq j} n_i n_j \ln a, \tag{5.4}$$

the latter equation implies that $\sum_{i \neq j} n_i n_j = 0$, which is equivalent to (5.2) due to the relation

$$\sum_{i \neq j} n_i n_j = -2Nmk + N(N-1)m^2. \tag{5.5}$$

We note that Eq. (5.3) implies that, if $\nabla H(\underline{a}_0) = 0$, then $\nabla H(\underline{a}) = 0$ for all $\underline{a}$ of forms $\underline{a} = s\underline{a}_0$, $s > 0$. The latter fact implies another proof of (5.2). Indeed, $H(\underline{a}/R)$ behaves as const $\cdot \ln R$ + const for large R. Hence, for an asymptotically force-free configuration (i.e., the one with $\nabla H(\underline{a}) = 0$), the constant in front of $\ln R$ is independent of the scale parameter a. This asymptotic scale invariance implies that the leading term

$$\pi(Nm - k)^2 \ln R$$

for the configuration with $a = 0$ (i.e., when all the vortices collapse to the center of the circle) is equal to the leading term

$$\pi(Nm^2 + k^2) \ln R$$

for the configuration with a very large a, and therefore the vortices in such a configuration can be treated as

virtually independent (see (4.4)). Hence,

$$(Nm-k)^2 = Nm^2 + k^2,$$

which implies (5.2).

We observe that Eq. (5.2) is equivalent to the relation

$$H\left(\frac{\underline{a}}{R}\right) = H\left(\frac{\underline{a}}{a}\right) = H(\underline{a}), \quad \text{independent of } a. \quad (5.6)$$

Indeed, this follows from Eqs. (5.4) and (5.5).

Relation (5.2) between k and m is assumed in what follows.

For the configuration above, we now introduce the energy differences

$$\Delta E(\underline{a}) := E_R(\underline{a}) - \pi(Nm-k)^2 \ln R, \quad (5.7)$$

where we recall that $Nm - k$ is the total vorticity of the configuration in question and $E_{n,R}$ is the energy of a single vortex of vorticity n, i.e., $E_{n,R} = \mathscr{E}_{\mathrm{ren}}(\psi_n)$. We let ΔE_n denote the energy difference for this vortex,

$$E_{n,R} = \pi n^2 \ln R + \Delta E_n. \quad (5.8)$$

Clearly,

$$E_R(\underline{0}) = E_{Nm-k,R} \text{ and } \Delta E(\underline{0}) = \Delta E_{Nm-k}. \quad (5.9)$$

This, together with (3.5), implies that (modulo $O(R^{-2})$)

$$\Delta E(\underline{0}) = -\pi(Nm-k)^2 \ln(Nm-k) + c(Nm-k). \quad (5.10)$$

On the other hand, for very large intervortex distances, Eqs. (5.7), (5.6), (4.6), and (3.5) imply that (modulo $O(R^{-2}) + o(a^{-2})$)

$$\begin{aligned}\Delta E(\underline{a}) \le &-\pi(Nm^2 \ln m + k^2 \ln k)\\ &+ Nc(m) + c(k) + H(\underline{a}) - Ca^{-2},\end{aligned} \quad (5.11)$$

where $C = A(\underline{a}/a)$. We compute $H(\underline{a})$ for the given configuration. Because the distances between the vortices on the circle are $2a\sin\frac{\pi}{N}, 2a\sin\frac{\pi}{N}, \ldots, 2a\sin\frac{(N-1)\pi}{N}$, we find

$$H(\underline{a}) = -\pi m^2 N \sum_{k=1}^{N-1} \ln\left(2\sin\frac{k\pi}{N}\right). \quad (5.12)$$

This equation, together with Eq. (5.11), yields, for large intervortex distances,

$$\begin{aligned}\Delta E(\underline{a}) \le &-\pi(Nm^2 \ln m + k^2 \ln k) + Nc(m)\\ &+ c(k) - \pi m^2 N \sum_{k=1}^{N-1} \ln\left(2\sin\frac{k\pi}{N}\right) - Ca^{-2}\end{aligned} \quad (5.13)$$

modulo $O(R^{-2}) + o(a^{-2})$.

In the next section, we establish the existence of points $\underline{a}_0$ such that $\nabla E(\underline{a}_0) = 0$ for given configurations by comparing $\Delta E(\underline{0})$ and $\Delta E(\underline{a})$ for large intervortex distances a.

6. SIMPLEST CASES: PROOF OF THEOREM 5.1

In this section, we consider some special, in fact the simplest, cases of the vortex configurations introduced in Section 5. We recall that every such configuration consists of a vortex of vorticity $-k$ placed at the origin and N vortices, each of vorticity m, distributed equidistantly on a circle of radius a with the center at the origin. Such a configuration is fixed by the symmetry group C_{Nv}, and hence the only remaining free parameter is the radius of the circle a. With a slight abuse of notation, we write $\Delta E(a) = \Delta E(\underline{a})$.

Proof of Theorem 5.1. The correlation coefficient C in Eq. (5.13) is computed for the specified configurations in Appendix 3:

$$\begin{aligned}C &= 8\pi, 20\pi\\ \text{for } &(N, m, k) = (2, 2, 1), (4, 2, 3).\end{aligned} \quad (6.1)$$

(We expect that, for general (N, m, k), $k = \frac{1}{2}(N-1)m$, C is of the form $\frac{\pi}{4} \cdot$ (integer).) Thus, $\Delta E(a)$ monotonically increases to

$$\begin{aligned}&\Delta E(a) \text{ monotonically increases to}\\ &\Delta E(\infty) \text{ as } a \longrightarrow \infty.\end{aligned} \quad (6.2)$$

Moreover, due to (3.6), we have

$$\Delta E(\infty) < \Delta E(0) \quad (6.3)$$

for the configurations $(N, m, k) = (2, 2, 1), (4, 2, 3)$ (explicit computations are given below). Hence, $\Delta E(a)$ has at least one minimum for these configurations as claimed.

Computation of (6.3)

(a) The case $N = 2$, $m = 2$, and $k = 1$ (we recall that $E_R(a) \equiv E_R(\underline{a})$, etc.). We have

$$\Delta E(0) \equiv \Delta E_3(0) = c(3) - 9\pi \ln 3 = -9.31\pi. \quad (6.4)$$

On the other hand, Eq. (5.11) implies that, for very large a,

$$\begin{aligned}\Delta E(a) \le &\, c(1) + (2c(2) - 8\pi \ln 2) - 8\pi \ln 2 - Ca^{-2}\\ &+ O\left(\frac{\ln a}{a^4}\right) = -9.64\pi - Ca^{-2} + O\left(\frac{\ln a}{a^4}\right).\end{aligned} \quad (6.5)$$

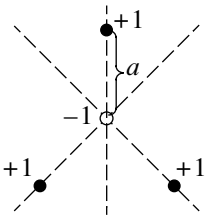


Fig. 3. (See Fig. 1; a critical value of the parameter a is not specified, but its existence is established).

(b) The case $N = 4$, $m = 2$, and $k = 3$ (see Fig. 1). In this case,

$$\Delta E(0) = \Delta E_5(0) = c(5) - 25\pi\ln 5 = -39.62\pi. \quad (6.6)$$

On the other hand, Eq. (5.11) implies that, for large a, we have the asymptotic behavior

$$\begin{aligned}\Delta E(a) &\le (4c(2) - 16\pi\ln 2) + (c(3) - 9\pi\ln 3)\\ &\quad - 32\pi\ln 2 - Ca^{-2} + O\left(\frac{\ln a}{a^4}\right)\\ &= -40.44\pi - Ca^{-2} + O\left(\frac{\ln a}{a^4}\right).\end{aligned} \quad (6.7)$$

Thus, (6.3) is shown.

Remark 1. We examine the case where $m = 1$, i.e., the vortices on the circle are simple. In this case, $k = (N - 1)/2$. Therefore, in the simplest case where $N = 3$ and $k = 1$, we take the ($m = 1$)-vortices as equally spaced (Fig. 3).

Equations (4.9), (4.12), and (3.6) imply that, in this case, $\Delta E(0) < \Delta E(\infty)$ (in fact, $\Delta E(0) = \Delta E_2(0) = -2.238\pi$ and $\Delta E(\infty) = -1.792\pi$). Numerical computations show (see Appendices 3 and 4) that $\Delta E'(\infty) > 0$ and $\Delta E'(0) > 0$ (in fact, for $a \gg 1$, $\Delta E(a) = 4c(1) - 3\pi\ln 3 - Ca^{-2} = -1.792\pi - Ca^{-2}$ with $C > 0$). In this case, we cannot therefore conclude that a critical point of $E_R(a)$ exists. But a more careful numerical analysis indicates that there probably exist two extremal points of $E_R(a)$, a minimum and a maximum, for $1/\sqrt{2} \le a \le 2$. Similar configurations for large (and odd) N are analyzed in Appendix 5.

Remark 2. The case where $N = 2$, $m = 2$, and $k = 1$ is the limiting case of $N = 4$, $m = 1$, and $k = 1$ (see Fig. 2). All three configurations have the same symmetry group C_{2v} generated by rotation by π and reflections in the vertical and horizontal axes passing through the vortex –1. After the symmetry group is fixed, the second and third configurations have two free parameters: the scale parameter a and the angle/distance α between two of its neighboring 1-vortices (see Fig. 3). As $\alpha \longrightarrow 0$, the second and third configurations are continuously transformed into the first one.

7. UPPER BOUND ON THE INTERVORTEX ENERGY

In this section, we prove inequality (4.4) for the energy $E_R(\underline{a})$ of vortex configurations.

Theorem 7.1. *We have the estimate*

$$E_R(\underline{a}) \le E_R^{(0)} + \mathrm{Rem} + O(\max|a_j|^2/R^2), \quad (7.1)$$

where $E_R^{(0)} = \sum_{k=1}^{k} E_{n_{i,R}} + H\left(\frac{\underline{a}}{R}\right)$ *and*

$$\mathrm{Rem} = \begin{cases} O(d_{\underline{a}}^{-2}) & \text{if } \nabla H(\underline{a}) = 0,\\ O(d_{\underline{a}}^{-2}\ln d_{\underline{a}}) & \text{otherwise.}\end{cases} \quad (7.2)$$

Moreover, if $\nabla H(\underline{a}) = 0$, then estimate (7.2) can be improved as

$$\mathrm{Rem} = -A(\underline{a}) + O(d_{\underline{a}}^{-8/3}) + O\left(\frac{1}{R^2}\right), \quad (7.3)$$

where $A(\underline{a})$, the correlation term, is a homogeneous degree-(–2) function explicitly given by the conditionally convergent integral

$$A(\underline{a}) = \frac{1}{4}\int\left[|\nabla\varphi_0|^4 - \sum_j |\nabla\varphi_j|^4\right] \quad (7.4)$$

(where $\nabla H(\underline{a}) = 0$ is assumed) with

$$\varphi_0 = \sum_j \varphi_j, \quad \varphi_j(x) = n_j\theta(x - a_j), \quad (7.5)$$

$\theta(x)$ is the polar angle of $x \in \mathbb{R}^2$.

Before proceeding to the proof of these estimates, we show that the integral in the right-hand side of (7.4) is conditionally convergent in the forceless case $\nabla H(\underline{a}) = 0$. Because the integrand has singularities at the points $a_1, \ldots, a_K$, it suffices to show that the integrals over the discs $D(a_k, \varepsilon)$ centered at a_k and of a radius $\varepsilon > 0$ converge. We consider the integral over the disc $D(a_k, \varepsilon)$. Let

$$\varphi_{(k)}(x) = \sum_{j \ne k} \varphi_j(x). \quad (7.6)$$

Because the function $\varphi_{(k)}(x)$ is harmonic in $D(a_k, \varepsilon)$, it has an expansion around the point a_k of the form

$$\varphi_{(k)}(x) = \sum_{m=0}^{\infty} c_m r_k^m \cos m(\theta_k - \theta^{(m)}), \quad (7.7)$$

where r_k and θ_k are the polar coordinates of $x_k = x - a_k$ and c_m and $\theta^{(m)}$ are some constants.

In the forceless case,

$$\nabla\varphi_{(k)}(a_k) = -\frac{1}{2\pi n_k} J\nabla_{a_k} H(\underline{a}) = 0, \tag{7.8}$$

and, therefore,

$$\nabla\varphi_{(k)}(x) = c_k(x_k\cos 2\theta_k - x_k^{\perp}\sin 2\theta_k) + O\left(\frac{r_k^2}{d_{\underline{a}}^3}\right), \tag{7.9}$$

where $ck = O(1/d_{\underline{a}}^2)$ is a constant, $r_k = |x_k|$, and $x^{\perp} = (-x_2, x_1)$. Now, writing

$$\int\limits_{D(a_k,\varepsilon)} (|\nabla\varphi|^4 - |\nabla\varphi_k|^4) = \int\limits_{D(a_k,\varepsilon)} (2|\nabla\varphi_k|^2\alpha_k + \alpha_k^2), \tag{7.10}$$

where

$$\alpha_k := 2\nabla\varphi_k\cdot\nabla\varphi_{(k)} + |\nabla\varphi_{(k)}|^2, \tag{7.11}$$

and using (7.9), we see that the singular part of the integral above is

$$\begin{aligned} &4\int\limits_{D(a_k,\varepsilon)} |\nabla\varphi_k|^2\nabla\varphi_k\cdot\nabla\varphi_{(k)} \\ &= 4\int\limits_{D(a_k,\varepsilon)} \frac{n_k^2}{r_k^2}(-c_k\sin 2\theta_k + O(r_k)) = \int\limits_{D(a_k,\varepsilon)} O\left(\frac{1}{r_k}\right) < \infty. \end{aligned} \tag{7.12}$$

Therefore, the integral in the right-hand side of (7.4) is conditionally convergent, in the sense that it is well defined as a limit of similar integrals with small discs around the points $a_1, \ldots, a_k$ excised, as the radii of those discs tend to 0.

Proof of Theorem 7.1. We prove the upper bound (7.1) using the variational inequality

$$E_R(\underline{a}) \le \mathscr{E}_R(\psi), \tag{7.13}$$

valid for any function ψ having the given vortex configuration $\underline{a}$, and by showing that, for an appropriate ψ, $\mathscr{E}_R(\psi)$ is of the form of the right-hand side of (7.6). Namely, we show that

$$\mathscr{E}_{\text{ren}}(\psi) = E_R^{(0)} + \text{Rem}, \tag{7.14}$$

where Rem is given by either (7.2) or (7.3), as appropriate. Then, (7.1) follows from (7.13) and (7.14).

We begin with proving estimate (7.1) with remainder (7.2). Let $\psi_i(x) = \psi^{(n_i)}(x_i)$, where $x_i = x - a_i$, and let $f_i \equiv |\psi_i|$. We consider the class of functions ψ of the form $\psi = fe^{i\varphi_0}$ with a function f such that

$$f = f_i + O\left(\frac{1}{rd_{\underline{a}}^n}\right) \text{ if } r_j \ll d_{\underline{a}}, \quad \forall i, \tag{7.15}$$

where $n = 2$ if $\nabla H(\underline{a})$ and $n = 1$ otherwise and $r_i = |x - a_i|$, and

$$f = 1 + O\left(\frac{1}{d^2(x,\underline{a})}\right) \text{ if } d(x,\underline{a}) \gg 1, \tag{7.16}$$

where

$$d(x,\underline{a}) = \min_j |x - a_j|,$$

with the corresponding estimates of their first derivatives.

We construct a function satisfying (7.15) and (7.16). Let $D(z, \rho)$ denote the disc of radius ρ centered at a point z. Let $\{\chi_j\}_1^K$ be a smooth partition of unity, i.e., $\sum_{l=1}^K \chi_j = 1$, having the properties

$$B\left(a_j, \frac{1}{3}d_{\underline{a}}\right) \subset \operatorname{supp}\chi_j \quad \forall j$$

and

$$\nabla^n\chi_j = O(d_{\underline{a}}^{-n}), \quad n = 0, 1, 2.$$

Then, the function $f = \sum f_j\chi_j$ satisfies (7.15) and (7.16). Indeed, (7.13) is obvious, while (7.14) follows from the relation

$$f_j = 1 + O(r_j^{-1}). \tag{7.17}$$

We prove the following lemma.

Lemma 7.1. *Let* ψ *satisfy* (7.15), (7.16). *Then,*

$$\mathscr{E}_R(\psi) = E_R^{(0)} + \text{Rem} + O\left(\frac{1}{R^2}\right), \tag{7.18}$$

where $E_R^{(0)}$ *is given in Theorem* 7.1 *and* Rem *is given by* (7.2).

Proof. Let $D_j = D(a_j, r_0)$, a disc with the center at a_j and of radius $r_0 = d_{\underline{a}}/3$. We decompose the energy functional as

$$\mathscr{E}_R(\psi) = \sum_j \int\limits_{D_j} e(\psi) + \int\limits_{D_R\setminus\cup D_j} e(\psi), \tag{7.19}$$

where $e(\psi)$ is the energy density,

$$e(\psi) = \frac{1}{2}|\nabla\psi|^2 + \frac{1}{4}(|\psi|^2 - 1)^2. \tag{7.20}$$

Let $e_1(\varphi) = \frac{1}{2}|\nabla\varphi|^2$ and $\langle f(\psi)\rangle = f(\psi) - \sum_k f(\psi_k)$. Equation (4.6) implies

$$\int_{D_R\setminus\cup D_k} e(\psi) = \int_{D_R\setminus\cup D_k} e_1(\varphi_0) + \int_{D_R\setminus\cup D_k} O(d(x, \underline{a})^{-4}). \quad (7.21)$$

Next, estimates (7.17) and

$$\nabla|\psi_i| = O(r_j^{-3}) \quad (7.22)$$

give

$$\int_{D_R\setminus\cup D_k} e_1(\varphi_i) = \int_{D_R\setminus\cup D_k} e(\psi_i) + O(r_0^{-2}). \quad (7.23)$$

Together with Eq. (7.10), this yields

$$\int_{D_R\setminus\cup D_k} \langle e(\psi)\rangle = \frac{1}{2}\sum_{i\neq j}\int_{D_R\setminus\cup D_k} \nabla\varphi_i\nabla\varphi_j + O(r_0^{-2}). \quad (7.24)$$

Next, in the region D_i, we have $\psi = e^{i\varphi_0} f_i$, where $f_i \equiv |\psi_i|$. Expansion (7.9) implies that

$$\int_{D_i} \nabla\varphi_i\nabla\varphi_{(i)} = 0. \quad (7.25)$$

Using this relation, we obtain

$$\int_{D_i} e(\psi) = \int_{D_i} e(\psi_i) + \int_{D_i} e_1(\varphi_{(i)}) + R,$$

where $R = \int_{D_i}(f_i^2 - 1)\alpha_i$. Expanding

$$\nabla\varphi_{(i)} = \nabla\varphi_{(i)}(a_i) + O\left(\frac{r_i}{d_{\underline{a}}^2}\right) \quad (7.26)$$

and using that $|\nabla\varphi_{(i)}(x)|^2 = O(d(x, \underline{a})^{-2})$, $\nabla\varphi_i(x) = O(r_i^{-1})$, and $\int_{D_i}(1 - f_i^2)\nabla\varphi_i = 0$, we obtain

$$R = O\left(\frac{\ln r_0}{d_{\underline{a}}^2}\right).$$

In the forceless case, we can improve this estimate using relation (7.9) again to show that, as in (7.12),

$$\int_{D_i}(f_i^2 - 1)\nabla\varphi_i\nabla\varphi_{(i)} = \int_{D_i}(f_i^2 - 1)\left(-c_i\sin 2\theta_i + O\left(\frac{r_i}{d_{\underline{a}}^3}\right)\right) = \int_{D_i}(f_i^2 - 1)O\left(\frac{r_i}{d_{\underline{a}}^3}\right) = O\left(\frac{r_0}{d_{\underline{a}}^3}\right).$$

This gives

$$R = O\left(\frac{r_0}{d_{\underline{a}}^3}\right) \text{ if } \nabla\varphi_i(a_i) = 0.$$

Finally, we observe that, due to (7.15),

$$\frac{1}{2}\int_{D_k}|\nabla\varphi_{(k)}|^2 = \sum_{j\neq k}\int_{D_k} e_1(\psi_j) + I_{D_k} = \sum_{j\neq k}\int e(\psi_j) + I_{D_k} + O(r_0^{-2}),$$

where

$$I_D := \frac{1}{2}\sum_{i\neq j}\int_D \nabla\varphi_i\cdot\nabla\varphi_j.$$

Collecting the estimates above, we arrive at

$$\int_{D_k}\langle e(\psi)\rangle = I_{D_k} + O\left(\frac{\ln r_0}{d_{\underline{a}}^2}\right) + O\left(\frac{1}{r_0^2}\right), \quad (7.27)$$

which, together with (7.9) and (7.16), yields

$$\mathscr{E}_R(\psi) = E + \mathrm{Rem}, \quad (7.28)$$

where Rem is given in (7.2) and

$$E = \int\left(g - \frac{(\deg\psi)^2}{n^2}\chi\right)$$

with

$$g = \sum_j e(\psi_j) + \frac{1}{2}\sum_{i\neq j}\nabla\varphi_i\nabla\varphi_j.$$

Now, by definition of the cutoff function χ ($\chi \geq 0$, $\chi = 1$ for $|x| \geq R$), we have

$$E \leq \int_{B(0,R)} g + \int_{B(0,R)^c}\left(g - \frac{n}{2r^2}\right), \quad (7.29)$$

where $n = \deg\psi$. We first compute the first integral in the right-hand side.

By definition of $E_{n,R}$ and because $a_i \ll R$, we have

$$\int_{D_R} e(\psi_i) = \int_{D_R + a_i} e(\psi^{(n_i)}) = E_{n_i,R} + O\left(\frac{1}{R^2}\right). \quad (7.30)$$

We now show that

$$I_{D_R} \equiv \frac{1}{2}\sum_{i\neq j}\int_{D_R} \nabla\varphi_i \nabla\varphi_j = -\sum_{i\neq j}\pi n_i n_j \ln\left(\frac{|a_{ij}|}{R}\right). \quad (7.31)$$

We compute

$$\int_{D_R} \nabla\varphi_i \nabla\varphi_j = n_i n_j \int_0^{2\pi R}\int_0 \frac{r - a\cos\theta}{r^2 + a^2 - 2ar\cos\theta} dr d\theta, \quad (7.32)$$

where $a = |a_{ij}|$. Furthermore, changing the integration variable as $\theta \longrightarrow z = e^{i\theta}$ and computing the residue, we find

$$\int_0^{2\pi} \frac{r - a\cos\theta}{r^2 + a^2 - 2ar\cos\theta} d\theta = \frac{\pi}{r} - \frac{r^2 - a^2}{2iar^2}\oint_{|z|=1} \frac{dz}{\left(z - \frac{r}{a}\right)\left(z - \frac{a}{r}\right)}$$

$$= \frac{\pi}{r} + \frac{\pi}{r}\frac{r^2 - a^2}{|r^2 - a^2|} = \frac{2\pi}{r}\begin{cases} 1 & \text{if } r > a, \\ 0 & \text{if } r < a. \end{cases}$$

The last two equations yield (7.24). We also observe that, up to a multiplicative constant, expression (7.24) can be found from the symmetry considerations: the invariance of the integral in the left-hand side under translations ($a_i \longrightarrow a_i + h$ and $a_j \longrightarrow a_j + h\ \forall h \in \mathbb{R}^2$) and rotations ($a_i \longrightarrow g a_i$ and $a_j \longrightarrow g a_j\ \forall g \in O(2)$) implies that it depends only on $|a_{ij}|$. Its scaling properties under the dilations ($a_i \longrightarrow \lambda a_i$ and $a_j \longrightarrow \lambda a_j\ \forall \lambda \in \mathbb{R}$) imply that it is a multiple of $\ln(|a_{ij}|/R)$.

Equations (7.30) and (7.31) imply that

$$\int_{B(0,R)} g = \sum E_{n_i,R} + H\left(\frac{\underline{a}}{R}\right) + O\left(\frac{1}{R^2}\right). \quad (7.33)$$

Next, we estimate the second integral in the right-hand side of (7.29). By Eqs. (7.17) and (7.22), we have

$$g = \frac{1}{2}|\nabla\varphi_0|^2 + O(d(x,\underline{a})^{-4}).$$

Furthermore, expanding the terms $\nabla\theta(x - a_j)$ in $\nabla\varphi_0(x) = \sum n_j \nabla\theta(x - a_j)$ around the point x, we obtain

$$\nabla\varphi_0(x) = n\nabla\theta(x) - \theta''(x)\sum n_j a_j + O\left(\frac{\sum n_j a_j^2}{d(x,\underline{a})^3}\right), \quad (7.34)$$

where $\theta''(x)$ is the Hessian of $\theta(x)$. Choosing the origin such that $\sum n_j a_j = 0$ eliminates the second term on the right-hand side. (Otherwise, we could use, by an explicit computation,

$$\theta''(x)\nabla\theta(x) = -\frac{x}{r^4},$$

the integral of which over the exterior of the ball $B(0, R)$ vanishes.) Hence,

$$\int_{B(0,R)^c} \left(f - \frac{n^2}{2r^2}\right) = \int_{B(0,R)^c} O\left(\frac{\sum n_j a_j^2}{d(x,\underline{a})^4}\right) = O\left(\frac{\sum n_j a_j^2}{R^2}\right). \quad (7.35)$$

Estimates (7.28), (7.29), (7.33), and (7.35) imply (7.7) with Rem given in (7.2).

Remark 7.3. The statement of Lemma 7.2 remains true for a wider class of functions defined by replacing (7.7) by the condition

$$f = f_i + O\left(\frac{1}{r d_{\underline{a}}^n}\right) \text{ and } \int_0^{2\pi} \mathrm{Re}(e^{-i\varphi_0}\psi - f_i) d\theta = O\left(\frac{1}{d_{\underline{a}}^{n+1}}\right) \text{ if } |x - a_i| \ll d_{\underline{a}}, \quad (7.36)$$

with the corresponding estimates of their first derivatives, where $n = 2$ if $\nabla H(\underline{a}) = 0$ and $n = 1$ otherwise.

To prove this, we write ψ in the region D_i as $\psi = e^{i\varphi_0}(f_i + \xi)$, where $f_i \equiv |\psi_i|$. Using relation (7.25) and

$$\int_{D_j} f_j \nabla\varphi_j \nabla \mathrm{Im}\xi = n_j \int_{D_j} f_j \frac{\partial}{\partial\theta}\mathrm{Im}\xi = 0, \quad (7.37)$$

we obtain

$$\int_{D_i} e(\psi) = \int_{D_i} e(\psi_i) + \int_{D_i} e_1(\varphi_{(i)}) + R + R', \quad (7.38)$$

where R is given above and

$$R' = \int_{D_i}\Big\{(|\nabla\varphi_0|^2 + f_i^2 - 1) f_i \mathrm{Re}\xi + f_i^2(\mathrm{Re}\xi)^2 + \frac{1}{2}|\nabla\varphi_0|^2|\xi|^2 + \frac{1}{2}|\nabla\xi|^2 + 2\nabla f_i \nabla \mathrm{Re}\xi + f_i \nabla\varphi_{(i)} \nabla \mathrm{Im}\xi + \mathrm{Im}(\xi\nabla\varphi_0 \cdot \nabla\xi) + \frac{1}{2}(f_i^2 - 1 + 2 f_i \mathrm{Re}\xi)|\xi|^2 + \frac{1}{4}|\xi|^4\Big\}. \quad (7.39)$$

Using

$$\xi = O\Big(\frac{1}{r d_{\underline{a}}}\Big)$$

and

$$\int_0^{2\pi} \mathrm{Re}\xi\, d\theta = O\Big(\frac{1}{d_{\underline{a}}^2}\Big) \text{ in } D_j$$

due to (7.36) and that $|\nabla\varphi_i|^2 + f_i^2 - 1 = O(r_i^{-4})$, we find

$$R' = O\Big(\frac{\ln r_0}{d_{\underline{a}}^2}\Big). \quad (7.40)$$

We now proceed to proving estimate (7.4) with Rem given by (7.3). First, we describe the class of test functions for which we prove this estimate: $\psi = e^{i\varphi_0} f$ with

$$f = \begin{cases} f_j - \frac{1}{2} f_j^{-1} \alpha_j \eta_j \text{ in } D\Big(a_j, \frac{1}{3} d_{\underline{a}}\Big) \ \forall j, & (7.41) \\ 1 - \frac{1}{2}|\nabla\varphi_0|^2 + O(d(x, \underline{a})^{-4}) \\ \quad \text{in } \Big(\bigcup_j D\Big(a_j, \frac{1}{4} D_{\underline{a}}\Big)\Big)^c, & (7.42) \end{cases}$$

where we used definition (7.11) and where η_j are smooth cutoff functions depending only on $r_j = |x_j|$ (i.e., radially symmetric in the x_j variables) satisfying

$$B\Big(a_j, \frac{1}{2} d_{\underline{a}}\Big)\backslash B(a_j, 2 d_{\underline{a}}^{\gamma}) \subset \mathrm{supp}\eta_j \subset B\Big(a_j, \frac{1}{2} d_{\underline{a}}\Big)\backslash B(a_j, d_{\underline{a}}^{\gamma}) \quad (7.43)$$

and

$$\nabla^n \eta_j = O(d_{\underline{a}}^{-\gamma n}), \quad n = 0, 1, 2, \quad (7.44)$$

for $\gamma = \frac{1}{3}$ (not optimal). (The f_j^{-1}'s in (7.41) play no important role and are chosen purely with a view of simplifying some expressions below.)

The function

$$f = \sum f_j \chi_j - \sum \frac{1}{2} f_j^{-1} \alpha_j \eta_j \quad (7.45)$$

satisfies Eqs. (7.41) and (7.42). To prove this, we use the expansion

$$f_j = 1 - \frac{1}{2}|\nabla\varphi_j|^2 + O(r_j^{-4}) \quad (7.46)$$

and the estimate

$$\alpha_j = O(d_{\underline{a}}^{-2}) \text{ in } D(a_j, d_{\underline{a}}), \quad (7.47)$$

which is shown by expanding the function $\nabla\varphi_{(j)}(x)$ around a_j and using

$$\nabla\varphi_{(j)}(a_j) = -\frac{1}{2\pi n_j} \nabla_{a_j} H(\underline{a}) = 0$$

and

$$\nabla\varphi_j(x) = O(r_j^{-1}).$$

Our next task is to prove the following lemma.

Lemma 7.4. *Let $\underline{a}$ be forceless in the sense that $\nabla H(\underline{a}) = 0$. Then, estimate* (7.7) *with* (7.3) *holds for any function ψ satisfying* (7.21), (7.22).

Proof. The proof follows the lines of the proof of Lemma 7.2, but with some subtle modifications considered below.

First of all, instead of $e_1(\psi) = \frac{1}{2}|\nabla\varphi|^2$ used in the proof of Lemma 7.2, we use the density

$$e_2(\varphi) = \frac{1}{2}|\nabla\varphi|^2 - \frac{1}{4}|\nabla\varphi|^4, \quad (7.48)$$

which is a better approximation to the density $e(\psi)$. We also use (7.27) instead of (7.17). In particular, we have

$$e(\psi_j) = e_2(\varphi_j) + O(r_j^{-6}). \quad (7.49)$$

We set $f_j := 1 - f_j^2 - |\nabla\varphi_j|^2$. For any k and for $u_k = e^{i\varphi_0}(f_k + \xi)$, where ξ is a real function, we have the identity

$$\langle e(u_k)\rangle = \frac{1}{2}\sum_{i \neq j} \nabla\varphi_i \nabla\varphi_j - A(\varphi) + B_k(\xi) + R_k, \quad (7.50)$$

where

$$B_k(\xi) := -\frac{1}{2}g_k(\alpha_k + 2f_k\xi) + \frac{1}{4}\alpha_k^2 + \alpha_k f_k\xi + f_k^2\xi_k^2 \tag{7.51}$$

and

$$R = \sum_{j\neq k}(e_2(\varphi_j) - e(\psi_j)) - \frac{1}{2}(g_k - \alpha_k)\xi^2 + f_k\xi_k + \frac{1}{4}\xi_k^4 + \frac{1}{2}(2\nabla f_k\nabla\xi + |\nabla\xi|^2). \tag{7.52}$$

We now take $\xi = -\frac{1}{2}f_k^{-1}\alpha_k\eta_k$. Then,

$$e(\psi) = e(u_k) \text{ on } D\left(a_k, \frac{1}{3}d_{\underline{a}}\right). \tag{7.53}$$

Due to (7.28) and the corresponding estimate for the derivatives of α_j and due to (7.25), (7.27), and (7.29), we have

$$R_k = O(d_{\underline{a}}^{-4\gamma-2}). \tag{7.54}$$

We note that the form of (7.21) is chosen such that

$$B_k(\xi) = 0 \text{ on } B\left(a_k, \frac{1}{2}d_{\underline{a}}\right)\backslash B(a_k, d_{\underline{a}}^{\gamma}) \subset \{\eta_k = 1\}.$$

Next, we estimate $B_k(\xi)$ on the entire disc $D\left(a_k, \frac{1}{3}d_{\underline{a}}\right)$. Expanding the function $\nabla\varphi_{(k)}(x)$ around the point a_k and using that

$$\nabla\varphi_{(k)}(a_k) = -\frac{1}{2\pi n_k}J\nabla_{a_k}H(\underline{a}) = 0,$$

we find

$$\alpha_k(x) = 2\nabla\varphi_k(x)\varphi''_{(k)}(a_k)x_k + O(r_k d_{\underline{a}}^{-3}), \tag{7.55}$$

where $x_k = -a_k$ and φ'' is the Hessian (the matrix of second derivatives) of a function φ. Using this expression in estimating $B_k(\xi)$, we find

$$B_k(\xi) = -g_k\nabla\varphi_k(x)\varphi''_{(k)}(a_k)x_k\bar{\eta}_k + O(r^{-3}d_{\underline{a}}^{-3} + d_{\underline{a}}^{-4})\bar{\eta}_k \text{ on } D\left(a_k, \frac{1}{3}d_{\underline{a}}\right), \tag{7.56}$$

where $\bar{\eta}_k = 1 - \eta_k$. The first term in the right-hand side of this expression is singular at $x_k = x - a_k = 0$, but the integral of it is conditionally convergent and equals 0. Indeed, because the function $\varphi_{(k)}(x)$ is harmonic in $D\left(a_k, \frac{1}{3}d_{\underline{a}}\right)$, we have (cf. (7.9))

$$\varphi''_{(k)}(a_k)x_k = c(x_k\cos 2\theta_k - x_k^{\perp}\sin 2\theta_k), \tag{7.57}$$

where $c = O(d_{\underline{a}}^{-2})$, $x^{\perp} = (-x_2, x_1)$, and θ_k is the polar angle of x_k (see Eq. (7.9)). Because g_k and $\bar{\eta}_k$ depend only on r_k (we write $(g_k\bar{\eta}_k)(r_k)$ for $g_k(x)\bar{\eta}_k(x)$), we have

$$\int(g_k\bar{\eta}_k)(r_k)\nabla\varphi_k(x)\varphi''_{(k)}(a_k)x_k = -c\int(g_k\bar{\eta}_k)(r_k)\sin 2\theta_k = 0 \tag{7.58}$$

(strictly speaking, we must first excise a small disc around $x_k = 0$ and then take the radius of this disc to zero).

Equations (7.32), (7.33), (7.35), and (7.37) imply that

$$\int_{D\left(a_k, \frac{1}{3}d_{\underline{a}}\right)}\langle e(\psi)\rangle = \int_{D\left(a_k, \frac{1}{3}d_{\underline{a}}\right)}\left(\frac{1}{2}\sum_{i\neq j}\nabla\varphi_i\nabla\varphi_j - A(\varphi)\right) + O(d_{\underline{a}}^{-3} + d_{\underline{a}}^{-2-4\gamma} + d_{\underline{a}}^{-4+2\gamma}). \tag{7.59}$$

Finally, we derive the estimate

$$\langle e(\psi)\rangle = \frac{1}{2}\sum_{i\neq j}\nabla\varphi_i\nabla\varphi_j - A(\varphi) + O(d(x, \underline{a})^{-6}) \tag{7.60}$$

on $\left(\bigcup_k D\left(a_k, \frac{1}{4}d_{\underline{a}}\right)\right)^c$. Indeed, Eq. (7.42) implies that

$$e(\psi) = e_2(\varphi_0) + O(d(x, \underline{a})^{-6}), \tag{7.61}$$

which, together with (7.49), implies (7.60).

Now, Eqs. (7.59) and (7.60) with $\gamma = 1/3$ imply

$$\mathcal{E}_R(\psi) = E - A(\underline{a}) + O(d_{\underline{a}}^{8/3}), \tag{7.62}$$

where the term E is defined after Eq. (7.28) and $A(\underline{a}) = \int A(\varphi)$. Equations (7.29), (7.33), (7.35), and (7.61) imply (7.14) with Rem given by (7.3).

Lemmas 7.2 and 7.4 and inequality (7.13) imply Theorem 7.1.

8. DISCUSSION

In this paper, we investigated the Ginzburg–Landau equation (1.1) that appears in condensed matter physics and nonlinear optics. Specifically, we presented careful arguments supporting the existence of nonradial-symmetric solutions corresponding to vortex configurations c with $N + 1$ vortices fixed by the symmetry group C_{Nv}.

In these configurations, N m vortices lie on the circle of radius a, one $(-k)$ vortex is placed at the center of the circle, and the only remaining free parameter is the overall size of the configuration, the radius of the circle a.

Our argument is based on reducing the problem of the existence of solutions corresponding to a given vortex configuration to the existence of critical points of the effective energy of the vortex configurations introduced in this paper. For C_{Nv} configurations, this effective energy is a function of a single variable, a. To prove the existence of critical points of this energy, we investigated it analytically and numerically at large and small values of parameter a. We found that there are critical points at the vortex configurations ($N = 2$, $m = 2$, $k = 1$) and ($N = 4$, $m = 2$, $k = 3$) and, consequently, we expect the existence of (static) solutions corresponding to these configurations. For the vortex configuration ($N = 3$, $m = 1$, $k = 1$), our numerical analysis indicates that it is very likely that such a critical point exists. Our numerical computations suggest that the critical a's are on the order of $O(1)$. Finding their true values requires rather elaborate numerical analysis, which would be desirable to develop but which is presently lacking. In addition, we have shown (see Appendix 5) that, for the vortex configurations (N, 1, $(N-1)/2$) with odd $N \gg 1$ and sufficiently large a, the energy is greater than the effective energy of a single N vortex.

All solutions considered are saddle points of the renormalized Ginzburg–Landau energy functional. Perturbations breaking the C_{Nv} symmetry group can lower the energy of the corresponding solution vortex configuration. However, we expect that, under small symmetry breaking perturbations, such solutions lead to long-living metastable states that can be observed experimentally. Moreover, even weak pinning centers can stabilize such solutions. Thus, to experimentally observe the static configurations found in this paper, one would need to create weak pinning potentials satisfying the suggested point symmetry, adjust the radius a at which these potentials are located, and then slowly reduce the strength of these potentials to zero.

ACKNOWLEDGMENTS

The authors are grateful to A. Knauf and L. Sadun for stimulating discussion. In fact, it was A. Knauf who suggested that there might be static multivortex configurations. The authors are also grateful to the referee for useful suggestions.

Research in this paper was supported by NSERC (grant no. NA7901) and by CRDF (grant no. 2565-MO-03 U.S.A.) and by the Russian Foundation for Basic Research.

APPENDIX 1

Computation of c(n)

In this appendix, we compute the constants $c(n)$ in expression (3.5) for the self-energy $E_{n,R}$ of the n vortex (see Eq. (3.6)). For this, we derive a convenient formula for $E_{n,R}$. Multiplying Eq. (3.2) by $r^2 f_n'(r)$, where $f'(r) = \partial f(r)/\partial r$, integrating the result over r, observing that the first two integrands are total derivatives, and integrating the last term by parts, we obtain the quantization relation (see [22])

$$\int_0^\infty (1 - f_n^2)^2 r dr = n^2.$$

This equation, together with Eq. (3.3), yields an expression for $E_{n,R}$,

$$E_{n,R} = -\frac{\pi}{2}n^2 + \pi\int_0^\infty \left(1 - f_n^2 - \frac{n^2}{r^2}\chi\right) r dr.$$

However, we prefer to use a different representation of $E_{n,R}$, which is obtained from above if we write $1 - f_n^2 = (1 - f_n^2) f_n^2 + (1 - f_n^2)^2$ and use the quantization formula above again:

$$E_{n,R} = \frac{\pi}{2}n^2 + \pi\int_0^\infty \left[(1 - f_n^2) f_n^2 - \frac{n^2}{r^2}\chi\right] r dr. \quad \text{(A.1.1)}$$

To avoid numerical evaluation of the integral in (A.1.1) over an infinite range, we use the expansion of $f_n(r)$ in $1/r$ for large r. However, $f_n(r)$ is not analytic at $r = \infty$; it has an essential singularity at this point. Hence, the resulting series is asymptotic. We truncate this series at the order $O\left(\frac{1}{r^6}\right)$. To compensate for this truncation, we add to the resulting polynomial in $1/r$ a multiple of the decaying solution $e^{-\sqrt{2}r}/\sqrt{r}$ of the linearization of Eq. (3.2) around 1. We should linearize Eq. (3.2) around the resulting polynomial, but the powers of $1/r^2$ lead to similar powers multiplying $e^{-\sqrt{2}r}/r$, and it therefore suffices to linearize around 1. The result is

$$f_n(r) = \left\{1 - \frac{n^2}{2r^2} - \frac{n^2(1 + n^2/8)}{r^4} - \frac{1}{r^6}\left(\frac{n^4}{2} + \frac{n^2 + 16}{2}\left(n^2 + \frac{n^4}{8}\right)\right) - \ldots\right\} - c\frac{e^{-\sqrt{2}r}}{\sqrt{r}}(1 + \ldots), \quad \text{(A.1.2)}$$

where c is a constant to be determined by a matching procedure. Inserting this expression in Eq. (A.1.1), we obtain

$$E_{n,R} - \pi n^2 \ln R = \frac{\pi n^2}{2} + \pi \int_0^{r_0} f_n^2(1 - f_n^2) r dr - \pi n^2 \left(\ln r_0 + \frac{n^2 - 2}{2r_0^2} + \frac{n^2 - 16}{4r_0^4} \right) + O(r_0^{-6}) \quad \text{(A.1.3)}$$

for any $r_0 > 0$. We choose $6 \le r_0 \le 10$. This relation together with Eq. (3.5) implies that

$$\frac{1}{\pi} c(n) = \frac{n^2}{2} + \int_0^{r_0} f_n^2(1 - f_n^2) r dr - n^2 \left(\ln \frac{r_0}{|n|} + \frac{n^2 - 2}{2r_0^2} + \frac{n^2 - 16}{4r_0^4} \right) + O(r_0^{-6}). \quad \text{(A.1.4)}$$

For numerical solution of Eq. (3.2), we take the interval $(0.3, r_0)$. Because Eq. (3.2) linearized around the function 1 has the solutions

$$\frac{1}{\sqrt{r}} e^{\pm \sqrt{2} r}, \quad \text{(A.1.5)}$$

we should apply the numerical iteration procedure starting from the upper limit, r_0. Then, the dangerous, exponentially growing solution would not affect our procedure.

In the range $0 < r \le 0.3$, we use the fact that, as Eq. (3.2) shows, the function $f_n(r)$ is analytic in a disc $|r| < O(1)$ and can therefore be presented by a convergent series

$$f_n(r) = \alpha r^n \left\{ 1 - \frac{r^2}{4(n+1)} + \frac{r^4}{8(n+2)} \times \left(\frac{1}{4(n+1)} + \alpha^2 \delta_{n,1} \right) + \frac{r^6}{12(n+3)} \times \left[\alpha^2 \left(\delta_{n,2} - \frac{3}{4(n+1)} \delta_{n,1} \right) - \frac{1}{8(n+2)} \times \left(\frac{1}{4(n+1)} + \alpha^2 \delta_{n,1} \right) \right] + \ldots \right\} \quad \text{(A.1.6)}$$

for some number $\alpha > 0$. Here, $\delta_{n,k}$ is the Kronecker symbol, $\delta_{n,k} = 1$ for $n = k$ and $\delta_{n,k} = 0$ for $n \ne k$. (We expect that the pole closest to the origin lies on the imaginary axis.)

To finish the computation of $c(n)$, we must find the value of the parameters α and c. This is done by matching solution (A.1.2) for small r with solution (A.1.6) for large r. Specifically, using Eq. (A.1.2), we compute $f_n(r_0)$ and $f_n'(r_0)$ for various values of the parameter c. Using these values as initial conditions, we integrate Eq. (3.2) backward to $r = 0.3$, which yields $f_{\text{right}}(0.3)$ and $f'_{\text{right}}(0.3)$. On the other hand, using Eq. (A.1.6), we compute $f_{\text{left}}(0.3)$ and $f'_{\text{left}}(0.3)$ for various values of the parameter α. We then match $f_{\text{right}}(0.3)$ and $f'_{\text{right}}(0.3)$ with $f_{\text{left}}(0.3)$ and $f'_{\text{left}}(0.3)$ by minimizing $[(f_{\text{right}}(0.3) - f_{\text{left}}(0.3))^2 + (f'_{\text{right}}(0.3) - f'_{\text{left}}(0.3))^2]^{1/2}$. This yields the values of the parameters c and α. After this, we compute $c(n)$ using formulas (A.1.4) and (A.1.6).

APPENDIX 2

Large-n Asymptotic Form of the Vortex (Self) Energy

In this appendix, we find the large-n asymptotic form of the constant $c(n)$ in expression (3.6) for the (self) energy of the n vortex. For this, we use the large-n asymptotic expression for the function $f_n(r)$ defined in (3.2),

$$f_n(r) = \begin{cases} \sqrt{1 - n^2/r^2} & \text{if } r - n \gg (n/2)^{1/3}, \\ (2/n)^{1/3} g(z) & \text{if } |r - n| \ll n, \end{cases} \quad \text{(A.2.1)}$$

where the variable z is defined by

$$r = n + \left(\frac{n}{2} \right)^{1/3} z, \quad \text{(A.2.2)}$$

and the function $g(z)$ is a solution of the equation

$$g'' + zg - g^3 = 0. \quad \text{(A.2.3)}$$

The function $g(z)$ has the asymptotic form

$$g(z) = z^{1/2} \text{ if } z \gg 1, \quad g(z) = \text{const}\,\phi(z) \text{ if } z \ll -1, \quad \text{(A.2.4)}$$

where $\phi(z)$ is the Airy function. In particular, we have

$$g(z) = \frac{0.39}{(-z)^{1/4}} e^{-2(-z)^{3/2}/3} \text{ for } z \ll -1. \quad \text{(A.2.5)}$$

Inserting expressions (A.2.1) and (A.2.2) in Eq. (A.1.1) and using (A.2.4) and (A.2.5), we find that

$$c(n) = \alpha n^{2/3} \pi + c + O(n^{-2/3}), \quad \text{(A.2.6)}$$

where c is some constant and

$$\alpha = 2^{1/3} \int_{-\infty}^{\infty} (g^2(z) - z\theta(z)) dz, \quad \text{(A.2.7)}$$

with $\theta(z) = 1$ for $z \geq 0$ and $\theta(z) = 0$ for $z < 0$. Multiplying Eq. (A.2.3) by $g'(z)$ and integrating the result, we find that $\alpha = 0$, and therefore,

$$c(n) = c + O(n^{-2/3}) \tag{A.2.8}$$

as $n \longrightarrow \infty$. A rough numerical computation yields the following value for the constant c:

$$c \approx 0.7\pi. \tag{A.2.9}$$

APPENDIX 3

Computation of Correlation Coefficients

In this appendix, we compute the correlation function

$$A = A(\underline{a}) = \frac{1}{4}\int\left[|\nabla\varphi_0|^4 - \sum_j |\nabla\varphi_j|^4\right], \tag{A.3.1}$$

with

$$\varphi_0 = \sum_j \varphi_j \text{ and } \varphi_j(x) = n_j\theta(x - a_j), \tag{A.3.2}$$

$$\theta(x) = \text{the polar angle of } x \in \mathbb{R}^2, \tag{A.3.3}$$

(see Eq. (4.4)) for configurations of $K = N + 1$ vortices with N vortices of vorticity m lying on the circle of radius a and one vortex of vorticity $-k$ at the center of this circle, such that $\nabla H(\underline{a}) = 0$.

We write $\underline{a} = a \cdot \underline{b}$, where $\underline{b}$ is a fixed configuration with N vortices on the unit circle and one at the center. Changing the integration variable in (A.3.4) as $x = ay$, we find

$$A(\underline{a}) = Ca^{-2}, \tag{A.3.4}$$

where C depends on $\underline{b}$ only. Our task now is to find the sign of C for the configurations of interest. We write $A = A(\underline{a})$.

1. $N = 2$, $m = 2$, and $k = 1$. In this case, there are two double vortices on the circle and one single vortex of the opposite vorticity at the center (see Fig. 1). Below, we use the dimensionless variable

$$\rho = \frac{|x|}{a}. \tag{A.3.5}$$

For the configuration under consideration, we have

$$A = \frac{1}{4a^2}\int_0^\infty \rho d\rho \int_0^{2\pi} d\theta\left\{\frac{48}{\alpha} - \frac{16\cos(2\theta)}{\alpha\rho^2}\right. \tag{A.3.6}$$
$$\left. + \frac{64\cos^2(2\theta)}{\alpha^2} - \frac{64}{\alpha^2}(1 + 2\rho^2 + 2\rho^2\cos(2\theta))\right\},$$

where

$$\alpha = \rho^4 + 1 + 2\rho^2\cos(2\theta). \tag{A.3.7}$$

(In general, for a_j, $j = 1, \ldots, N$, distributed equidistantly on the circle of radius a, $\alpha = \prod_{j=1}^N (x - a_j)^2/a^{2N}$.) First, we take the integral over θ. For this, we change the integration variable as $\theta \longrightarrow z = \exp(2i\theta)$; i.e., we write the inner integral in (A.3.8) as an integral over the unit circle. A simple calculation gives

$$\int_0^{2\pi} \frac{d\theta}{\alpha^2} = \frac{2\pi(1 + \rho^4)}{|1 - \rho^4|^3},$$
$$\int_0^{2\pi} \frac{d\theta}{\alpha^2}\cos(2\theta) = -\frac{4\pi\rho^2}{|1 - \rho^4|^3}, \tag{A.3.8}$$

$$\int_0^{2\pi} \frac{d\theta}{\alpha} = \frac{2\pi}{|1 - \rho^4|},$$
$$\int_0^{2\pi} \frac{d\theta}{\alpha}\cos(2\theta) = -\frac{2\pi}{|1 - \rho^4|}\min\left\{\rho^2, \frac{1}{\rho^2}\right\}, \tag{A.3.9}$$

$$\int_0^{2\pi} \frac{d\theta}{\alpha^2}\cos^2(2\theta) = \frac{\pi}{|1 - \rho^4|^3}$$
$$\times\begin{cases} 1 + 4\rho^4 - \rho^8 & \text{for } \rho < 1, \\ (\rho^8 + 4\rho^4 - 1)/\rho^4 & \text{for } \rho > 1. \end{cases} \tag{A.3.10}$$

Inserting expressions (A.3.7)–(A.3.10) in Eq. (A.3.6), we obtain

$$A = \frac{4\pi}{a^2}\left\{2\int_0^1 dx\frac{1 - x}{(1 + x)^3}\right.$$
$$\left. + \int_1^\infty \frac{dx}{(1 + x)^3}\left(3x + 1 + \frac{3}{x} + \frac{1}{x^2}\right)\right\}.$$

This gives

$$A = \frac{8\pi}{a^2}. \tag{A.3.11}$$

Hence, in the configuration under consideration, the energy $E_R(\underline{a})$ is given by

$$\frac{1}{\pi}E_R(\underline{a}) - 9\ln R = -9.64 - \frac{8}{a^2} + O\left(\frac{\ln a}{a^4}\right). \tag{A.3.12}$$

2. $N = 3$, $m = 1$, and $k = 1$. Similarly to Eq. (A.3.8), we obtain

$$A = \frac{1}{4a^2}\int_0^\infty d\rho\rho\int_0^{2\pi} d\theta\left\{\frac{6}{\alpha}(1+2\rho^2)\right.$$
$$-\frac{12\sin(3\theta)}{\rho\alpha}-\frac{9(1+\rho^2)}{\alpha^2}(1+\rho^2+2\rho^4) \qquad \text{(A.3.13)}$$
$$\left.+\frac{36\rho^2\sin^2(3\theta)}{\alpha^2}-\frac{36\rho^5\sin(3\theta)}{\alpha^2}\right\},$$

where $\alpha = \rho^6 + 1 + 2r^3\sin(3\theta)$. The integrals in Eq. (A.3.13) can be taken explicitly. To do this, we set $z = \exp(3i\theta)$, and then

$$\int_0^{2\pi}\frac{d\theta}{\alpha} = \frac{2\pi}{|1-\rho^6|},$$
$$\int_0^{2\pi}\frac{d\theta}{\alpha}\sin(3\theta) = -\frac{2\pi}{|1-\rho^6|}\min\left(\rho^3,\frac{1}{\rho^3}\right), \qquad \text{(A.3.14)}$$

$$\int_0^{2\pi}\frac{d\theta}{\alpha^2} = \frac{2\pi(1+\rho^6)}{|1-\rho^6|^3},$$
$$\int_0^{2\pi}\frac{d\theta}{\alpha^2}\sin(3\theta) = -\frac{4\pi\rho^3}{|1-\rho^6|^3}, \qquad \text{(A.3.15)}$$

and

$$\int_0^{2\pi}\frac{d\theta}{\alpha^2}\sin^2(3\theta) = \frac{\pi}{|1-\rho^6|^3}$$
$$\times\begin{cases}1+4\rho^6-\rho^{12} & \text{for } \rho<1,\\ (\rho^{12}+4\rho^6-1)/\rho^6 & \text{for } \rho>1.\end{cases} \qquad \text{(A.3.16)}$$

Inserting expressions (A.3.14)–(A.3.16) in Eq. (A.3.13), we obtain

$$A = \frac{3\pi}{4a^2}\left\{\int_0^1 dx\frac{5x+9x^2-1-2x^3-2x^4}{(1+x+x^2)^3}\right.$$
$$+\int_1^\infty dx\left(\frac{4}{1+x+x^2}-\frac{9}{(1+x+x^2)^2}\right. \qquad \text{(A.3.17)}$$
$$\left.\left.+\frac{10x+18}{(1+x+x^2)^3}+\frac{6x+2}{x^2(1+x+x^2)^3}\right)\right\}.$$

A simple calculation of integrals in Eq. (A.3.16) gives explicit answers for A:

$$A = \frac{2\pi}{a^2}. \qquad \text{(A.3.18)}$$

Hence, the energy for such configurations is given by

$$\frac{1}{\pi}E_R(a) - 4\ln R = -1.792 - \frac{2}{a^2}. \qquad \text{(A.3.19)}$$

3. $N = 4$, $m = 2$, and $k = 3$. In this case, there are four double vortices in the corners of a rectangle and a (–3)-vortex at the center. For this configuration, we have

$$A = \frac{16}{a^2}\int_0^{2\pi} d\theta\int_0^\infty\frac{d\rho\rho}{\alpha}\left\{\frac{4\rho^{12}}{\alpha}+\frac{36\rho^4\cos^2(4\theta)}{\alpha}\right.$$
$$+4.5\rho^4+13.5\cos(4\theta)+\frac{24\rho^8}{\alpha}\cos(4\theta) \qquad \text{(A.3.20)}$$
$$-\frac{1}{\alpha}[(\rho^2+1)^6-2\rho^2(\rho^2+1)^2(\rho^4+1)+4\rho^6]$$
$$\left.-2\rho^4\cos(4\theta)(3(\rho^2+1)^2-2\rho^2)/\alpha\right\},$$

where

$$\alpha = \rho^8 + 1 - 2\rho^4\cos(4\theta).$$

The change of variables $2\theta \longrightarrow \tilde{\theta} + \pi/2$, $\rho^8 \longrightarrow \tilde{\rho}^4$ reduces the integrals over θ in Eq. (A.3.20) to those in Eqs. (A.3.8)–(A.3.10). As a result, we obtain

$$A = \frac{16\pi}{a^2}\left\{\int_0^1 dx\left[\frac{1-3x}{1+x+x^2+x^3}\right.\right.$$
$$+\frac{2(5x^5+23x^4+18x^3+6x^2-3x-1)}{(1+x+x^2+x^3)^3}$$
$$+\int_1^\infty dx\left(\frac{7.5}{1+x^2}-\frac{1.5}{x^2(1+x^2)}-\frac{4(1+x+x^2)}{x^2(1+x+x^2+x^3)^3}\right. \qquad \text{(A.3.21)}$$
$$-\frac{2}{(1+x+x^2+x^3)^3}\left(x^5+11x^4-2x^3-22x^2\right.$$
$$\left.\left.\left.\left.-31x-21-\frac{12}{x}-\frac{4}{x^2}\right)\right)\right]\right\}.$$

Direct calculation of the integrals in Eq. (A.3.11) gives

$$A = \frac{80\pi}{a^2}, \qquad \text{(A.3.22)}$$

and therefore, the energy of the configuration in question is

$$\frac{1}{\pi}E_R(a) - 25\ln R = -40.44 - \frac{80}{a^2}. \qquad \text{(A.3.23)}$$

We note that, for all the configurations under consideration, the correlation term A is given by

$$A = \frac{\pi}{4a^2}M,$$

where M is an integer; i.e., the quantity given by the integral in A is quantized. Moreover, the "quantization" takes place separately for the integrals over regions $r < 1$ and $r > 1$. We conjecture that this property is general and holds for any forceless configuration.

APPENDIX 4

Inequality $E_R'(0) > 0$

In this appendix, we show that $E_R(a) - E_R(0) > 0$ for the configuration consisting of N 1-vortices equidistributed on the circle of radius a and one $\left(-\frac{N-1}{2}\right)$ vortex at the center and for sufficiently small a. We assume that N is odd but otherwise arbitrary.

For $a = 0$, the configuration in question collapses to a single $\frac{N+1}{2}$-vortex, $\psi_{\frac{N+1}{2}}$, sitting at the origin. Let L be the Hessian of $\mathscr{E}_{\text{ren}}(\psi)$ at $\psi = \psi_{\frac{N+1}{2}}$. It was shown in [13] that the subspaces

$$\left\{u_1(r)e^{im\theta} + u_2(r)e^{i\left(2\frac{N+1}{2}-m\right)\theta}\middle| u_k \in L^2(rdr), \quad k = 1, 2\right\}, \qquad \text{(A.4.1)}$$

$m = \frac{N+1}{2}, \frac{N+1}{2} + 1, \ldots$, which are orthogonal to each other and span the entire Hilbert space $L^2(\mathbb{R}^2)$, are invariant under the action of the operator L. Moreover, it was shown that, in the sectors with $m \geq 3\frac{N-1}{2} - 1$, L is nonnegative and 0 is not its eigenvalue (actually, the statement in [13] is formulated for $m \geq 3\frac{N-1}{2}$, but the proof works also for $m = 3\frac{N-1}{2} - 1$), while in the sectors

$$\frac{N+1}{2} + 2 \leq m \leq 2\frac{N+1}{2},$$

the operator L has negative eigenvalues. We now observe that the sectors with $\frac{N+1}{2} \leq m \leq 3\frac{N-1}{2} - 2$ do not have the C_{Nv} symmetry and, consequently, are forbidden in our case. Therefore, on the subspace invariant under the action of the group C_{Nv}, $L \geq 0$ and 0 is now its eigenvalue. The latter implies that

$$E_R(a) - E_R(0) > 0 \qquad \text{(A.4.2)}$$

for any odd N and for sufficiently small a.

APPENDIX 5

Large-N Asymptotic Forms

In this appendix, we find the asymptotic behavior of the energy of the circular asymptotically forceless configurations, i.e., the ones with $\nabla H(\underline{a}) = 0$, for large values of N. More precisely, the configurations we consider consist of N 1-vortices equally spaced on a circle of radius a and with the center at the origin and one $(-k)$ vortex at the center. We recall that the condition $\nabla H(\underline{a}) = 0$ is equivalent to the relation $k = -(N-1)/2$. We assume in addition that N is odd and $a \gg N$.

According to Eq. (5.10) and because

$$\sin\frac{\pi k}{N} = \sin\frac{\pi(N-k)}{N},$$

the energy of the above configuration is

$$E_R(\underline{a}) = \pi\left(\frac{N+1}{2}\right)^2 \ln R - \pi\left(\frac{N-1}{2}\right)^2 \ln\left(\frac{N-1}{2}\right) + Nc(1) - 2\pi N\sum_{k=1}^{\frac{N-1}{2}} \ln\left(2\sin\frac{\pi k}{N}\right), \qquad \text{(A.5.1)}$$

where we use the notation $E_R(a) = E_R(\underline{a})$. For $a = 0$ (the "initial state"), the energy is given by Eq. (3.5),

$$E_R(0) = \pi\left(\frac{N+1}{2}\right)^2 \ln R - \pi\left(\frac{N+1}{2}\right)^2 \ln\left(\frac{N+1}{2}\right). \qquad \text{(A.5.2)}$$

To calculate the sum in Eq. (A.5.1), we use the Euler expansion

$$\sum_{k=M}^{L} f(k) = \int_{M-\frac{1}{2}}^{L+\frac{1}{2}} f(x)dx - \frac{1}{24}\left(f'\left(L+\frac{1}{2}\right) - f'\left(M-\frac{1}{2}\right)\right) \quad \text{(A.5.3)}$$

and

$$\int_0^{\pi/2} \ln(2\sin x)dx = 0, \qquad \sum_{k=1}^{M} \ln k = \ln\Gamma(M+1), \quad \text{(A.5.4)}$$

where $\Gamma(x)$ is the Euler gamma function,

$$\sum_{k=1}^{\frac{N-1}{2}} \ln\left(2\sin\frac{\pi k}{N}\right) = \sum_{k=1}^{M} \ln\left(2\sin\left(\frac{\pi k}{N}\right)\right) - \frac{N}{\pi}\int_0^{\frac{\pi M}{N}} dz \ln(2\sin Z) + \sum_{k=M+1}^{\frac{N-1}{2}} \ln\left(2\sin\left(\frac{\pi k}{N}\right)\right)$$

$$-\frac{N}{\pi}\int_{\frac{\pi M}{N}}^{\pi/2} dz \ln(2\sin Z) = \sum_{k=1}^{M} \ln\left(\frac{2\pi k}{N}\right) - M\left(\ln\left(\frac{2\pi M}{N}\right) - 1\right) - \frac{1}{2}\ln\left(\frac{2\pi M}{N}\right) + \frac{1}{24M},$$

where $1 \ll M \ll N$. For $N \gg 1$, this yields

$$\sum_{k=1}^{\frac{N-1}{2}} \ln\left(2\sin\frac{\pi k}{N}\right) = \frac{1}{2}\ln N \quad \text{(A.5.5)}$$

modulo terms $O(1)$ in N. As a result, we have the energy difference

$$E_R(a) - E_R(0) = N\left[c(1) + \left(\frac{1}{2} - \ln 2\right)\pi\right] = 0.183\pi N. \quad \text{(A.5.6)}$$

Thus, for ($N \gg 1$) vortices placed equidistantly on a circle of radius $a \gg N$, the energy is greater than the effective energy of a single N vortex.

REFERENCES

1. F. Bethuel, H. Brezis, and F. Hélein, Calc. Variat. PDE **1**, 123 (1993).
2. F. Bethuel, H. Brezis, and F. Hélein, *Ginzburg–Landau Vortices* (Birkhäuser, Basel, 1994).
3. F. Pacard and T. Rivière, *Linear and Nonlinear Aspects of Vortices: The Ginzburg–Landau Model* (Birkhäuser, Boston, 2000), Progress in Nonlinear Differential Equations and Their Applications, Vol. 39.
4. Yu. N. Ovchinnikov and I. M. Sigal, Physica A (Amsterdam) **261**, 143 (1998).
5. T. Rivière, Manuscr. Math. **108**, 217 (2002).
6. F. Dalfovo, S. Giogini, L. P. Pitaevskii, and S. Stringari, Rev. Mod. Phys. **71**, 463 (1999).
7. A. L. Fetter and A. A. Svidzinsky, cond-mat/0102003 (2001).
8. V. L. Ginzburg and L. Pitaevskii, Sov. Phys. JETP **7**, 585 (1958).
9. P. Hagan, SIAM J. Appl. Math. **42**, 762 (1982).
10. M. Hervé and R. Hervé, Ann. Inst. Henri Poincaré Anal. Nonlin. **11**, 427 (1994).
11. Y. Chen, C. Elliot, and T. Qui, Proc. R. Soc. Edinburgh, Sect. A **124**, 1068 (1994).
12. P. Fife and L. A. Peletier, Quart. Appl. Math. **54**, 85 (1996).
13. Yu. N. Ovchinnikov and I. M. Sigal, CRM Proc. Res. Notes **12**, 199 (1997).
14. S. Chanillo and M. K.-H. Kiessling, C. R. Acad. Sci. Paris, Ser. I: Math. **321**, 1023 (1995).
15. P. Mironescu, J. Funct. Anal. **130**, 334 (1995).
16. P. Mironescu, C. R. Acad. Sci. Paris, Ser. I: Math. **323**, 593 (1996).
17. E. Lieb and M. Loss, Math. Res. Lett. **1**, 701 (1994).
18. Yu. N. Ovchinnikov and I. M. Sigal, Eur. J. Appl. Math. **13**, 153 (2002).
19. I. Shafrir, C. R. Acad. Sci. Paris, Ser. I: Math. **318**, 327 (1994).
20. Yu. N. Ovchinnikov and I. M. Sigal, Rev. Math. Phys. **12**, 287 (2000).
21. Yu. N. Ovchinnikov and I. M. Sigal, Nonlinearity **11**, 1277 (1998).
22. H. Brezis, F. Merle, and T. Rivière, Arch. Ration. Mech. Anal. **126**, 35 (1994).

Journal of Experimental and Theoretical Physics, Vol. 99, No. 5, 2004, pp. 1108–1116.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 5, 2004, pp. 1267–1276.

SOLIDS
Electronic Properties

Static and Dynamic Charge Inhomogeneity and Crystal-Field Fluctuations for 4*f* Ions in High-T_c Cuprates¶

A. S. Moskvin*, Yu. D. Panov, and N. V. Mel'nikova
Ural State University, Yekaterinburg, 620083 Russia
**e-mail: alexandr.moskvin@usu.ru*
Received June 16, 2004

Abstract—The main mechanism in homogeneous broadening and relaxation of crystal-field excitations for R-ions in cuprates is believed to be provided by the fluctuations of crystalline electric field induced by a static and dynamic charge inhomogeneity generic to doped cuprates. Such an inhomogeneity is assumed to be the result of topological phase separation. We address the generalized granular model as one of the model scenarios for describing the static and dynamic charge inhomogeneity in cuprates. The charge subsystem is believed to be similar to that of a Wigner crystal with melting transition and phononlike positional excitation modes. We consider a simple model of charge inhomogeneity that allows us to elucidate the main universal features of the density of CF states and the respective inhomogeneous broadening. The formal description of R-ion relaxation mainly coincides with that of the recently suggested magnetoelastic mechanism by Lovesey and Staub. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Inelastic neutron scattering (INS) spectroscopy is a powerful tool that makes it possible to unambiguously determine the Stark multiplet structure and crystal-field (CF) potential in rare-earth (R) based high-T_c superconducting materials such as $Y_{1-x}R_xBa_2Cu_3O_{6+y}$ [1, 2]. This technique provides detailed information on the electronic ground state of the R-ions, which is important for understanding the thermodynamic magnetic properties and the observed coexistence between superconductivity and long-range magnetic ordering of the R-ion sublattice at low temperatures. Moreover, INS spectroscopy may be effectively used for quantitative monitoring of the decay of the antiferromagnetic state of the parent compound and the evolution of the superconducting state upon doping, because the linewidths of CF transitions are believed to directly probe the electronic susceptibility. The relaxation behavior appears to be extremely dependent on the energy at which the susceptibility is probed. The CF INS spectroscopy is widely used to reveal the opening of an electronic gap in the normal state of underdoped superconductors [1] and to examine its anisotropy [3, 4]. Recently, Ho^{3+} CF-INS spectroscopy was used to investigate the oxygen and copper isotope effects on the pseudogap in Ho-124 and $(LaHoSr)_2CuO_4$ high-temperature superconductors [5, 6]. However, the mechanism of the relaxation of R-ions in cuprates has become the issue of heated debates [7, 8] that question the current interpretation of information detected by INS spectroscopy.

In the normal state, the excited CF levels of an R-ion interact with phonons, spin fluctuations, and charge carriers. These interactions limit the lifetime of the excitation; thus, the observed CF transitions exhibit line broadening. Similarly to the case of conventional Fermi-liquid metals, interaction with charge carriers is considered the predominant relaxation mechanism in cuprates. This interaction is usually assumed to be an isotropic exchange coupling with the effective spin Hamiltonian $H_{ex} = -2I(g_J - 1)(\mathbf{s} \cdot \mathbf{J})$, where I is an exchange integral that should be nearly independent of the particular R-ion under consideration; g_J is the Lande factor; **s** is the spin moment of a charge carrier; and **J** is the total momentum of the R-ion. Such a scenario seems to be rather natural if the predominant spin channel of neutron scattering is taken into account. The detailed theory of the respective relaxation mechanism was developed by Becker, Fulde, and Keller (BFK-model) [9]. The corresponding intrinsic linewidth appears to increase almost linearly with temperature ($\Gamma(T) \propto \rho^2 T$) according to the well-known Korringa law [10]. Here, ρ is the coupling constant, $\rho = I(g_J - 1)N(E_F)$, where $N(E_F)$ is the density of states (DOS) at the Fermi level. The deviation from a linear temperature dependence at low temperatures has been usually interpreted in terms of the opening of a (pseudo)gap and the associated reduction in damping. Fitting the high-temperature linewidth data in the framework of the simple or modified Korringa law, one

¶ This article was submitted by authors in English.

obtains coupling constant values that typically vary from 0.003 to 0.006 [1, 3–6].

We emphasize that the spin channel of relaxation directly implies the relevance of the Fermi-liquid scenario for cuprates, with many signatures of non-Fermi-liquid behavior ignored. However, the spin-exchange model has a number of visible inconsistencies, firstly, as concerns the magnitude of the coupling constant. Indeed, a linear temperature dependence of the relaxation time above T_c observed in EPR studies of S-ion Gd^{3+} in $YBa_2Cu_3O_7$ after Korringa fitting yields a magnitude of the exchange integral of $I \approx 3 \times 10^{-4}$ eV [12], which directly points to unrealistically large values of the spin coupling constants ρ found in all INS experiments on CF transitions. Certain problems exist with the Lande factor scaling approximately as $(g_J - 1)$. In studying the system $Y_{1-x}R_xBa_2Cu_3O_{6+y}$ (R = Er, Ho, Tm), Mukherjee *et al.* [11] found $|\rho(Tm)/\rho(Ho)| \approx 2$ instead of the theoretically expected value $(g_{Tm} - 1)/(g_{Ho} - 1) = 2/3$, and $|\rho(Tm)/\rho(Er)| \approx 4.5$, instead of the expected $(g_{Tm} - 1)/(g_{Er} - 1) = 5/6$. This clear disagreement evidences against the exchange mechanism. The spin-exchange scenario fails to explain the "strange" doping dependence of Tm^{3+} relaxation in Tm-123 [13] and Nd^{3+} relaxation in $(LaSrNd)_2CuO_4$ [14].

Finally, Staub *et al.* [15] found that the Lorentzian linewidth of the quasi-elastic neutron scattering for Tb^{3+} in $YBa_2Cu_3O_7$ can be properly described by the simple $(\exp(\Delta/k_BT) - 1)^{-1}$ law typical of Orbach processes governed by lattice vibrations. They have shown that such an interpretation also describes the results obtained earlier for Ho^{3+} and Tm^{3+}. They conclude that interactions with charge carriers are negligible and that interactions with lattice vibrations are responsible for the relaxation behavior of the 4*f* electrons in cuprates. Therefore, the INS results that claim to probe the superconducting gap or the pseudo-gap should be reexamined in terms of Orbach processes. A similar conclusion was drawn in [14] for Nd^{3+} relaxation in $(LaSrNd)_2CuO_4$. Lovesey and Staub [16] have shown that the dynamic properties of the lanthanide ions (Tb^{3+}, Ho^{3+}, and Tm^{3+}) are adequately described by a simple three-state model, not unlike the one introduced by Orbach for interpreting electron paramagnetic resonance signals from a lanthanide ion in dilute concentration in a salt. The cross section for inelastic scattering of neutrons by the lanthanide ion is derived by constructing a pseudospin $S = 1$ model and treating the magnetoelastic interaction as a perturbation of the three CF states. The scattering of neutrons is thus a quasielastic process and the relaxation rate is proportional to $(\exp(\Delta/k_BT) - 1)^{-1}$, where Δ is the energy of the intermediate CF state at which the density of phonon states is probed. However, this very attractive scenario also faces some visible difficulties in explaining, for instance, the unusual nonmonotonic temperature dependences and too large oxygen isotope effect in the INS spectra of Ho-124 and Ho-214 systems [5, 6], and certain doping dependences in Y-123 and Nd-214 systems [14]. The origin of the anomalously large low-temperature inhomogeneous broadening remains unclear. The magnetoelastic mechanism yields very small magnitudes of $\Gamma(T = 0)$, one or two orders smaller than that found in experiment.

In comparing the two mechanisms, we emphasize a difference in them that seems to be of primary importance: the spin-channel mechanism takes the fluctuations of the effective magnetic field on R-ions into account, while the phonon (magnetoelastic) mechanism deals with fluctuations of the electric field. Moreover, the conventional spin-channel mechanism actually probes spin fluctuations rather than charge fluctuations, although its contribution to the linewidth $\Gamma(T) \propto (IN(E_F))^2$ is believed to strongly depend on the carrier density. However, this relationship is derived in the framework of the Fermi-liquid scenario and should be modified if one addresses the typical antiferromagnetic insulating state. Interestingly, in [8, 15, 16], the phonon (magnetoelastic) mechanism is addressed as an alternative to charge fluctuations. As an example, the authors point to insulating materials in which "…the density of carriers is essentially zero…" [8], which forbids the charge fluctuation channel of relaxation.

We emphasize that both groups of researchers have underestimated the role of the conventional spinless charge fluctuation channel. Indeed, the CF Hamiltonian for an R-ion in cuprates can be written in its standard form as

$$H_{CF} = \sum_{k = 2, 4, 6} \sum_{-k \le q \le k} B^*_{kq} \hat{O}^q_k,$$

where $\hat{O}^q_k$ are Stevens equivalent operators; $B_{kq} = b_{kq}\langle r^k \rangle \gamma_k$, where b_{kq} are CF parameters; $\gamma_2 = \alpha$, $\gamma_4 = \beta$, $\gamma_6 = \gamma$ (α, β, γ are Stevens parameters); and

$$b_{kq} = \langle b_{kq} \rangle + \Delta b_{kq},$$

which may be expressed within the well-known point-charge model as

$$\Delta b_{kq} = \sum_i \frac{qC^k_q(\mathbf{R}_i)}{R_i^{k+1}} (\hat{n}_i(t) - \langle n_i \rangle),$$

where C^k_q is the tensorial spherical harmonics and $\hat{n}_i(t)$ is the charge number operator. Conventional metals are characterized by very short-time charge dynamics, which makes it possible to neglect the contribution of charge fluctuations to the inhomogeneous broadening and relaxation of R-ions in the low-energy range of CF energies, and to consider a mean homogeneous charge distribution. An altogether different picture emerges in

the case of cuprates where we are dealing with various manifestations of static and dynamic charge inhomogeneity (see, e.g., [17, 18] and references therein). Moreover, the INS spectroscopy of CF excitations itself yields an impressive picture of charge inhomogeneity in the 123 system [1, 2], where it was found that the observed CF spectra separate into different local components whose spectral weights distinctly depend on the doping level; i.e., there is clear experimental evidence for cluster formation. The onset of superconductivity can be shown to result from percolation, which means that the superconductivity is a property of inhomogeneous materials. It seems probable that the dynamic rearrangement of the charge system at temperatures above T_c somehow affects both the inhomogeneous broadening of CF transitions and R-ion relaxation.

2. CHARGE INHOMOGENEITY IN CUPRATES: TOPOLOGICAL PHASE SEPARATION

At present, the stripe model of inhomogeneity [18] is most popular in cuprate physics. It is worth noting that this model is based on the more universal idea of topological phase separation, with the doped particles assumed to be localized inside the domain walls of a bare phase.

Below, we address one of the topological phase separation scenarios that may be termed a generalized granular model for doped cuprates. We assume that the CuO_2 layers in parent cuprates may gradually lose their stability under electron–hole doping, while a new self-organized multigranular 2D phase becomes stable.

The new scenario implies that the unconventional phase state evolves from the parent insulating cuprate as a result of self-trapping of charge transfer excitons (CT) accompanied by a self-consistent lattice polarization and the appearance of the "negative-U" effect. Parent insulating cuprates appear to be unstable with regard to self-trapping of low-energy one- and two-center CT excitons [19, 20], with the nucleation of electron–hole droplets being actually a system of coupled electron CuO_4^{7-} and hole CuO_4^{5-} centers glued in the lattice due to strong electron–lattice polarization effects. Such a system can be regarded as an electron–hole Bose liquid described by the generalized Bose–Hubbard Hamiltonian. Doping, or deviation from half-filling in an electron–hole Bose liquid is accompanied by formation of multicenter topological defects such as charge-order (CO) bubble domain(s) with Bose superfluid (BS) and extra bosons localized in domain wall(s), or a topological CO + BS phase separation, rather than a uniform mixed CO + BS supersolid phase [21, 22]. Such a situation partly resembles that of granular superconductivity.

The most probable possibility is that every micrograin accumulates one or two particles. Then the number of such entities in a multigranular texture nucleated with doping has to depend on the doping in a nearly linear way. Generally speaking, each individual micrograin may be characterized by its position, nanoscale size, and the orientation of the U(1) degree of freedom. In contrast to uniform states, the phase of the superfluid order parameter for a micrograin is assumed to be unordered. The granular structure must be considered largely dynamic in nature.

In the long-wavelength limit, off-diagonal ordering can be described by an effective Hamiltonian in terms of the U(1) (phase) degree of freedom associated with each micrograin. Such a Hamiltonian contains a repulsive, long-range Coulomb part and a short-range contribution related to the phase degree of freedom. The latter term can be written in the form of a so-called Josephson coupling, standard for the *XY* model,

$$H_J = -\sum_{\langle i,j\rangle} J_{ij}\cos(\varphi_i - \varphi_j), \qquad (1)$$

where φ_i and φ_j are global phases for micrograins centered at the respective points i and j, and J_{ij} is the Josephson coupling parameter. The Josephson coupling gives rise to the long-range ordering of the phase of the superfluid order parameter in such a multicenter texture. Such a Hamiltonian represents a starting point for the analysis of disordered superconductors, granular superconductivity, and the insulator–superconductor transition with an $\langle i, j\rangle$ array of superconducting islands with phases φ_i, φ_j.

To account for the Coulomb interaction and allow for quantum corrections, we introduce the charging energy [23]

$$H_{ch} = -\frac{1}{2}q^2\sum_{i,j} n_i(C^{-1})_{ij}n_j$$

into the effective Hamiltonian, where n_i is the number operator for particles bound in the ith micrograin; it is quantum-mechanically conjugate to φ, $n_i = -i\partial/\partial\varphi_i$, $(C^{-1})_{ij}$ is the capacitance matrix, and q is the particle charge.

Such a system appears to reveal a tremendously rich quantum-critical structure [24, 25]. In the absence of disorder, the $T = 0$ phase diagram of the multigranular system implies either triangular or square crystalline arrangements with a possible melting transition to a liquid. We note that the analogy with the charged 2D Coulomb gas implies Wigner crystallization of the multigranular system with a Wigner crystal (WC) to the Wigner liquid melting transition. Naturally, additional degrees of freedom of the micrograin provide a richer physics of such lattices. For a system to be an insulator,

disorder is required that pins the multigranular system and also causes the crystalline order to have a finite correlation length. The traditional approach to Wigner crystallization implies the formation of a WC for densities lower than the critical density, when the Coulomb energy is greater than the kinetic energy. The effect of quantum fluctuations leads to a (quantum) melting of the solid at high densities or at a critical lattice spacing. The critical properties of a two-dimensional lattice without any internal degrees of freedom are successfully described by applying the BKT theory to dislocations and disclinations of the lattice. This description proceeds in two steps. The first step implies the transition to a liquid-crystal phase with a short-range translational order, and the second involves the transition to an isotropic liquid. In such a system, if the micrograin positions are fixed at all temperatures, the long-wavelength physics is described by an (anti)ferromagnetic *XY* model with an expectable BKT transition and a gapless *XY* spin-wave mode.

The low-temperature physics in a multigranular system is governed by an interplay of two BKT transitions, for the U(1) phase and the positional degrees of freedom, respectively [25]. Dislocations lead to a mismatch in the U(1) degree of freedom, which makes the dislocations bind fractional vortices and leads to a coupling of translational and phase excitations. The BKT temperatures either coincide (square lattice) or the melting temperature is higher (triangular lattice) [25].

Quantum fluctuations can substantially affect these results. Quantum melting can destroy the U(1) order at sufficiently low densities where the Josephson coupling becomes exponentially small. A similar situation is expected to occur in the vicinity of structural transitions in a multigranular crystal. With increasing micrograin density, the quantum effects result in a significant decrease in the melting temperature compared with the classical square-root dependence. The resulting melting temperature can reveal an oscillating behavior as a function of the particle density with zeros at critical (magic) densities associated with structural phase transitions.

In terms of our model, the positional order corresponds to an incommensurate charge density wave, while the U(1) order corresponds to superconductivity. In other words, we arrive at a subtle interplay between two orders. The superconducting state evolves from a charge order with $T_C \le T_m$, where T_m is the temperature of the melting transition, which could be termed the temperature of the opening of the insulating gap.

The normal modes of a dilute multigranular system include the pseudo-spin waves propagating in between the micrograins; the positional fluctuations, or quasiphonon modes, which are gapless in a pure system but are gapped when the lattice is pinned; and, finally, fluctuations in the U(1) order parameter.

The orientational fluctuations of the multigranular system are governed by the gapless *XY* model [24]. The relevant model description is most familiar as an effective theory of the Josephson junction array. An important feature of the model is that it displays a quantum critical point.

The low-energy collective excitations of a multigranular liquid includes the usual longitudinal acoustic phononlike branch. The liquid crystal phases differ from the isotropic liquid in that they have massive topological excitations, i.e., disclinations. We note that liquids do not support transverse modes, and these could survive in a liquid state only as overdamped modes. It is therefore reasonable to assume that solidification of the bubble lattice is accompanied by stabilization of transverse phononlike modes with their sharpening below the melting transition. In other words, the instability of transverse phononlike modes signals the onset of melting. The phononlike modes in the bubble crystal have much in common with the usual phonon modes, but because of their electronic nature they can hardly be detected by inelastic neutron scattering.

A generic property of the positionally ordered bubble configuration is the sliding mode, which is usually pinned by the disorder. The depinning of sliding mode(s) can be detected in a low-frequency and low-temperature optical response.

We note that as regards CF fluctuations, there is no principal difference between the contributions of real phonon modes and quasiphonon modes of a multigranular system. Moreover, it is worth noting that the charge inhomogeneity in a multigranular system is prone to be closely coupled with lattice structural distortions. However, stabilization of transverse phononlike modes in multigranular system that accompanies its solidification at the temperatures above T_c may strongly affect the CF relaxation due to a mechanism identical to the magnetoelastic mechanism proposed by Lovesey and Staub. In a sense, such a conclusion reconciles the "old" spin-fluctuation [1, 2] and the "new" magnetoelastic phonon [15, 16] approaches to INS spectroscopy of cuprates with R-ions.

Above, we addressed a simplified model of "rigid" bubbles and neglected any possible internal or conformational degree of freedom. However, the bubble can actually be characterized by a subtle interplay of orbital degrees of freedom with a pseudo-Jahn–Teller effect. In other words, we may anticipate a set of different conformational states of the bubble.

3. CHARGE INHOMOGENEITY AND INHOMOGENEOUS LINE BROADENING OF CF TRANSITIONS IN CUPRATES

Neutron spectroscopy involves energies of several meV and is therefore susceptible to dynamical effects of the order of 10^{-13} s. All slower processes contribute to inhomogeneous broadening of CF transitions. The

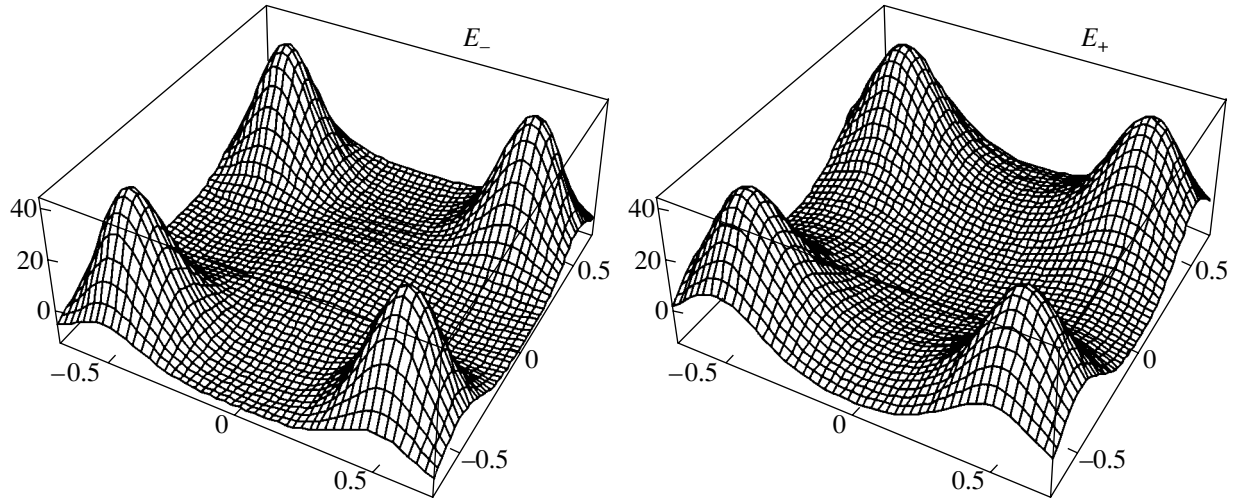


Fig. 1. Energy surfaces $E_{\pm}(x, y)$ for a model doublet ($z = 0.3$).

effects of inhomogeneous line broadening are clearly seen for CF excitation in $Ho_{1-x}Y_xBa_2Cu_3O_7$ with an energy near 0.5 meV [4]. Although the CF transition is between two singlets, it reveals an intrinsic multiple-peak structure at low temperatures, comprising a dominant central peak with shoulders on each side, and a tail on the higher energy side. The key assumption on which the traditional analysis is based is that the observed line shape arises from R-ions distributed in slightly different local environments, but subject to the same relaxation processes. In other words, one assumes that the observed line shape reflects a temperature-independent inhomogeneous broadening and a universal temperature-dependent relaxation mechanism. The spectrum measured at a particular temperature is then given by the convolution of a broadening function characteristic of that temperature and the residual line shape at absolute zero. Hence, the relaxation is assumed to be described by a single broadening function whose position, width, and amplitude depend on temperature. However, this approach fails to explain an unusual low-temperature line shape of the CF transition with an unexpectedly large ($\Gamma_0 \approx 0.2$ meV) residual linewidth of the central peak. The low-temperature experimental spectra are likely to reveal some sort of continuous distribution of CFs, rather than a simple superposition of only three components whose spectral weights distinctly depend on the doping level, as was assumed in [1, 2].

In the continuum approximation, the resultant CF transition line shape in the static case corresponds to the density of the local CF distribution convoluted with the individual line shape. The density of the local CF distribution has a number of universal features typical of a rather wide range of inhomogeneous 2D potentials. Under certain conditions, we can easily predict the character and number of such peculiarities in the complex structure of the CF transition and even the line shape itself.

3.1. Simple Point Charge Model of Electron Inhomogeneity

To make our consideration more quantitative, we consider a simple model of charge inhomogeneity centers organized into an incommensurate square superlattice in the CuO_2 plane with a parameter a, and hypotetical R-ions with the momentum $J = 1$ positioned above (under) the CuO_2 plane (the z-coordinate in units of a: $z = Z/a$), as in the R-123 structure, and having the $M = 0$ ground singlet state. In the approximation of a strong tetragonal CF component, $|B_0^2| \gg |B_2^2|$, the energies of two excited states with $|M| = 1$ are shifted by

$$\Delta E_{\pm}(x, y) \propto \left[\sqrt{\frac{3}{2}}\Delta B_0^2(x, y) \pm \left|\Delta B_2^2(x, y)\right|\right].$$

Hence, we can introduce two energy surfaces $E_{\pm}(x, y)$, where x and y are the plane coordinates of a R-ion. The surfaces osculate at points with tetragonal symmetry. The point-charge model for the CF parameters $B^0_{0,\pm 2}$ allows us to easily compute these surfaces. For $z = 0.3$, they are shown in Fig. 1. At first sight, these surfaces differ insignificantly, but the isoenergetic curves reveal a distinction. In Fig. 2, we marked different singular points and some isoenergetic curves for both modes $E_{\pm}(x, y)$. Four points of type M at the corners of the square cell with the tetragonal symmetry correspond to sharp maxima for both $E_-(x, y)$ and $E_+(x, y)$. The Γ point at the center with the tetragonal local symmetry corresponds to a smooth local maximum of the energy $E_-(x, y)$ and a minimum of the energy $E_+(x, y)$. Four

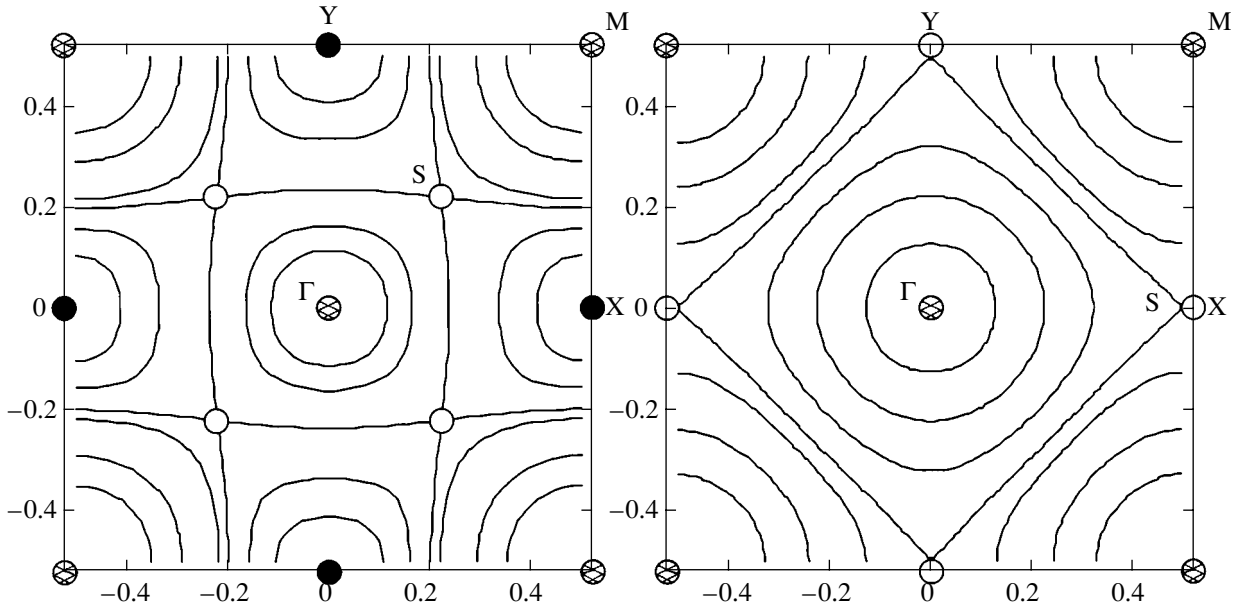


Fig. 2. Specific points in an elementary cell of incommensurate superstructure. Singular points and some isoenergetic curves for the $E_-(x, y)$ (left-hand side) and $E_+(x, y)$ (right-hand side) branches of a model doublet (see text for details).

minima of $E_-(x, y)$ are situated at the points $X(Y)$ on the boundaries where the $E_+(x, y)$ surface has saddle points. The saddle points of the $E_-(x, y)$ surface are situated inside the elementary cell. By varying the lattice separation, we may simulate the effect of varying the concentration of charge inhomogeneity centers. The energy surfaces $E_\pm(x, y)$ can be described by the DOS defined as $\rho(E) \propto [dE/dS]^{-1}$, where $S(E)$ is the area of the cross section $E(x, t) = E = \text{const}$. The R-ions are assumed to be uniformly distributed in the x, y plane, and their number is proportional to the cross-sectional area: $dN_R(E) = \rho(E)dE \propto dS(E)$. The DOS has several singularities associated with extremal points (minima, maxima, and saddle points). The saddle points are of primary importance because they are known to yield a logarithmic divergence of the DOS in two-dimensional systems. Near the minima and maxima, the $E(S)$ dependence can be approximated as $E(S) \approx E(0) + aS^n$. It is clear that for $n \le 1$, the DOS is finite at the extremum point, while for $n > 1$, it diverges at the extremum point. It is worth noting that sharp extrema with small n correspond to a small DOS. It is interesting to note the strong resemblance of the procedure to that of the conventional two-dimensional band model, where one deals with a **k**-momentum space.

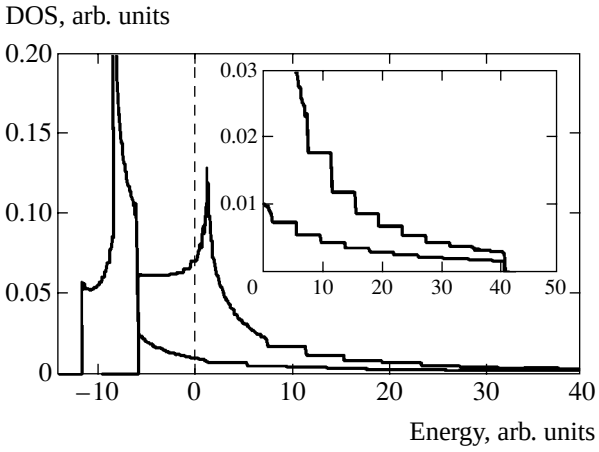


Fig. 3. Results of a numerical calculation of the density of states for the lower and upper branches $E_\pm(x, y)$ of a model doublet with $z = 0.3$.

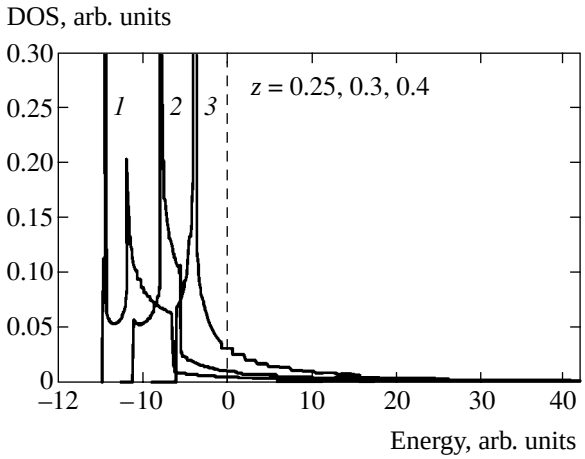


Fig. 4. Results of a numerical calculation of the density of states for the lower branch $E_-(x, y)$ of a model doublet for different values of the z parameter: $z = 0.25$ (*1*); 0.3 (*2*); 0.4 (*3*).

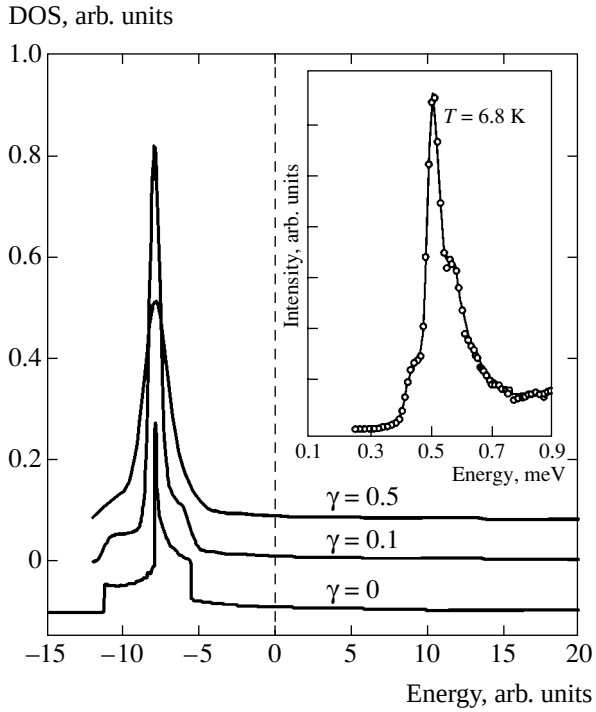


Fig. 5. Inhomogeneous broadening of the model CF transition. The results of a Gaussian convolution of the calculated DOS. The inset shows the experimental inelastic neutron scattering spectra of the ground state to the first excited state CF transition of Ho^{3+} in $Ho_{0.1}Y_{0.9}Ba_2Cu_3O_7$ in the energy range 0.1 to 0.9 meV [4].

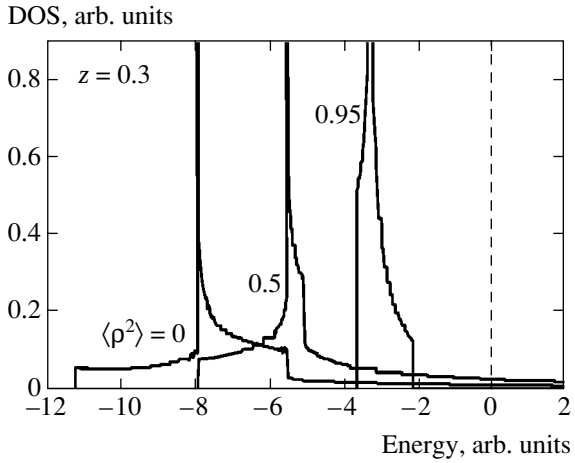


Fig. 6. Results of a numerical calculation of the density of states for the lower branch $E_-(x, y)$ of a model doublet given different values of the $\langle\rho^2\rangle$ parameter: 0; 0.5; 0.95.

The results of numerical calculation of the DOS for both the low-energy $|-\rangle$ and high-energy $|+\rangle$ modes are shown in Fig. 3.[1] The inset to Fig. 3 shows the fine structure of the DOS near the maximum of the energies $E_\pm$. The dotted line shows the energy position of the $|\pm\rangle$ doublet failing the inhomogeneity potential. We note that both DOSs reveal features typical of two-dimensional systems. This figure yields a nice illustration of the effects of charge incommensurability, in particular, the splitting effect resulting from local breaking of the tetragonal symmetry. It is worth noting that our model DOS obeys the a^{-3} scaling law.

Figure 4 shows the effect of varying the distance z of the R-ion from the CuO_2 plane. We see the change in the DOS shape with the expected narrowing and blue shift for larger z.

Our model approach yields a simple illustration of the concentration effects. Indeed, if we assume the generic square lattice for the inhomogeneity centers, we obtain a simple relation between the lattice parameter and concentration: $a^2 \propto 1/x$. Hence, given a fixed absolute magnitude of the Z parameter, we see that the dimensionless parameter $z = Z/a$ varies with the concentration of the charge inhomogeneity centers. In other words, Fig. 4 with additional a^{-3} scaling corrections yields an example of a change in the DOS with a rise in concentration. As expected, the rise in concentration results in a smoothing of the energy surfaces with a narrowing of the energy distribution and a sizeable shift of the main peak. Positional disorder due to conventional defects such as substituted ions, unconventional topological defects such as dislocations and disclinations (which are inherent for two-dimensional materials, however), and slow positional motion of bubbles result in an inhomogeneous broadening, which implies a weighted superposition of different energy surfaces $E(x, y)$. Such a broadening can be easily taken into account if we simply assume the Gaussian distribution of different $E(x, y)$ values near a mean value $\langle E(x, y)\rangle$. An illustrative example of a Gaussian broadening is shown in Fig. 5, where we have included both raw numerical data and the results of a convolution with the Gaussian function with a half-width of $\gamma = 0.1, 0.5$.

Fast positional motion of the charge inhomogeneity centers results in averaging of the CF potential acting on the R-ion. The simplest model of such an averaging in the framework of the point-charge approximation for our square superlattice assumes the distribution of the point charges near mean positions **R** with a probability of $W(\rho) \propto e^{-\rho^2/\langle\rho^2\rangle}$, where ρ specifies the displacement from the mean position and $\langle\rho^2\rangle$ is a mean-square displacement. In general, the mean-square displacement is believed to be strongly anisotropic, with a predominant in-plane component. For simplicity, however, the numerical calculations of the energy surfaces $E_\pm$ were performed with an isotropic displacement. The averaged potential differs from the bare Coulomb potential in that it has smoothed and lower maxima, and hence the bounds of the DOS spectra shrink with a simultaneous shift in the center of gravity to higher energies (see Fig. 6). Interestingly, the shift in the center of DOS gravity is very sensitive to the z component of the

[1] The numerical calculations were performed by E. Zenkov.

charge displacement and in a sense can be used as its measure.

However, this is not the only effect of averaging. As the extremum regions of $E_{\pm}(r)$ become flatter and their areas extend, the contribution of a greater number of adjacent sites to the extrema becomes important, each of them coming with its own phase. This leads to the specific interference phenomena. In particular, as the dispersion $\langle\rho^2\rangle$ increases, the extremum points of $E_{\pm}$ reveal a clear tendency to splitting. For example, the maximum of the bare E_{-} surface at the Γ point splits into four maxima, which are shifted towards the four corresponding M points, while a shallow minimum appears at the Γ point. Thus, unexpectedly enough, averaging can result in some complication of the energy surfaces in general. The same effect is obtained alternatively by increasing the z parameter.

As the charge distribution in the CuO_2 plane becomes more uniform, the $E_{\pm}$ energy separation progressively vanishes, because the potential of a uniformly charged plane yields no electric field gradient. This may be achieved either by increasing z or by amplifying fluctuations of the in-plane sites. Then the energy spectrum of the model consists of the $M = 0$ singlet ground state and the $M = \pm 1$ doublet, and the DOS spectrum of the only excited state reduces to a δ-peak that resides at zero energy in the adopted units. The numerical results (see Figs. 4 and 6) confirm this conclusion.

To summarize, analysis of the real-space charge inhomogeneity makes it possible to approach the interpretation of the typical features of experimental spectra from a novel angle in terms of the "real-space DOS" singularities, which reflect some essential topological properties of the inhomogeneity-induced spatial distribution of the relevant physical parameters such as the CF for R-ions. The basic properties of these DOS singularities (their number, kinds, etc.) are rather stable against variations in the charge distribution and admit a simple classification scheme. For example, extremum points of the distribution correspond to jumps in the DOS spectra, while saddle points give rise to sharp divergences that manifest themselves as spectral peaks. The observed experimental spectrum should be regarded as a convolution of an individual line profile, the intensity factor, and the DOS function, together contributing to a complex resultant line shape. However, it is sometimes possible to discriminate between different sources of the spectral features. In particular, the proposed DOS mechanism should be addressed in the case of "extra lines," where the number of spectral features observed exceeds that predicted from symmetry considerations.

3.2. Implications for CF Transitions in Cuprates

In the inset to Fig. 5, we present the low-temperature CF excitation spectrum in $Ho_xY_{1-x}Ba_2Cu_3O_7$ ($x = 0.1$) with the energy near 0.5 meV [4]. It may be concluded that the spectrum exhibits all the features found in our model simulation, and hence precisely the inhomogeneous broadening governs the line shape. The central peak seems to reflect the contribution of saddle-point R-ions, while the left-hand and right-hand shoulders are associated with R-ions exposed to extremal CF magnitudes.

Bubbles in a crystal or liquid state of the cuprate participate in both slow and fast motion, and we should therefore expect a rather complicated interplay of inhomogeneous broadening and averaging/narrowing, which can strongly depend on temperature. Simple classical considerations imply the T-linear high-temperature dependence of both $\langle\rho^2\rangle$ and the concentration of the topological defects in the bubble system such as dislocations and disclinations. However, the low-temperature behavior of $\langle\rho^2\rangle$ is governed mainly by quantum effects. It is worth noting that the contribution of topological defects changes when crossing the BKT transition temperature, which is accompanied by binding/unbinding of topological defects and the change in the behavior of correlation functions. At first glance, the rise in the temperature has to suppress the inhomogeneous broadening due to a faster motion of bubbles. However, we are actually dealing with two competing T-dependent effects: the rise of the concentration of topological defects on the one hand and the rise of their mean velocity on the other.

Slow conformational motion can be described in terms of a finite diffusion, resulting in a linear-in-T dependence of the respective inhomogeneous broadening. Actually, we deal with a combined effect of different sources of static and dynamic factors governing the line shape of CF transitions. Its separation requires both further experimental information and a refinement of theoretical models.

4. CONCLUSIONS

We have argued that the main mechanism of inhomogeneous broadening and relaxation of CF excitations for R-ions in cuprates can be provided by fluctuations of the crystalline electric field induced by a static and dynamic charge inhomogeneity generic to doped cuprates. Such an inhomogeneity is assumed to be the result of topological phase separation. We have considered the generalized granular model as one of the model scenarios for describing the static and dynamic charge inhomogeneity in cuprates. The charge subsystem is believed to be similar to that of a Wigner crystal with the melting transition and phononlike positional excitation modes. We have considered a simple model of charge inhomogeneity organized into an incommensurate square superlattice; this model makes it possible to elucidate the main universal features of the real-space density of CF states. It is worth noting that both static and dynamic effects are considered on an equal footing.

We see that the studies of line narrowing for CF transitions for 4*f* ions in high-T_c cuprates provides an informative tool to investigate the charge rearrangement that accompanies the onset of high-T_c superconductivity. Our model approach based on the analysis of the real-space DOS can be easily generalized to study other manifestations of the electron inhomogeneity in cuprates such as inhomogeneous broadening of NMR–NQR signals.

ACKNOWLEDGMENTS

This paper was supported in part by INTAS (grant no. 01-0654), the CRDF (grant no. REC-005), RME (grant nos. 02.-3.4-392 and UR.01.01.062), and the Russian Foundation for Basic Research (project no. 04-02-96077). A. S. M. has benefited from stimulating discussions with A. T. Boothroyd, A. Mirmelstein, and J. Mesot.

REFERENCES

1. J. Mesot and A. Furrer, J. Supercond. **10**, 623 (1997).
2. J. Mesot and A. Furrer, in *Neutron Scattering in Layered Copper-Oxide Superconductors*, Ed. by A. Furrer (Kluwer, Dordrecht, 1998), p. 335.
3. J. Mesot, G. Böttger, H. Mutka, and A. Furrer, Europhys. Lett. **44**, 498 (1998).
4. A. T. Boothroyd, A. Mukherjee, and A. P. Murani, Phys. Rev. Lett. **77**, 1600 (1996).
5. D. Rubio Temprano, J. Mesot, S. Janssen, *et al.*, Phys. Rev. Lett. **84**, 1990 (2000).
6. D. Rubio Temprano, K. Conder, A. Furrer, *et al.*, Phys. Rev. B **66**, 184506 (2002).
7. A. T. Boothroyd, Phys. Rev. B **64**, 066501 (2001).
8. S. W. Lovesey and U. Staub, Phys. Rev. B **64**, 066502 (2001).
9. K. W. Becker, P. Fulde, and J. Keller, Z. Phys. B **28**, 9 (1977).
10. J. Korringa, Physica C (Amsterdam) **16**, 601 (1950).
11. A. Mukherjee, A. T. Boothroyd, D. McK. Paul, *et al.*, Phys. Rev. B **49**, 13089 (1994).
12. D. Shaltiel, C. Noble, J. Pilbrow, *et al.*, Phys. Rev. B **53**, 12430 (1996).
13. R. Osborn and E. A. Goremychkin, Physica C (Amsterdam) **185**–**189**, 1179 (1991).
14. M. Roepke, E. Holland-Moritz, B. Büchner, *et al.*, Phys. Rev. B **60**, 9793 (1999).
15. U. Staub, M. Gutmann, F. Fauth, and W. Kagunya, J. Phys.: Condens. Matter **11**, L59 (1999).
16. S. W. Lovesey and U. Staub, Phys. Rev. B **61**, 9130 (2000).
17. J. Burgy, M. Mayr, V. Martin-Mayor, *et al.*, Phys. Rev. Lett. **87**, 277202 (2001).
18. E. W. Carlson, V. J. Emery, S. A. Kivelson, and D. Orgad, cond-mat/0206217.
19. A. S. Moskvin, R. Neudert, M. Knupfer, *et al.*, Phys. Rev. B **65**, 180512(R) (2002).
20. A. S. Moskvin, J. Málek, M. Knupfer, *et al.*, Phys. Rev. Lett. **91**, 037001 (2003).
21. A. S. Moskvin, I. G. Bostrem, and A. S. Ovchinnikov, JETP Lett. **78**, 772 (2003).
22. A. S. Moskvin, Phys. Rev. B **69**, 214505 (2004).
23. S. A. Kivelson and B. Z. Spivak, Phys. Rev. B **45**, 10490 (1992).
24. A. G. Green, Phys. Rev. B **61**, R16299 (2000).
25. C. Timm, S. M. Girvin, and H. A. Fertig, Phys. Rev. B **58**, 10634 (1998).

Journal of Experimental and Theoretical Physics, Vol. 99, No. 5, 2004, pp. 891–897.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 5, 2004, pp. 1027–1033.

GRAVITATION, ASTROPHYSICS

A Macroscopic System of Einstein–Maxwell Equations for a System of Interacting Particles with Different Masses

A. V. Zakharov and R. K. Mukharlyamov

Kazan State University, ul. Lenina 19, Kazan, 420008 Tatarstan, Russia

e-mail: Alexei.Zakharov@ksu.ru; Ruslan.Muharlyamov@ksu.ru

Received December 25, 2003

Abstract—A macroscopic system of Einstein–Maxwell equations for systems of particles with different masses is derived up to the second order in the interaction. The dominant type of interaction in this system are electromagnetic interactions between particles (for example, a radiation-dominated cosmological plasma in the expanding universe before the moment of recombination). The results of [1], which can only be applied to systems of interacting particles with equal masses, are generalized. *© 2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

It is well known that the Maxwell equations for continua can be obtained from the microscopic Maxwell–Lorentz equations by statistical averaging over ensembles [2]. It is natural to assume that the Einstein equations (or their generalizations) in a medium can also be obtained by averaging the microscopic Einstein field equations (the Einstein equations whose right-hand side contains a sum of energy–momentum tensors of individual particles). However, since the left-hand side of the Einstein equations is nonlinear, the averaging of the microscopic Einstein equations presents a much more complicated problem compared with the derivation of the macroscopic Einstein equations in special theory of relativity.

The problem of constructing macroscopic Einstein equations was first posed by Shirokov [3]. In [4, 5], a method of averaging over an ensemble of microscopic Einstein equations was developed for a system of self-gravitating particles with equal masses. As a result, macroscopic Einstein equations were obtained for continua up to the second order in the gravitational interaction. In [6], we generalized the results of [4, 5] for a system of gravitationally interacting particles with different masses. The equations obtained differ from the classical Einstein equations by additional terms attributed to the interaction between particles. These terms are proportional to the cube of the Einstein constant and are expressed in terms of a two-particle correlation function of particles.

The paper [1] is devoted to the derivation (up to the second order in the interaction) of a macroscopic system of the Einstein–Maxwell equations for a relativistic plasma in which the dominant type of interaction are electromagnetic interactions.

The present paper generalizes the results of [1], which can only be applied to a system of interacting particles with equal masses.

It turns out that macroscopic Einstein equations for a relativistic plasma differ from the classical Einstein equations by additional terms on the left-hand side that are associated with both electromagnetic and gravitational interactions simultaneously. These terms represent a symmetric bivalent traceless tensor with zero divergence. In the present paper, we obtain explicit expressions for these terms as integrals, in the momentum space, of expressions containing one-particle distribution functions of interacting particles of the plasma.

It turns out that macroscopic Maxwell equations for a system of electromagnetically and gravitationally interacting particles are also different from the classical Maxwell equations. This difference manifests itself in the additional terms on the left-hand side of the Maxwell equations, which are attributed simultaneously to the effects of general gravity theory and the effects of interaction.

2. MICROSCOPIC EQUATIONS

Let us write out a microscopic system of Einstein and Maxwell equations as

$$\tilde{G}^{ij} = \chi\tilde{T}^{ij}_{(\mathrm{m})} + \chi\tilde{T}^{ij}_{(\mathrm{el})}, \tag{1}$$

$$\tilde{\nabla}_k\tilde{F}^{ik} = -\frac{4\pi}{c}\tilde{J}^i. \tag{2}$$

Here, $\tilde{G}^{ij}$ is the Einstein tensor of a Riemannian space with metric $\tilde{g}_{ij}$, $\tilde{T}^{ij}_{(\mathrm{m})}$ is a microscopic energy–momentum tensor of particles of the medium, $\chi = 8\pi k/c^4$ is the Einstein constant, k is the Newton gravitational con-

stant, c is the velocity of light, $\tilde{F}^{ik}$ is the tensor of electromagnetic field (the Maxwell tensor), $\tilde{J}^i$ is the microscopic current four-vector, and $\tilde{T}^{ij}_{(\mathrm{el})}$ is the energy–momentum tensor of the electromagnetic field. The operations of raising and lowering indices are performed by the metric tensor $\tilde{g}_{ij}$ and its inverse $\tilde{g}^{ij}$. $\tilde{\nabla}_k$ denotes a covariant derivative in a Riemannian space with metric $\tilde{g}_{ij}$.

The tensor $\tilde{T}^{ij}_{(\mathrm{el})}$ is expressed in terms of $\tilde{F}^{ik}$ as follows:

$$\tilde{T}^{ij}_{(\mathrm{el})} = \frac{1}{4\pi}\left(-\tilde{F}^i_{\cdot l}\tilde{F}^{lj} + \frac{1}{4}\tilde{g}^{ij}\tilde{F}_{lm}\tilde{F}^{lm}\right). \tag{3}$$

The tensors $\tilde{T}^{ij}_{(\mathrm{m})}$ and $\tilde{J}^i$ are expressed in terms of the Klimontovich random function [7] as

$$\tilde{T}^{ij}_{(\mathrm{m})} = \sum_a m_a c^2 \int \frac{d^4\tilde{p}_a}{\sqrt{-\tilde{g}}}\tilde{u}^i_a\tilde{u}^j_a\tilde{N}_a(q^i, \tilde{p}_i), \tag{4}$$

$$\tilde{J}^i = \sum_b e_b c \int \frac{d^4\tilde{p}}{\sqrt{-\tilde{g}}}\tilde{u}^i_b\tilde{N}_b(q, \tilde{p}). \tag{5}$$

Here, e_b is the charge of particles of kind "b"; $\tilde{g}$ is the determinant of the metric $\tilde{g}_{ij}$; m_a and $\tilde{p}^i_a$ are the mass and the momentum of particles of kind "a";

$$\tilde{u}^i_a = \frac{\tilde{p}^i_a}{\sqrt{\tilde{g}_{kj}\tilde{p}^k_a\tilde{p}^j_a}},$$

$$\frac{d^4\tilde{p}}{\sqrt{-\tilde{g}}}$$

is the invariant volume element in the four-dimensional space with coordinates $\tilde{p}_i$ [8]; and $\tilde{N}_a(q^i, \tilde{p}^i_a)$ is the Klimontovich random function [7]:

$$\tilde{N}_a(q^i, \tilde{p}_j) = \sum_{i=1}^{n_a}\int d\tilde{s}\,\delta^4(q^i - q^i_{(l)})\delta^4(\tilde{p}_j - \tilde{p}^{(l)}_j(\tilde{s})), \tag{6}$$

where n_a is the number of particles of kind "a" and $\tilde{s}$ is a canonical parameter along the trajectory of particles:

$$d\tilde{s} = \sqrt{g_{ij}dq^i dq^j},$$

$q^i_{(l)}$ and $\tilde{p}^{(l)}_j$ are the coordinates and the momentum of the lth particle of kind "a" that are determined from the equations of motion

$$\frac{dq^i_{(l)}}{d\tilde{s}} = \frac{\tilde{p}^i_{(l)}}{m_a c},$$
$$\frac{d\tilde{p}^{(l)}_i}{d\tilde{s}} = \frac{1}{m_a c}\left(\tilde{\Gamma}_{j,ik}\tilde{p}^j_{(l)}\tilde{p}^k_{(l)} + \frac{e_a}{c}\tilde{F}_{ik}\tilde{p}^k_{(l)}\right) \tag{7}$$

($\tilde{\Gamma}_{j,ik}$ are the Christoffel symbols of the first kind calculated by the metric $\tilde{g}_{ij}$).

According to Eqs. (7), function (6) satisfies the following equation:

$$\tilde{p}^i\frac{\partial\tilde{N}_a}{\partial q^i} + \tilde{\Gamma}_{j,ik}\tilde{p}^j\tilde{p}^k\frac{\partial\tilde{N}_a}{\partial\tilde{p}_i} + \frac{e_a}{c}\tilde{F}_{ik}\tilde{p}^k\frac{\partial\tilde{N}_a}{\partial\tilde{p}_i} = 0. \tag{8}$$

Let us represent the metric $\tilde{g}_{ij}$ of the gravitational field produced by all the particles as a sum of the generalized metric g_{ij} and the contribution h_{ij} due to microscopic interactions between particles:

$$\tilde{g}_{ij} = g_{ij} + h_{ij}, \tag{9}$$

where $g_{ij} = \langle\tilde{g}_{ij}\rangle$ is the metric $\tilde{g}_{ij}$ averaged over the ensemble [1]. Note that $\langle h_{ij}\rangle \equiv 0$. The tensor of electromagnetic field $\tilde{F}_{ik}$ can also be represented as

$$\tilde{F}_{ik} = F_{ik} + \omega_{ik}, \tag{10}$$

where $F_{ik} = \langle\tilde{F}_{ik}\rangle$ is the Maxwell tensor averaged over the ensemble and ω_{ik} is a microscopic tensor of electromagnetic field due to interactions between particles. Note that $\langle\omega_{ik}\rangle \equiv 0$.

Along with the momenta

$$\tilde{p}^i_{(l)} = m_a c\frac{dq^i_{(l)}}{d\tilde{s}},$$

we will also use the momenta p^i measured in the metric g_{ij}:

$$p^i_{(l)} = \alpha^{-1}(q, p)\tilde{p}^i_{(l)},$$
$$\alpha(q, p) = \frac{ds}{d\tilde{s}} = \frac{(g_{ij}p^i p^j)^{1/2}}{(\tilde{g}_{lk}\tilde{p}^l\tilde{p}^k)^{1/2}}. \tag{11}$$

Here, s is a canonical parameter calculated by the metric g_{ij}.

Let us pass from $\tilde{p}_i$ to p_i by the rule

$$\tilde{p}_j = \tilde{g}_{jk}\tilde{p}^k = \alpha\tilde{g}_{jk}g^{ki}p_i. \tag{12}$$

The Jacobian of transformation (12) is equal to (see [9])

$$\left|\frac{\partial\tilde{p}_i}{\partial p_j}\right| = \alpha^4\frac{\tilde{g}}{g}, \tag{13}$$

where g is the determinant of the metric g_{ij}. Let us introduce a function $N_a(q^i, \tilde{p}_j)$ defined in the eight-dimensional phase space with coordinates (q, p):

$$N_a(q, p) = \sum_{l=1}^{n_a} \int ds \delta^4(q^i - q^i_{(l)}(s)) \delta^4(p_j - p_j^{(l)}(s)). \quad (14)$$

The functions $q^i_{(l)}$ and $p_j^{(l)}$ in (14) are determined from the equations obtained from (7) by the change of variables (12) ($p^i = g^{ij}p_j$).

Note that the functions $\tilde{N}_a$ and N_a are related by the formula

$$\tilde{N}_a(q, \tilde{p}) = \frac{g}{\tilde{g}\alpha^5} N_a(q, p). \quad (15)$$

An equation for $N_a(q, p)$ is obtained from the Liouville equation (8) by the change of variables (12) and (15):

$$p^i \frac{\partial N_a}{\partial q^i} + \Gamma_{j,ik} p^j p^k \frac{\partial N_a}{\partial p_i} + \frac{e_a}{c} F_{ik} p^k \frac{\partial N_a}{\partial p_j} = \frac{\partial}{\partial p_i}\left[\left(\Omega^m_{jk}\Delta_{mi} p^j p^k - \frac{e_a}{c}\psi^l_{.k}\Delta_{lj} p^k\right) N_a\right]. \quad (16)$$

Here,

$$\Delta_{ki} = g_{ki} - u_k u_i, \quad u^k = \frac{p^k}{\sqrt{p^l p_l}}, \qquad \Omega^m_{kj} = \tilde{\Gamma}^m_{kj} - \Gamma^m_{kj} \quad (17)$$

is the difference of the Christoffel symbols of the second kind for the metrics $\tilde{g}_{ij}$ and g_{ij}, and

$$\psi^l_{.k} = \frac{1}{\alpha(q, p)}\tilde{F}^l_{.k} - F^l_{.k} = \frac{1}{\alpha(q, p)}\tilde{g}^{lm}\tilde{F}_{mk} - g^{lm}F_{mk}. \quad (18)$$

Passing to the momenta p_i and the function N_a in (4) and (5), we obtain

$$\tilde{T}^{ij} = \sum_a m_a c^2 \int \frac{d^4 p_a}{\sqrt{-g}} \alpha(q, p) \sqrt{\frac{g}{\tilde{g}}} u^i_a u^j_a N_a(q, p_a), \quad (19)$$

$$\tilde{J}^i = \sum_b e_b c \int \frac{d^4 p_b}{\sqrt{-g}} \sqrt{\frac{g}{\tilde{g}}} u^i_b N_b(q, p_b), \quad (20)$$

where $d^4p/\sqrt{-g}$ is the invariant volume element in the unperturbed momentum space.

For further consideration, it is convenient to rewrite the Einstein equations as

$$R_{ij} + \nabla_m \Omega^m_{ij} - \nabla_j \Omega^m_{im} + \Omega^m_{mn}\Omega^n_{ij} - \Omega^m_{jn}\Omega^n_{im} = \chi \sum_a m_a c^2 \int \frac{d^4 p_a}{\sqrt{-g}} \alpha \sqrt{\frac{g}{\tilde{g}}} \left(\tilde{g}_{ik}\tilde{g}_{jm} - \frac{1}{2}\tilde{g}_{ij}\tilde{g}_{km}\right) u^k_a u^m_a N_a(q, p_a) + \chi \tilde{T}^{(el)}_{ij}. \quad (21)$$

Here, R_{ij} is the Ricci tensor of a Riemannian space with metric g_{ij} and ∇_m is a covariant derivative in this space.

Substitute (9) and (10) into the Maxwell equations (2). Taking into account that the gravitational interactions are weak, we expand the microscopic Maxwell equations up to the first-order terms in h_{ij}:

$$\nabla_k F^{ik} + \nabla_k (h^i_m F^{km} - h^k_m F^{im}) + \frac{1}{2}F^{ik}\nabla_k h + \nabla_k \omega^{ik} + \nabla_k (h^i_m \omega^{km} - h^k_m \omega^{im}) + \frac{1}{2}\omega^{ik}\nabla_k h = -4\pi \sum_a e_a \int \frac{d^4 p_a}{\sqrt{-g}} u^i_{(a)} \left(1 - \frac{1}{2}h\right) N_a(q, p_a). \quad (22)$$

In expressions (21) and (22) and below, raising and lowering indices is performed by the averaged metric g_{ij}, $h = h^l_l$.

Now, let us expand Eq. (6) up to the second-order terms in h_{ij} and ω_{ij} and average the equations obtained over the ensemble. When we restrict ourselves to the averaged equations up to the second-order terms in the interaction, we can obtain a closed system of equations for the one-particle distribution function

$$f_a(q, p) = \frac{\langle N_a \rangle}{n_a},$$

the averaged metric g_{ij}, and the macroscopic tensor of electromagnetic field (the Maxwell tensor). An equation for f_a was obtained earlier in [5, 6].

3. MACROSCOPIC SYSTEM OF EINSTEIN AND MAXWELL EQUATIONS

Upon averaging over ensembles of microscopic equations by the scheme that was described in detail in [1], we arrived at macroscopic Einstein and Maxwell equations. These equations are now expressed as

$$G_{ij} + \nabla_k \varphi^k_{ij} + \mu_{ij} - \chi \tau^{(\mathrm{gr})}_{ij} = \chi T_{ij}, \quad (23)$$

$$\nabla_k F^{ik} + \nabla_k \varphi^{ik} + \mu^i = -\frac{4\pi}{c} J^i. \quad (24)$$

Here, G_{ij} is the Einstein tensor of a Riemannian space with a macroscopic metric g_{ij}, F^{ik} is the Maxwell tensor, J^i is the macroscopic four-vector of electric current density, and T_{ij} is the macroscopic tensor of energy–momentum. The last tensor represents a sum of the macroscopic energy–momentum tensors $T_{ij}^{(\mathrm{m})}$ of the medium, the macroscopic tensor of electromagnetic field $T_{ij}^{(\mathrm{el})}$, and the macroscopic energy–momentum tensor $\tau_{ij}^{(\mathrm{r})}$ of electromagnetic radiation in a plasma. (As applied to the cosmological plasma, by the latter we will mean the energy–momentum tensor of the relic radiation.)

The macroscopic Einstein equations differ from the classical Einstein equations by the additional terms $\nabla_k \varphi_{ij}^k$, μ_{ij}, and $-\chi\tau_{ij}^{(\mathrm{gr})}$ on the left-hand side. These tensors are explicitly expressed in terms of one-particle distribution functions by the formulas given below. The last of these additional terms represents a correction to the macroscopic tensor of electromagnetic radiation due to gravitational interaction that is multiplied by the Einstein constant and carried over, with minus sign, from the right- to the left-hand side of the Einstein equations.

The macroscopic Maxwell equations in general relativity theory also differ from the classical Maxwell equations due to the additional terms $\nabla_k \varphi^{ki} + \mu^i$ on their left-hand side. These additional terms are associated both with the effects of interaction and the effects of general relativity theory. They are also explicitly expressed in terms of one-particle distribution functions.

Let us write out the additional terms in the explicit form:

$$\varphi_{ij}^k = \sum_{b,c} \frac{\chi^2 e_b e_c m_b^2 m_c^2 c^5}{16\pi^2} \int \frac{d^3p'}{p'^0\sqrt{-g}}$$

$$\times \int \frac{d^3p''}{p''^0\sqrt{-g}} \left[\frac{1}{2} g^{fk} u''_i u''_j + u'^k (u'u'')(\delta^f_j u''_i + \delta^f_i u''_j) \right] \quad (25)$$

$$\times (u'u'') K_{f\alpha}(u', u'') \left(F_c(x'') \frac{\partial F_b(x')}{\partial p'_\alpha} - F_b(x') \frac{\partial F_c(x'')}{\partial p''_\alpha} \right),$$

$$\mu_{ij} = \sum_{b,c} \frac{\chi^2 e_b e_c m_b^2 m_c^2 c^5}{16\pi^2} \int \frac{d^3p'}{p'^0\sqrt{-g}}$$

$$\times \int \frac{d^3p''}{p''^0\sqrt{-g}} \left[\left(z^2 + \frac{1}{2} \right) \right.$$

$$+ (u''_i u''_j + u'_i u'_j) g^{\mathrm{qr}} + \left(z^2 - \frac{1}{2} \right) g_{ij} g^{\mathrm{qr}}$$

$$- 2z(u'_i u''_j + u''_i u'_j) g^{qr} - \left(z^2 - \frac{1}{2} \right) \quad (26)$$

$$\left. \times (\delta^q_i \delta^r_j + \delta^q_j \delta^r_i) \right] (z\delta^m_\alpha - u''_\alpha u'^m) J_{rqm}(u', u'')$$

$$\times F_c(x'') \frac{\partial F_b(x')}{\partial p'_\alpha},$$

$$\varphi^{ik} = \sum_{b,c} \frac{\chi e_b e_c^2 m_b^2 m_c c^3}{2\pi} \int \frac{d^3p'}{p'^0\sqrt{-g}} \int \frac{d^3p''}{p''^0\sqrt{-g}} (u'u'')$$

$$\times K_{f\alpha}(u', u'') [(u'u'')(u'^i g^{kf} - u'^k g^{if}) \quad (27)$$

$$- (u''^i g^{kf} - u''^k g^{if})]$$

$$\times \left(F_c(x'') \frac{\partial F_b(x')}{\partial p'_\alpha} - F_b(x') \frac{\partial F_c(x'')}{\partial p''_\alpha} \right),$$

$$\mu^i = \sum_{b,c} \frac{\chi e_b^2 e_c m_b m_c^2 c^3}{4\pi} \int \frac{d^3p'}{p'^0\sqrt{-g}} \int \frac{d^3p''}{p''^0\sqrt{-g}} \quad (28)$$

$$\times [(u'u'')\delta^s_\alpha - u''_\alpha u'^s] u'^k J^i_{ks}(u', u'') F_c(x'') \frac{\partial F_b(x')}{\partial p'_\alpha},$$

$$\tau_{ij}^{(\mathrm{gr})} = \sum_{b,c} \frac{\chi e_b e_c m_b^2 m_c^2 c^5}{16\pi^2} \int \frac{d^3p'}{p'^0\sqrt{-g}}$$

$$\times \int \frac{d^3p''}{p''^0\sqrt{-g}} [2z\delta^p_i \delta^q_j - z g_{ij} g^{pq} - (\delta^q_i u''_j + \delta^q_j u''_i) u'^p$$

$$+ g^{pq}(u'_i u''_j + u''_i u'_j)] J^{(\mathrm{gr})}_{pqf}(u', u'') F_c(x'') \quad (29)$$

$$\times \frac{\partial}{\partial p'_n} \left\{ F_b(x') \left[\left(z^2 - \frac{1}{2} \right) \delta^f_n \right. \right.$$

$$\left. \left. + \left(z^2 + \frac{1}{2} \right) u'_n u'^f - 2z u''_n u'^f \right] \right\}.$$

Here,

$$\frac{d^3p'}{p'^0\sqrt{-g}} \quad \text{and} \quad \frac{d^3p''}{p''^0\sqrt{-g}}$$

are the invariant volume elements in the three-dimensional momentum space of particles of kinds "*b*" and "*c*," respectively. The Greek index "α" in formulas (25)–(28) runs only through the values 1, 2, and 3

(the space index). The derivative with respect to p'_f in (29) should be calculated as if all four components of the momentum were independent. The dependence of p'_0 on p'_α is taken into account after differentiation with respect to p'_f.

These expressions coincide with the corresponding expressions for the additional terms in the macroscopic Maxwell and Einstein equations obtained in [1] under the assumption that the system consists of particles of equal masses. However, they become different for systems with different masses. These differences are displayed explicitly by the tensor $J_{ijk}(u', u'')$ in formulas (26), (28), and (29).

In [5], explicit expressions for these tensors (see (28), (29), and (37)) were presented in a locally Lorentz reference frame. It was pointed out that one should take the center-of-momentum frame as such a reference frame. To facilitate the calculations, it was assumed in [1, 5] that all the particles in the system have equal rest masses; hence, one should assume that, in the center-of-momentum frame, the velocities of interacting particles have equal magnitudes but opposite directions:

$$\mathbf{v}'' = -\mathbf{v}',$$

where $\mathbf{v}'$ and $\mathbf{v}''$ are three-dimensional velocities of interacting particles.

If we do not require that the masses of interacting particles should be equal, then the following equality holds in the center-of-momentum frame:

$$\mathbf{p}'' = -\mathbf{p}',$$

where $\mathbf{p}'$ and $\mathbf{p}''$ are the three-dimensional momenta of interacting particles.

In this frame of reference,

$$K_{00} = K_{0\alpha} = 0,$$

$$K_{\alpha\beta} = \frac{4\pi^2 c}{v' u'_0 u''_0 k^2_{\min}(1 + m_b u'_0/m_c u''_0)}\left(\delta_{\alpha\beta} - \frac{v'_\alpha v'_\beta}{v'^2}\right). \quad (30)$$

Here,

$$v' = \sqrt{v'^2_1 + v'^2_2 + v'^2_3}, \quad v'_\alpha = v'^\alpha = u'^\alpha/u'^0$$

are the space components of the vector $\mathbf{v}'$.

A covariant generalization of (21) is expressed as

$$K_{ij}(u', u'') = \frac{4\pi^2}{k^2_{\min}[(u'u'')^2 - 1]^{3/2}}\{-[(u'u'')^2 - 1]g_{ij} - u'_i u'_j - u''_i u''_j + (u'u'')(u'_i u''_j + u''_i u'_j)\}. \quad (31)$$

The expression for $K_{ij}(u', u'')$ turns out to be divergent as $k \longrightarrow 0$, i.e., at large target distances. This divergence is associated with the fact that we integrate over an infinite domain, whereas, in fact, we should restrict the integration to the domain of correlation, where the metric is assumed to be slowly varying. To surmount this difficulty, one should introduce, just as when deriving a kinetic equation, a truncation in the divergent integral

$$\int_0^\infty \frac{dk}{k^3}.$$

We set the lower limit of integration equal to

$$k_{\min} = 1/r_{\max}$$

rather than zero; here, $r_{\max}$ is the size of the correlation domain (the correlation length). Then, the above integral reduces to

$$\frac{1}{2k^2_{\min}} = r^2_{\max}/2.$$

Tensor (31) has the following properties:

$$K_{ij}(u', u'') = K_{ij}(u'', u'),$$
$$K_{ij}u'^i = K_{ij}u''^i = 0, \quad K_{ij} = K_{ji}. \quad (32)$$

Let us write out formula (27) from [5] in the center-of-momentum frame, in which $\mathbf{p}'' = -\mathbf{p}'$. In this frame of reference, the components of $J^{(1)}_{lmn}(u', u'')$ are expressed as (the space indices of the three-dimensional velocity vector v'^α are lowered by the three-dimensional Kronecker delta $\delta_{\alpha\beta}$)

$$J_{000} = -\frac{2}{(1 + m_b u'_0/m_c u''_0)}\frac{1}{u'_0 u''_0}\left(\frac{m_b u'_0}{m_c u''_0}\right)^3 \times \alpha\left(\frac{m_b u'_0}{m_c u''_0}v'\right)\frac{v'^2}{c^2}, \quad (33)$$

$$J_{00\alpha} = -\frac{2}{(1 + m_b u'_0/m_c u''_0)}\frac{1}{u'_0 u''_0}\left(\frac{m_b u'_0}{m_c u''_0}\right)^2 \times \alpha\left(\frac{m_b u'_0}{m_c u''_0}v'\right)\frac{v'_\alpha}{c}, \quad (34)$$

$$J_{0\alpha\beta} = -\frac{2}{(1 + m_b u'_0/m_c u''_0)}\frac{1}{u'_0 u''_0}\left(\frac{m_b u'_0}{m_c u''_0}\right)\alpha\left(\frac{m_b u'_0}{m_c u''_0}v'\right)\delta_{\alpha\beta} + \frac{2}{(1 + m_b u'_0/m_c u''_0)}\frac{1}{u'_0 u''_0}\left(\frac{m_b u'_0}{m_c u''_0}\right)\beta\left(\frac{m_b u'_0}{m_c u''_0}v'\right) \quad (35)$$

$$\times\left(\delta_{\alpha\beta}-\frac{v'_\alpha v'_\beta}{v'^2}\right),$$

$$J_{\alpha\beta\gamma} = -\frac{2}{(1+m_b u'_0/m_c u''_0)}\frac{1}{u'_0 u''_0}\alpha\left(\frac{m_b u'_0}{m_c u''_0}v'\right)\frac{c^2}{v'^2}$$

$$\times\left[\delta_{\alpha\beta}\frac{v'_\gamma}{c}+\delta_{\alpha\gamma}\frac{v'_\beta}{c}+\delta_{\beta\gamma}\frac{v'_\alpha}{c}-2\frac{v'_\alpha v'_\beta v'_\gamma}{c v'^2}\right]$$

$$+\frac{2}{(1+m_b u'_0/m_c u''_0)}\frac{1}{u'_0 u''_0} \tag{36}$$

$$\times\beta\left(\frac{m_b u'_0}{m_c u''_0}v'\right)\frac{c^2}{v'^2}\left[\left(\delta_{\alpha\beta}-\frac{v'_\alpha v'_\beta}{v'^2}\right)\frac{v'_\gamma}{c}\right.$$

$$\left.+\left(\delta_{\alpha\gamma}-\frac{v'_\alpha v'_\gamma}{v'^2}\right)\frac{v'_\beta}{c}+\left(\delta_{\beta\gamma}-\frac{v'_\beta v'_\gamma}{v'^2}\right)\frac{v'_\alpha}{c}\right].$$

The functions α and β in formulas (33)–(36) depend only on the argument

$$w = \frac{m_b u'_0}{m_c u''_0}v'$$

and are represented in the following explicit forms:

$$\alpha = \frac{\pi c^3}{w^3 k_{\min}}\left[\frac{2\frac{w}{c}\left(1+\frac{w^2}{c^2}\right)}{\left(1-\frac{w^2}{c^2}\right)^2}+\ln\left(\frac{1-\frac{w}{c}}{1+\frac{w}{c}}\right)\right], \tag{37}$$

$$\beta = \frac{\pi c^3}{2w^3 k_{\min}}\left[\frac{2\frac{w}{c}\left(3-2\frac{w^2}{c^2}+3\frac{w^4}{c^4}\right)}{\left(1-\frac{w^2}{c^2}\right)^2}\right.$$

$$\left.+3\left(1+\frac{w^2}{c^2}\right)\ln\left(\frac{1-\frac{w}{c}}{1+\frac{w}{c}}\right)\right]. \tag{38}$$

Here, we introduced the following notation for the integral:

$$\frac{1}{k_{\min}} = \int_{k_{\min}}^{\infty}\frac{dk}{k^2}.$$

By the arguments given above, here we again set the lower index equal to $k_{\min} = 1/r_{\max}$.

A covariant generalization of these results, which are obtained in a locally Lorentz center-of-mass frame of reference, to arbitrary reference frames has the form

$$J^{(1)}_{ijk}(u', u'') = J^{(2)}_{ijk}(u'', u') = J_{ijk}(u', u''), \tag{39}$$

$$J_{ijk}(u', u'') = A[(g_{ij}u'_k + g_{ik}u'_j + g_{jk}u'_i)$$

$$-z(g_{ij}u''_k + g_{ik}u''_j + g_{jk}u''_i)$$

$$-(u'_i u''_j u''_k + u''_i u'_j u''_k + u''_i u''_j u'_k) + 3z u''_i u''_j u''_k] \tag{40}$$

$$+C[u'_i u'_j u'_k - z(u'_i u'_j u''_k + u'_i u''_j u'_k + u''_i u'_j u'_k)$$

$$+z^2(u'_i u''_j u''_k + u''_i u'_j u''_k + u''_i u''_j u'_k) - z^3 u''_i u''_j u''_k].$$

Here,

$$z = (u'u'') = (u'^i u''_i),$$

$$A = \frac{\pi}{k_{\min}}\frac{(\mu^2+2\mu z+1)^{1/2}(1+\mu z)^2}{\mu^3(z^2-1)^{5/2}}$$

$$\times\left[\frac{2\mu\sqrt{z^2-1}(1+3\mu^2+2\mu z-2\mu^2 z^2)}{(1+\mu z)(\mu^2+2\mu z+1)}\right.$$

$$+\frac{(1-3\mu^2+2\mu z+4\mu^2 z^2)}{(1+\mu z)^2} \tag{41}$$

$$\left.\times\ln\left(\frac{1+\mu z-\mu\sqrt{z^2-1}}{1+\mu z+\mu\sqrt{z^2-1}}\right)\right],$$

$$C = \frac{\pi}{k_{\min}}\frac{(\mu z+1)}{(1+2\mu z+\mu^2)^{1/2}\mu^3(z^2-1)^{7/2}}$$

$$\times[2\mu\sqrt{z^2-1}(5+7\mu^2+10\mu z-2\mu^2 z^2)$$

$$+\frac{(1+2\mu z+\mu^2)}{1+\mu z} \tag{42}$$

$$\left.\times(5-7\mu^2+10\mu z+12\mu^2 z^2)\ln\left(\frac{1+\mu z-\mu\sqrt{z^2-1}}{1+\mu z+\mu\sqrt{z^2-1}}\right)\right],$$

$$\mu = m_b/m_c.$$

When $\mu = 1$, these results coincide with the results of [1].

Thus, we have generalized the results of [1] to the case of a multicomponent system of electromagnetically and gravitationally interacting particles with different masses.

The tensor $J_{ijk}(u', u'')$ satisfies the identity

$$J_{ijk}(u', u'')u''^{k} = 0. \quad (43)$$

The tensors φ_{ij}^{k}, μ_{ij}, $\tau_{ij}^{(\mathrm{gr})}$, and μ^{i} must satisfy the following additional conditions

$$g^{lj}\nabla_{l}(\nabla_{k}\varphi_{ij}^{k} + \mu_{ij} - \chi\tau_{ij}^{(\mathrm{gr})}) = 0, \quad (44)$$

$$\nabla_{i}\mu^{i} = 0, \quad (45)$$

because the divergences of all the other tensors in the macroscopic Einstein and Maxwell equations are identically zero.

Equations (44) and (45) impose certain constraints on the dependence of the parameters r_{D} and r_{g}, which enter the expressions for $J_{rpq}^{(\mathrm{el})}$ and $J_{rpq}^{(\mathrm{gr})}$, respectively, on the coordinates and the relative velocity of particles.

The macroscopic energy–momentum tensor of the plasma particles and the current four-vector can also be expressed in terms of seven-dimensional distribution functions:

$$T_{ij}^{(\mathrm{m})} = \sum_{a} c\int \frac{d^{3}p}{p^{0}\sqrt{-g}} p_{i}p_{j}F_{a}(p), \quad (46)$$

$$J^{i} = \sum_{a} e_{a}c\int \frac{d^{3}p}{p^{0}\sqrt{-g}} p^{i}F_{a}(p). \quad (47)$$

One should supplement the system of equations obtained with a kinetic equation for F_{b}. For a relativistic plasma, this equation was obtained in [10, 11].

4. POSSIBLE APPLICATIONS OF THE THEORY

The equations of the gravitational field for continua obtained in this paper differ from the classical Einstein equations by the additional terms

$$\nabla_{k}\varphi_{ij}^{k} + \mu_{ij} - \chi\tau_{ij}^{(\mathrm{gr})}$$

on the left-hand side.

These terms are proportional to the squared Einstein constant; however, they are also proportional to the squared density of particles. Hence, these additional terms may only be effective in continua with sufficiently high density. Such densities may take place at early stages of evolution of the universe, as well as inside objects that are close to the state of gravitational collapse. Therefore, it is natural to apply the equations obtained to the theory of early stages of stellar evolution and to the theory of gravitational collapse.

In [3], the authors suggested that the macroscopic Einstein equations should be applied to the construction of cosmological models. In [1, 12], the possibility of applying the macroscopic Einstein equations to the early stages of stellar evolution was assessed.

REFERENCES

1. A. V. Zakharov, Teor. Mat. Fiz. **125**, 107 (2000).
2. Yu. L. Klimontovich, *The Kinetic Theory of Electromagnetic Processes* (Nauka, Moscow, 1980; Springer, Berlin, 1983).
3. M. F. Shirokov and I. Z. Fisher, Astron. Zh. **39**, 899 (1962) [Sov. Astron. **6**, 699 (1962)].
4. A. V. Zakharov, Zh. Éksp. Teor. Fiz. **110**, 3 (1996) [JETP **83**, 1 (1996)].
5. A. V. Zakharov, Zh. Éksp. Teor. Fiz. **112**, 1153 (1997) [JETP **85**, 627 (1997)].
6. A. V. Zakharov and R. K. Mukharlyamov, Zh. Éksp. Teor. Fiz. **123**, 665 (2003) [JETP **96**, 587 (2003)].
7. Yu. L. Klimontovich, Zh. Éksp. Teor. Fiz. **37**, 735 (1959) [Sov. Phys. JETP **10**, 535 (1959)].
8. N. A. Chernikov, Nauchn. Dokl. Vyssh. Shkoly Fiz. Mat., No. 1, 168 (1959).
9. A. V. Zakharov, Zh. Éksp. Teor. Fiz. **96**, 769 (1989) [Sov. Phys. JETP **69**, 437 (1989)].
10. Yu. L. Klimontovich, Zh. Éksp. Teor. Fiz. **38**, 1212 (1960) [Sov. Phys. JETP **11**, 876 (1960)].
11. A. V. Zakharov, Zh. Éksp. Teor. Fiz. **86**, 3 (1984) [Sov. Phys. JETP **59**, 1 (1984)].
12. A. V. Zakharov, gr-qc/0002087.

Translated by I. Nikitin

Journal of Experimental and Theoretical Physics, Vol. 99, No. 5, 2004, pp. 898–914.
From Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 5, 2004, pp. 1034–1050.

NUCLEI, PARTICLES, AND THEIR INTERACTION

Radiative Corrections to Deep Inelastic Electron–Deuteron Scattering: The Case of Tensor-Polarized Deuteron¶

G. I. Gakh and O. N. Shekhovtsova

National Science Center Kharkov Institute of Physics and Technology, Kharkov, 61108 Ukraine
e-mail: shekhovtsova@kipt.kharkov.ua

Received April 6, 2004

Abstract—Model-independent radiative corrections to deep inelastic scattering of an unpolarized electron beam off the tensor-polarized deuteron target are considered. The contribution to the radiative corrections due to the hard photon emission from the elastic electron–deuteron scattering (the so-called elastic radiative tail) is also investigated. The calculation is based on the covariant parametrization of the deuteron quadrupole polarization tensor. Radiative corrections to the polarization observables are estimated numerically for the kinematical conditions of the current experiment at HERA. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The flavor structure of nucleons is described in terms of parton distribution functions. Most of the information on these functions has up to now come from inclusive deep inelastic scattering processes: experiments where only the scattered lepton is detected. Investigation of the nucleon spin structure involves new types of reactions. For example, the HERMES experiment was specifically designed to perform accurate measurements of semi-inclusive reactions, where, in addition to the scattered lepton, some of the hadrons produced are also detected [1].

The polarized nuclei of deuterium and helium-3 are used to extract information on the neutron spin-dependent structure function $g_1(x)$ [2]. In analyzing the experimental data on inclusive spin asymmetries for deuterium, a small effect due to a possible tensor polarization in this spin-1 target must be taken into account in order to deduce the spin-dependent structure function g_1^d. This is connected with the presence of additional tensor-polarized structure functions in a deuteron target [1]. So far, spin-structure studies have focused on the spin-1/2 nucleon. Different spin physics, such as the tensor structure in the deuteron, exists for higher-spin hadrons. Measurement of these additional spin-dependent structure functions provides important information about nonnucleonic components in spin-1 nuclei and tensor structures at the quark–parton level [3]. A general formalism of deep inelastic electron–deuteron scattering was discussed in [4], where new four tensor structure functions $b_i(x)$, $i = 1, \ldots, 4$ were introduced. They can be measured using a tensor-polarized target and an unpolarized electron beam. Among these new structure functions, only one, b_1, is the leading twist in QCD [4], and it was found that this function is small for a weakly coupled system of nucleons (for example, the deuteron). Therefore, the measurement of b_1 for a deuteron can provide information on its possible exotic components.

From the theoretical standpoint, the spin-dependent structure function $b_1(x)$ was investigated in a number of papers. The available fixed targets with $J \geq 1$ are only nuclei (deuteron is the most commonly used nucleus). If the nucleons in the deuteron are in the S state, then $b_1(x) \equiv 0$. For nucleons in the D state, $b_1(x) \neq 0$ in general [4]. It was found [5] that, in the quark–parton model, the sum rule

$$\int dx b_1(x) = 0$$

is generally true if the sea of quarks and antiquarks is unpolarized (and it was shown how this sum rule is modified in the presence of a polarized sea). Mankiewicz [6] studied $b_1(x)$ for the ρ meson and noticed empirically that

$$\int dx b_1(x) = 0$$

in his model. It was shown in [7] that multiple scattering terms at low x can still lead to $b_1(x) \neq 0$ even in the case where only the S-wave component is present. Various twist-two structure functions of deuterons (in particular, b_1) have been calculated in a version of the convolution model that incorporates relativistic and binding energy corrections [8]. Simple parametrizations of these structure functions are given in terms of few deuteron wavefunction parameters and the free nucleon structure functions. The tensor structure functions were

¶ This article was submitted by authors in English.

discussed in [9] in the case of lepton scattering and in hadron reactions such as the polarized proton–deuteron Drell–Yan process.

As is known, the HERMES experiment has been designed to measure the nucleon spin-dependent structure functions from deep inelastic scattering of longitudinally polarized positrons and electrons from polarized gaseous targets (H, D, ^{3}He). In 2000, HERMES collected a data set with a tensor-polarized deuterium target for the purpose of making the first measurement of the tensor structure function $b_1(x)$. The preliminary results on this structure function are presented in [10] for the kinematic range $0.002 < x < 0.85$ and $0.1 < Q^2 < 20$ GeV^2. The preliminary result for the tensor asymmetry is small enough to produce an effect of more than 1% on the measurement of g_1^d. The dependence of b_1 on the x variable is in qualitative agreement with the expectations based on coherent double-scattering models [11–13] and favors a sizeable value of b_1 in the low-x region. This suggests a significant tensor polarization of the sea quarks, violating the Close–Kumano sum rule [5].

The radiative corrections to deep inelastic scattering of unpolarized and longitudinally polarized electron beams on a polarized deuteron target were considered in [14] in a particular case of the deuteron polarization (which can be obtained from the general covariant spin-density matrix [15] when spin functions are eigenvectors of the spin projection operator). The leading-log model-independent radiative corrections in deep inelastic scattering of an unpolarized electron beam off the tensor-polarized deuteron target were considered in [16]. The calculation is based on the covariant parametrization of the deuteron quadrupole polarization tensor and uses a Drell–Yan-like representation.

Current experiments at modern accelerators have reached a new level of precision, and this circumstance requires a new approach to data analysis and inclusion of all possible systematic uncertainties. One of the important sources of such uncertainties is the electromagnetic radiative effect caused by physical processes occurring in higher orders of perturbation theory with respect to electromagnetic interaction. In the present paper, we give a covariant description of the deep inelastic scattering of an unpolarized electron beam off the tensor-polarized deuteron target (the polarization state of the target is described by the spin-density matrix of the general form) with the radiative corrections

$$e^-(k_1) + d(p) \longrightarrow e^-(k_2) + X(p_x) \tag{1}$$

taken into account.

The corresponding approach is based on the covariant parametrization of the deuteron quadrupole polarization tensor in terms of the 4-momenta of the particles in process (1) [16]. We also performed numerical calculations of the radiative corrections for the kinematical conditions of the experiment [10]. The contribution of the radiative tail from the elastic *ed* scattering is considered separately.

2. BORN APPROXIMATION

The standard set of variables used for the description of deep inelastic scattering processes is

$$x = \frac{-q^2}{2pq}, \quad y = \frac{2pq}{V}, \qquad V = 2pk_1, \quad q^2 = -Vxy, \quad q = k_1 - k_2, \tag{2}$$

where q is the 4-momentum of the intermediate heavy photon that probes the deuteron structure. We first define the deep inelastic scattering cross section of process (1) in terms of the contraction of the leptonic $L^B_{\mu\nu}$ and hadronic $W_{\mu\nu}$ tensors (in the Born approximation, we can neglect the electron mass)

$$\frac{d\sigma}{dx dQ_B^2} = \frac{\pi\alpha^2}{VQ_B^4}\frac{y}{x}L^B_{\mu\nu}W_{\mu\nu}. \tag{3}$$

We note that, only in the Born approximation (without accounting for radiative corrections),

$$q = k_1 - k_2, \quad Q_B^2 = -q^2 = 2k_1k_2.$$

The Born leptonic tensor (in the unpolarized case) is

$$L^B_{\mu\nu} = q^2 g_{\mu\nu} + 2(k_{1\mu}k_{2\nu} + k_{1\nu}k_{2\mu}). \tag{4}$$

The hadronic tensor is defined as

$$W_{\mu\nu} = (2\pi)^3 \sum_X \delta^{(4)}(k_1 + p - k_2 - p_x)\overline{J_\mu J_\nu^*},$$

where J_μ is the electromagnetic current for the $\gamma^* + d \longrightarrow X$ transition (γ^* is the virtual photon). The sum means summation over the final states, and the bar means averaging over the polarizations of the target and summation over the polarizations of the final particles. To write the hadron tensor in terms of the structure functions, we first define the deuteron spin-density matrix (we do not consider the effect caused by the vector polarization of the deuteron in what follows)

$$\rho_{\mu\nu} = -\frac{1}{3}\left(g_{\mu\nu} - \frac{p_\mu p_\nu}{M^2}\right) - \frac{i}{2M}\varepsilon_{\mu\nu\lambda\rho}s_\lambda p_\rho + Q_{\mu\nu}, \qquad Q_{\mu\nu} = Q_{\nu\mu}, \quad Q_{\mu\mu} = 0, \quad p_\mu Q_{\mu\nu} = 0, \tag{5}$$

where s_μ and $Q_{\mu\nu}$ are the target deuteron polarization 4-vector and the deuteron quadrupole polarization tensor. The corresponding hadron tensor has both the polarization-independent and polarization-dependent parts and in the general case can be written as

$$W_{\mu\nu} = W_{\mu\nu}(0) + W_{\mu\nu}(V) + W_{\mu\nu}(T), \tag{6}$$

where $W_{\mu\nu}(0)$ corresponds to the unpolarized case and $W_{\mu\nu}(V)(W_{\mu\nu}(T))$ corresponds to the case of the vector

(tensor) polarization of the deuteron target. The $W_{\mu\nu}(0)$ term has the form

$$W_{\mu\nu}(0) = -W_1\tilde{g}_{\mu\nu} + \frac{W_2}{M^2}\tilde{p}_\mu\tilde{p}_\nu,$$
$$\tilde{g}_{\mu\nu} = g_{\mu\nu} - \frac{q_\mu q_\nu}{q^2}, \quad \tilde{p}_\mu = p_\mu - \frac{pq}{q^2}q_\mu, \tag{7}$$

where M is the deuteron mass and $W_{1,2}$ are the unpolarized structure functions depending on two independent variables x and q^2. The part of the hadron tensor that depends on the quadrupole polarization tensor can be represented as

$$W_{\mu\nu}(T) = \frac{M^2}{(pq)^2}\left\{Q_{\alpha\beta}q_\alpha q_\beta\left(B_1\tilde{g}_{\mu\nu} + \frac{\tilde{B}_2}{pq}\tilde{p}_\mu\tilde{p}_\nu\right)\right.$$
$$\left. + B_3 q_\alpha(\tilde{p}_\mu Q_{\tilde{\nu}\alpha} + \tilde{p}_\nu Q_{\tilde{\mu}\alpha}) + pqB_4\tilde{Q}_{\mu\nu}\right\}. \tag{8}$$

Here, B_i ($i = 1, 2, 3, 4$) are the spin-dependent structure functions (caused by the tensor polarization of the target). They are also functions of the two variables q^2 and x. Because the hadron tensor $W_{\mu\nu}(T)$ is symmetric under $\mu \longleftrightarrow \nu$, measuring these new structure functions does not require the electron beam to be polarized.

We used the following notation in formula (8):

$$Q_{\mu\tilde{\nu}} = Q_{\mu\nu} - \frac{q_\nu q_\alpha}{q^2}Q_{\mu\alpha}, \quad Q_{\mu\tilde{\nu}}q_\nu = 0,$$
$$\tilde{Q}_{\mu\nu} = Q_{\mu\nu} + \frac{q_\mu q_\nu}{q^4}Q_{\alpha\beta}q_\alpha q_\beta \tag{9}$$
$$-\frac{q_\nu q_\alpha}{q^2}Q_{\mu\alpha} - \frac{q_\mu q_\alpha}{q^2}Q_{\nu\alpha}, \quad \tilde{Q}_{\mu\nu}q_\nu = 0.$$

We note that the deuteron spin-dependent structure functions B_i are also related to the structure functions b_i introduced in [4] as

$$B_1 = -b_1, \quad B_2 = \frac{b_2}{3} + b_3 + b_4,$$
$$B_3 = \frac{b_2}{6} - \frac{b_4}{2}, \quad B_4 = \frac{b_2}{3} - b_3. \tag{10}$$

In calculating radiative corrections, it is convenient to parametrize the polarization state of the deuteron target in terms of the 4-momenta of the particles participating in the reaction under consideration. Therefore, first, we have to find the set of the axes and write them in a covariant form in terms of the 4-momenta. If we choose, in the laboratory system of reaction (1), the longitudinal direction **l** along the electron beam and the transverse one **t** in the plane ($\mathbf{k}_1$, $\mathbf{k}_2$) and perpendicular to **l**, then

$$S_\mu^{(l)} = \frac{2\tau k_{1\mu} - p_\mu}{M},$$
$$S_\mu^{(t)} = \frac{k_{2\mu} - (1 - y - 2xy\tau)k_{1\mu} - xyp_\mu}{d}, \tag{11}$$
$$S_\mu^{(n)} = \frac{2\varepsilon_{\mu\lambda\rho\sigma}p_\lambda k_{1\rho}k_{2\sigma}}{Vd},$$
$$d = \sqrt{Vxyb}, \quad b = 1 - y - xy\tau, \quad \tau = M^2/V.$$

We chose one of the axes along the direction **l** because, in the experiment on measuring the b_1 structure function [10], the direction of the magnetic field used for polarization of the deuteron target is along the positron beam line. The direction of the magnetic field provides the quantization axis for the nuclear spin in the target.

It can be verified that the set of the 4-vectors $S_\mu^{(l,t,n)}$ has the properties

$$S_\mu^{(\alpha)}S_\mu^{(\beta)} = -\delta_{\alpha\beta}, \quad S_\mu^{(\alpha)}p_\mu = 0, \quad \alpha, \beta = l, t, n, \tag{12}$$

and that in the rest frame of the deuteron (the laboratory system),

$$S_\mu^{(l)} = (0, \mathbf{l}), \quad S_\mu^{(t)} = (0, \mathbf{t}), \quad S_\mu^{(n)} = (0, \mathbf{n}),$$
$$\mathbf{l} = \mathbf{n}_1, \quad \mathbf{t} = \frac{\mathbf{n}_2 - (\mathbf{n}_1\cdot\mathbf{n}_2)\mathbf{n}_1}{\sqrt{1 - (\mathbf{n}_1\cdot\mathbf{n}_2)^2}}, \tag{13}$$
$$\mathbf{n} = \frac{\mathbf{n}_1\times\mathbf{n}_2}{\sqrt{1 - (\mathbf{n}_1\cdot\mathbf{n}_2)^2}}, \quad \mathbf{n}_{1,2} = \frac{\mathbf{k}_{1,2}}{|\mathbf{k}_{1,2}|}.$$

Adding one more 4-vector $S_\mu^{(0)} = p_\mu/M$ to set (11), we obtain a complete set of orthogonal 4-vectors with the properties

$$S_\mu^{(m)}S_\nu^{(m)} = g_{\mu\nu},$$
$$S_\mu^{(m)}S_\mu^{(n)} = g_{mn}, \quad m, n = 0, l, t, n. \tag{14}$$

This allows us to express the deuteron quadrupole polarization tensor in the general case as

$$Q_{\mu\nu} = S_\mu^{(m)}S_\nu^{(n)}R_{mn} \equiv S_\mu^{(\alpha)}S_\nu^{(\beta)}R_{\alpha\beta},$$
$$R_{\alpha\beta} = R_{\beta\alpha}, \quad R_{\alpha\alpha} = 0, \tag{15}$$

because the components R_{00}, $R_{0\alpha}$, and $R_{\alpha 0}$ vanish by the condition $Q_{\mu\nu}p_\nu = 0$.

In the Born approximation, the components R_{ln} and R_{tn} do not contribute to the cross section (because the 4-momenta q_μ and $k_{1\mu}$ are orthogonal to the 4-vector

$S_\mu^{(n)}$) and expansion (15) can be rewritten in the standard form

$$Q_{\mu\nu} = \left[S_\mu^{(l)}S_\nu^{(l)} - \frac{1}{2}S_\mu^{(t)}S_\nu^{(t)}\right]R_{ll} + \frac{1}{2}S_\mu^{(t)}S_\nu^{(t)}(R_{tt} - R_{nn}) + (S_\mu^{(l)}S_\nu^{(t)} + S_\mu^{(t)}S_\nu^{(l)})R_{lt}, \quad (16)$$

where we took into account that

$$R_{ll} + R_{tt} + R_{nn} = 0.$$

In what follows, we consider the deep inelastic scattering of the unpolarized electron beam from the tensor-polarized deuteron target. Thus, we have to calculate only the contraction of the Born leptonic tensor $L_{\mu\nu}^B$ and the hadronic tensor $W_{\mu\nu}(T)$ caused by the tensor polarization of the target,

$$S^B(T) = L_{\mu\nu}^B W_{\mu\nu}(T) = 8\frac{\tau}{y}\left\{-\frac{1}{y^2}[xy^2B_1 + (a-1+y)B_2 + yB_3]Q_0 + \frac{1}{y}[(2-y)B_3 - yB_4]Q_1 + B_4Q_{11}\right\}, \quad (17)$$

where

$$a = xy\tau, \quad Q_0 = Q_{\alpha\beta}q_\alpha q_\beta,$$
$$Q_1 = Q_{\alpha\beta}q_\alpha k_{1\beta}, \quad Q_{11} = Q_{\alpha\beta}k_{1\alpha}k_{1\beta}.$$

Using the formulas for the vectors $S_\mu^{(\alpha)}$, we can calculate the contractions. After simple calculation, we have

$$\frac{d\sigma_B(T)}{dxdQ_B^2} = \frac{2\pi\alpha^2}{xQ_B^4} \times [S_{ll}R_{ll} + S_{tt}(R_{tt} - R_{nn}) + S_{lt}R_{lt}], \quad (18)$$

with

$$S_{ll} = [2xb\tau - y(1+2x\tau)^2]G + 2b(1+3x\tau)B_3 + (b-a)B_4,$$
$$S_{lt} = 2\sqrt{\frac{xb\tau}{y}} \times [2(y+2a)G + (2-y-4b)B_3 + yB_4], \quad (19)$$
$$S_{tt} = -2xb\tau(G + B_3), \quad G = xyB_1 - \frac{b}{y}B_2.$$

Therefore, in the general case, the cross section of deep inelastic scattering of an unpolarized electron beam from a tensor-polarized target is determined, in the Born approximation, by the components of the quadrupole polarization tensor R_{ll}, R_{lt}, and the combination $(R_{tt} - R_{nn})$.

We now consider just one more, commonly used choice of the coordinate axes: components of the deuteron polarization tensor are defined in the coordinate system with the axes along the directions **L**, **T**, and **N** in the rest frame of the deuteron, where

$$\mathbf{L} = \frac{\mathbf{k}_1 - \mathbf{k}_2}{|\mathbf{k}_1 - \mathbf{k}_2|}, \quad \mathbf{T} = \frac{\mathbf{n}_1 - (\mathbf{n}_1\cdot\mathbf{L})\mathbf{L}}{\sqrt{1-(\mathbf{n}_1\cdot\mathbf{L})^2}}, \quad \mathbf{N} = \mathbf{n}. \quad (20)$$

The corresponding covariant form of set (20) is given by

$$S_\mu^{(L)} = \frac{2\tau(k_1-k_2)_\mu - yp_\mu}{M\sqrt{yh}},$$
$$S_\mu^{(T)} = \frac{(1+2x\tau)k_{2\mu} - (1-y-2x\tau)k_{1\mu} - x(2-y)p_\mu}{\sqrt{Vxbh}}, \quad (21)$$
$$S_\mu^{(N)} = S_\mu^{(n)}, \quad h = y + 4x\tau,$$

and the expansion of the deuteron polarization tensor is defined by full analogy with (16),

$$Q_{\mu\nu} = \left[S_\mu^{(L)}S_\nu^{(L)} - \frac{1}{2}S_\mu^{(T)}S_\nu^{(T)}\right]R_{LL} + \frac{1}{2}S_\mu^{(T)}S_\nu^{(T)}(R_{TT} - R_{NN}) + (S_\mu^{(L)}S_\nu^{(T)} + S_\mu^{(T)}S_\nu^{(L)})R_{LT}. \quad (22)$$

These two sets of orthogonal 4-vectors are connected by an orthogonal matrix that describes a rotation in the plane perpendicular to the direction **n** = **N**,

$$S_\mu^{(L)} = \cos\theta S_\mu^{(l)} + \sin\theta S_\mu^{(t)},$$
$$S_\mu^{(T)} = -\sin\theta S_\mu^{(l)} + \cos\theta S_\mu^{(t)}, \quad (23)$$
$$\cos\theta = \frac{y(1+2x\tau)}{\sqrt{yh}}, \quad \sin\theta = -2\sqrt{\frac{xb\tau}{h}}.$$

In this set of axes, the part of the differential cross section that depends on the tensor polarization can be written as

$$\frac{d\sigma_B(T)}{dxdQ_B^2} = \frac{2\pi\alpha^2}{xQ_B^4} \times [S_{LL}R_{LL} + S_{TT}(R_{TT} - R_{NN}) + S_{LT}R_{LT}], \quad (24)$$
$$S_{LL} = -hG + 2bB_3 + \frac{B_4}{h}[(1-y)(y-2x\tau) - 2a(y+x\tau)],$$

$$S_{TT} = \frac{2xb\tau}{h}B_4,$$
$$S_{LT} = 2\sqrt{\frac{xb\tau}{y}}(2-y)\left(B_3 + \frac{y}{h}B_4\right). \quad (25)$$

3. RADIATIVE CORRECTIONS

In this paper, we consider only the QED radiative corrections to the deep inelastic scattering process (1). We confine ourselves to calculation of the so-called model-independent radiative corrections, corresponding to photons radiated from a lepton line with the vacuum polarization taken into account. The reason is that it gives the leading contribution to radiative corrections due to the smallness of the electron mass and can be calculated without any additional assumptions. Nevertheless, these radiative corrections depend on the shape of the deuteron structure functions (both spin-independent and spin-dependent) via their dependence on variables x and Q^2.

There exist two contributions to radiative corrections when we take the corrections of the order α into account. The first one is caused by virtual and soft photon emission that cannot affect the kinematics of process (1). The second one arises due to the radiation of a hard photon,

$$e^-(k_1) + d(p) \longrightarrow e^-(k_2) + \gamma(k) + X(p_x). \qquad (26)$$

The leptonic tensor corresponding to the hard-photon radiation is well known [17, 18]. For an unpolarized electron beam, it can be written as

$$L^\gamma_{\mu\nu} = A_0\tilde{g}_{\mu\nu} + A_1\tilde{k}_{1\mu}\tilde{k}_{1\nu} + A_2\tilde{k}_{2\mu}\tilde{k}_{2\nu}, \qquad (27)$$

where

$$A_0 = -\frac{(q^2+\chi_1)^2 + (q^2-\chi_2)^2}{\chi_1\chi_2} - 2m^2q^2\left(\frac{1}{\chi_1^2} + \frac{1}{\chi_2^2}\right),$$

$$A_1 = -4\left(\frac{q^2}{\chi_1\chi_2} + \frac{2m^2}{\chi_2^2}\right),$$

$$A_2 = -4\left(\frac{q^2}{\chi_1\chi_2} + \frac{2m^2}{\chi_1^2}\right),$$

$$\tilde{k}_{i\mu} = k_{i\mu} - \frac{qk_i}{q^2}q_\mu, \quad i = 1, 2,$$

with

$$\chi_{1,2} = 2kk_{1,2},$$

m is the electron mass,

$$q^2 = \chi_2 - \chi_1 - Q_B^2,$$

and

$$q = k_1 - k_2 - k$$

in this section. The hadronic tensor in this case has the same form as the hadronic tensor in the Born approximation, but the momentum transfer q differs from the Born one and the structure functions B_i depend on the new momentum q. Here and in what follows, we neglect the terms vanishing as $m \longrightarrow 0$.

We consider the hard photon (with the energy $\omega > \Delta\varepsilon$, where $\Delta \ll 1$) emission process using the approach in [19], where it was applied to the process of deep inelastic scattering on an unpolarized target. We introduce the variables suitable for this process,

$$z = \frac{M_x^2 - M^2}{V} = \frac{q^2 + 2pq}{V}, \quad r = -\frac{q^2}{Q_B^2},$$

$$x' = \frac{-q^2}{2pq} = \frac{xyr}{xyr + z}, \quad \chi_{1,2} = 2kk_{1,2},$$

where M_x is the invariant mass of the hadron system produced in scattering of the photon (with the virtuality q^2) by the target.

We note the physical meaning of the z variable: it shows the degree of deviation from the elastic process ($ed \longrightarrow ed$). Therefore, the value $z = 0$ corresponds to the elastic ed scattering threshold, and the value $z = \varepsilon_d/\varepsilon_1$ (where ε_d is the deuteron bound energy and ε_1 is the electron beam energy in the laboratory system) corresponds to the $ed \longrightarrow enp$ reaction threshold (quasielastic ed scattering).

The contraction of the leptonic and hadronic tensors can be represented as

$$S^\gamma(T) = AA_0 + BA_1 + CA_2, \qquad (28)$$

$$A = NQ_0\left[3B_1 + \frac{2\tau}{c}B_2 + \frac{c}{2xyr}(B_2 + 2B_3 + B_4)\right],$$

$$B = N\left\{Q_0\left[\frac{V}{2c}B_2 - V\frac{Q_B^2 + \chi_1}{2rQ_B^2}(B_2 + B_3)\right.\right.$$

$$\left. + \frac{(Q_B^2 + \chi_1)^2}{4rQ_B^2}\left(B_1 + \frac{Vc}{2rQ_B^2}(B_2 + 2B_3 + B_4)\right)\right]$$

$$\left. + VQ_1\left[B_3 - \frac{Q_B^2 + \chi_1}{2rQ_B^2}c(B_3 + B_4)\right] + \frac{V}{2}cQ_{11}B_4\right\},$$

$$C = N\left\{Q_0\left[\frac{V(1-y)^2}{2}\frac{1}{c}B_2\right.\right.$$

$$+ V\frac{Q_B^2 - \chi_2}{2rQ_B^2}(1-y)(B_2 + B_3)$$

$$\left. + \frac{(Q_B^2 - \chi_2)^2}{4rQ_B^2}\left(B_1 + \frac{Vc}{2rQ_B^2}(B_2 + 2B_3 + B_4)\right)\right]$$

$$\left. + VQ_2\left[B_3(1-y) + \frac{Q_B^2 - \chi_2}{2rQ_B^2}c(B_3 + B_4)\right] + \frac{V}{2}cQ_{22}B_4\right\},$$

where

$$N = 4\tau/Vc^2, \quad c = z + xyr.$$

The quantities Q_0, Q_1, Q_2, Q_{11}, and Q_{22} are the contractions of the deuteron quadrupole polarization tensor and 4-momenta. They can be expressed in terms of the scalar products of the 4-momenta of the particles participating in the reaction and the set of 4-vectors $S_\mu^{(l,t,n)}$. Therefore, these contractions are

$$\begin{aligned}
Q_0 &= Q_{\alpha\beta}q_\alpha q_\beta = \left[(lq)^2 - \frac{1}{2}(tq)^2 - \frac{1}{2}(nq)^2\right]R_{ll} \\
&+ 2lqtqR_{lt} + 2nqlqR_{ln} + 2nqtqR_{tn} \\
&+ \frac{1}{2}[(tq)^2 - (nq)^2](R_{tt} - R_{nn}), \\
Q_1 &= Q_{\alpha\beta}q_\alpha k_{1\beta} = \left(lk_1lq - \frac{1}{2}tk_1tq\right)R_{ll} \\
&+ (lk_1tq + tk_1lq)R_{lt} + lk_1nqR_{ln} \\
&+ tk_1nqR_{tn} + \frac{1}{2}tk_1tq(R_{tt} - R_{nn}), \\
Q_{11} &= Q_{\alpha\beta}k_{1\alpha}k_{1\beta} = \left[(lk_1)^2 - \frac{1}{2}(tk_1)^2\right]R_{ll} \\
&+ 2lk_1tk_1R_{lt} + \frac{1}{2}(tk_1)^2(R_{tt} - R_{nn}), \\
Q_2 &= Q_1(k_1 \longrightarrow k_2), \quad Q_{22} = Q_{11}(k_1 \longrightarrow k_2), \\
ia &= S_\mu^{(i)}a_\mu, \quad i = l, t, n,
\end{aligned} \tag{29}$$

where we used the conditions

$$R_{ll} + R_{tt} + R_{nn} = 0, \quad nk_1 = nk_2 = 0.$$

For the set of the 4-vectors $S_\mu^{(l,t,n)}$, we also have $tk_1 = 0$.

It is convenient to separate the poles in the term $(\chi_1\chi_2)^{-1}$ using the relation

$$\frac{1}{\chi_1\chi_2} = \frac{1}{Q_B^2}\frac{1}{1-r}\left(\frac{1}{\chi_1} - \frac{1}{\chi_2}\right).$$

Then, the radiative correction (caused by the hard-photon emission) to the differential cross section of deep inelastic scattering of an unpolarized electron beam by the tensor-polarized target has the form

$$\frac{d\sigma^\gamma}{dxdQ_B^2} = \frac{\alpha y}{Vx}\int\frac{d^3k}{2\pi\omega}\Sigma(z, r), \tag{30}$$

where ω is the energy of the hard photon and

$$\begin{aligned}
\Sigma(z, r) &= \frac{\alpha^2(q^2)}{Q_B^4}\left\{R_0(z, r) + \left(\frac{1}{\chi_1} - \frac{1}{\chi_2}\right)R_1(z, r)\right. \\
&\left. + \frac{m^2}{\chi_1^2}R_{1m}(z, r) + \frac{m^2}{\chi_2^2}R_{2m}(z, r)\right\},
\end{aligned} \tag{31}$$

$$R_0 = -\frac{2}{r^2}A,$$

$$R_1 = \frac{1}{r-1}\left[\left(1 + \frac{1}{r^2}\right)Q_B^2A - \frac{4}{r}(B + C)\right],$$

$$R_{1m} = 2\left(\frac{Q_B^2}{r}A - \frac{4}{r^2}C\right),$$

$$R_{2m} = 2\left(\frac{Q_B^2}{r}A - \frac{4}{r^2}B\right).$$

It is convenient to write the integral in Eq. (30) as

$$I = \int\frac{d^3k}{2\pi\omega}\Sigma(z, r) = I_{1m} + I_{2m} + I_R, \tag{32}$$

where we separate the contributions proportional to m^2,

$$\begin{aligned}
I_{1m} &= \int\frac{d^3k}{2\pi\omega}\frac{\alpha^2(q^2)}{Q_B^4}\frac{m^2}{\chi_1^2}R_{1m}(z, r), \\
I_{2m} &= \int\frac{d^3k}{2\pi\omega}\frac{\alpha^2(q^2)}{Q_B^4}\frac{m^2}{\chi_2^2}R_{2m}(z, r).
\end{aligned} \tag{33}$$

We first consider the integrals I_{im}, $i = 1, 2$. The numerator of the integrands in $I_{1m}(I_{2m})$ is then calculated in the approximation $\chi_1 = 0$ ($\chi_2 = 0$) [19]. The integration measure over the hard-photon phase space is written as

$$\frac{d^3k}{2\pi\omega} = \frac{dz}{z_+ - z}\frac{\omega^2 d\Omega_k}{2\pi}, \quad z_+ = y(1-x). \tag{34}$$

Using the invariance of $\omega^2 d\Omega_k$, we can integrate over the angular variables $d\Omega_k$ in the most suitable coordinate system, namely, in the coordinate frame, where

$$\mathbf{k}_1 - \mathbf{k}_2 + \mathbf{p} = 0$$

(the center-of-mass system of the scattered electron and the produced hadronic system). We obtain

$$\int\frac{\omega^2 d\Omega_k}{2\pi}\frac{m^2}{\chi_{1,2}^2} = \frac{1}{2}.$$

3.1. Integral I_{1m}

We calculate the integrand in the approximation where $\chi_1 = 0$ (except in the denominator). This approximation corresponds to the emission of a collinear photon along the initial-electron momentum. In this case, the variables take the values

$$r_1 = \frac{1-y+z}{1-xy}, \quad q_1^2 = -r_1 Q_B^2, \quad x_1' = \frac{xyr_1}{z+xyr_1}.$$

After integrating over the hard-photon angular variables, the integral I_{1m} can be represented as

$$I_{1m} = \frac{1}{Q_B^4}\int_0^{z_m} \frac{dz}{z_+ - z}\alpha_1^2 N_1 \Sigma_1(z), \tag{35}$$

$$\Sigma_1(z) = \Sigma_{1ll}R_{ll} + \Sigma_{1lt}R_{lt} + \Sigma_{1tt}(R_{tt} - R_{nn}),$$

$$\Sigma_{1tt} = b\frac{Q_B^2}{r_1}(G_t + B_{3t}),$$

$$\Sigma_{1lt} = -\frac{V}{r_1}\sqrt{\frac{xyb}{\tau}}[(y-1+r_1)B_{4t}$$

$$+ (a-3b+r_1)B_{3t} + 2(a-b+r_1)G_t],$$

$$\Sigma_{1ll} = \frac{V}{2\tau r_1}\{(a-b)(y-1+r_1)B_{4t}$$

$$+ 2b(b-2a-r_1)B_{3t} - [2ab-(a-b+r_1)^2]G_t\},$$

$$G_t = xyB_{1t} - \frac{b}{y-1+r_1}B_{2t}, \quad \alpha_1 = \alpha(q_1^2),$$

$$N_1 = \frac{4\tau}{(z+xyr_1)^2},$$

$$z_m = z_+ - \rho, \quad \rho = 2\Delta\varepsilon\sqrt{(\tau+z_+)/V},$$

$$B_{it} = B_i(q_1^2, x_1'), \quad i = 1\text{–}4.$$

It is convenient to explicitly extract the contribution containing the infrared divergence. For this, we add to the numerator of the integrand and subtract from it its value at $z = z_+$. At this value, we have

$$r_1 = 1, \quad \alpha_1 = \alpha, \quad N_1 = 4\tau/y^2, \quad x_1' = x.$$

The integral I_{1m} can thus be written as

$$I_{1m} = \frac{1}{Q_B^4}\int_0^{z_+} \frac{dz}{z_+ - z}$$

$$\times\left[\alpha_1^2 N_1\Sigma_1(z) - \alpha^2\frac{4\tau}{y^2}\Sigma_1(z_+)\right] + \frac{Vx}{\pi y}\ln\frac{\rho}{z_+}\frac{d\sigma_B}{dx\,dQ_B^2}. \tag{36}$$

3.2. Integral I_{2m}

Calculation of the integrand is performed in the approximation $\chi_2 = 0$, which corresponds to the emission of a collinear photon along the final-electron momentum. In this case, the variables take the values

$$r_2 = \frac{1-z}{1-z_+}, \quad q_2^2 = -r_2 Q_B^2, \quad x_2' = \frac{xyr_2}{1-r_2(1-y)}.$$

After integrating over the hard-photon angular variables, the integral I_{2m} is represented as

$$I_{2m} = \frac{1}{Q_B^4}\int_0^{z_m} \frac{dz}{z_+ - z}\alpha_2^2 N_2 \Sigma_2(z), \tag{37}$$

$$\Sigma_2(z) = \Sigma_{2ll}R_{ll} + \Sigma_{2lt}R_{lt} + \Sigma_{2tt}(R_{tt} - R_{nn}),$$

$$\Sigma_{2tt} = bQ_B^2(r_2 G_s + B_{3s}),$$

$$\Sigma_{2lt} = -V\sqrt{\frac{xyb}{\tau}}\left\{\left(y-1+\frac{1}{r_2}\right)B_{4s}\right.$$

$$\left.+\left(a-3b+\frac{1}{r_2}\right)B_{3s} + 2[1+(a-b)r_2]G_s\right\},$$

$$\Sigma_{2ll} = \frac{V}{2\tau r_2}\{(a-b)[1-r_2(1-y)]B_{4s}$$

$$-2b[1+(2a-b)r_2]B_{3s}$$

$$-[2abr_2^2 - (1+ar_2-br_2)^2]G_s\},$$

$$G_s = xyB_{1s} - \frac{b}{1-r_2(1-y)}B_{2s}, \quad \alpha_2 = \alpha(q_2^2),$$

$$N_2 = \frac{4\tau}{(z+xyr_2)^2}, \quad B_{is} = B_i(q_2^2, x_2'), \quad i = 1, 2, 3, 4.$$

The contribution containing the infrared divergence is extracted explicitly in a similar manner as for the I_{1m} integral. At a value of $z = z_+$, we have

$$r_2 = 1, \quad \alpha_2 = \alpha, \quad N_2 = 4\tau/y^2, \quad x_2' = x.$$

The integral I_{2m} is then rewritten as

$$I_{2m} = \frac{1}{Q_B^4}\int_0^{z_+} \frac{dz}{z_+ - z}$$

$$\times\left[\alpha_2^2 N_2\Sigma_2(z) - \alpha^2\frac{4\tau}{y^2}\Sigma_2(z_+)\right] + \frac{Vx}{\pi y}\ln\frac{\rho}{z_+}\frac{d\sigma_B}{dx\,dQ_B^2}. \tag{38}$$

The radiative corrections due to the virtual photon exchange and real soft-photon emission (with energy

$\omega < \Delta\varepsilon$) can be related to the Born cross section as[1]

$$\frac{d\sigma^{(S+V)}}{dxdQ_B^2} = \delta^{SV}\frac{d\sigma_B}{dxdQ_B^2}, \tag{39}$$

where the factor δ^{SV} is [19]

$$\delta^{SV} = \frac{\alpha}{\pi}\Big[(L-1)\ln\frac{(\Delta\varepsilon)^2}{\varepsilon_1\varepsilon_2} + \frac{3}{2}L - \frac{1}{2}\ln^2\frac{\varepsilon_1}{\varepsilon_2} - \frac{\pi^2}{6} - 2 - f\Big(\cos^2\frac{\theta}{2}\Big)\Big], \quad L = \ln\frac{Q_B^2}{m^2}, \tag{40}$$

$\varepsilon_1(\varepsilon_2)$ is the initial (final) electron energy, and θ is the electron scattering angle in the coordinate frame, where

$$\mathbf{k}_1 - \mathbf{k}_2 + \mathbf{p} = 0.$$

The function f is defined as

$$f(x) = \int_0^x \frac{dt}{t}\ln(1-t).$$

The quantities ε_1, ε_2, and θ can be expressed in terms of the invariant variables as

$$\varepsilon_1 = \frac{V(1-xy)}{2\sqrt{V(\tau+z_+)}}, \quad \varepsilon_2 = \frac{V(1-z_+)}{2\sqrt{V(\tau+z_+)}}, \quad \cos^2\frac{\theta}{2} = \frac{1-y-xy\tau}{(1-xy)(1-z_+)}. \tag{41}$$

The radiative correction δ^{SV} is finally rewritten as

$$\delta^{SV} = \frac{\alpha}{2\pi}\Big\{-1-\frac{\pi^2}{3} - 2f\Big[\frac{1-y-xy\tau}{(1-xy)(1-z_+)}\Big] - \ln^2\frac{1-xy}{1-z_+} + (L-1)\Big(3 + 2\ln\frac{\rho^2}{(1-xy)(1-z_+)}\Big)\Big\}. \tag{42}$$

3.3. Integral I_R

To calculate this integral, we use the results in [19]. In addition to the integrals calculated in that paper, we need the integrals

$$\int\frac{d^3k}{2\pi\omega}F(z,r)\chi_1, \quad \int\frac{d^3k}{2\pi\omega}F(z,r)\chi_1^2. \tag{43}$$

To calculate these integrals, we write the hard-photon phase space measure as

$$\frac{d^3k}{2\pi\omega} = \frac{Q_B^2}{2\sqrt{y^2+4a}}\frac{d\varphi}{2\pi}dzdr. \tag{44}$$

Because the function F is independent of the φ variable in our case, we can integrate over this variable. We do this in the coordinate frame specified above. The results are

$$i_1 = \int\frac{d\varphi}{2\pi}\chi_1 = \frac{Q_B^2}{y^2+4a} \times[(2-y)(y-c) - (1-r)(y+2a)], \qquad i_2 = \int\frac{d\varphi}{2\pi}\chi_1^2 = \frac{1}{2}\Big[3i_1^2 - \frac{Q_B^4(1-xy)^2}{y^2+4a}(r-r_1)^2\Big]. \tag{45}$$

After simple calculations, the integral I_R is (with the contributions proportional to the R_{ln} and R_{tn} components omitted)

$$\begin{aligned} I_R = \frac{1}{2Q_B^4}\sum_{i=1}^{4}\sum_{m,n}R_{mn}\Big\{&\frac{L_1}{1-xy}\int_0^{z_m}\frac{dz}{1-r_1}G_i^{mn}(z,r_1) + \frac{L_2}{1-z_+}\int_0^{z_m}\frac{dz}{1-r_2}\tilde{G}_i^{mn}(z,r_2) \\ &+ \frac{1}{1-xy}\int_0^{z_m}dz\int_{r_-}^{r_+}\frac{dr}{|r-r_1|}\Big[\frac{G_i^{mn}(z,r)}{1-r} - \frac{G_i^{mn}(z,r_1)}{1-r_1}\Big] \\ &+ \frac{1}{1-z_+}\int_0^{z_m}dz\int_{r_-}^{r_+}\frac{dr}{|r-r_2|}\Big[\frac{\tilde{G}_i^{mn}(z,r)}{1-r} - \frac{\tilde{G}_i^{mn}(z,r_2)}{1-r_2}\Big] \\ &+ \frac{Q_B^2}{\sqrt{y^2+4a}}\int_0^{z_m}dz\int_{r_-}^{r_+}dr\frac{\alpha^2}{r^2} \times B_i[C_{0i}^{mn}(z,r) + i_1C_{1i}^{mn}(z,r) + i_2C_{2i}^{mn}(z,r)]\Big\}, \end{aligned} \tag{46}$$

where

$$L_1 = \ln\frac{Q_B^2(1-xy)^2}{m^2xy(\tau+z_+)}, \quad L_2 = \ln\frac{Q_B^2(1-z_+)^2}{m^2xy(\tau+z_+)}, \tag{47}$$

$$r_\pm(z) = \frac{1}{2xy(\tau+z_+)} \times[2xy(\tau+z) + (z_+-z)(y\pm\sqrt{y^2+4a})],$$

$$G_i^{mn}(z,r) = \frac{\alpha^2}{r^2}(1-r)B_iA_i^{mn}(z,r),$$

$$\tilde{G}_i^{mn}(z,r) = \frac{\alpha^2}{r^2}(1-r)B_iB_i^{mn}(z,r),$$

[1] We note that the vacuum polarization effects are included in the Born cross section through the dependence of the coupling constant α on the virtual-photon momentum.

with $m, n = l, t, n$. We note that the structure functions B_i are functions of two independent variables

$$q^2 = -rQ_B^2, \quad x' = \frac{xyr}{z + xyr}.$$

The expressions for the coefficients A_i^{mn}, B_i^{mn}, C_{ki}^{mn}, $k = 0, 1, 2$, are given in Appendix A. The contributions proportional to the R_{ln} and R_{tn} components are considered in more detail in Appendix B.

We now briefly discuss the singularities in the I_R integral. The value $r = 1$ corresponds to the real soft-photon emission (there is an infrared divergence at this point), and the value $r = r_1(r_2)$ corresponds to the emission of a collinear photon along the initial- (final-) electron momentum (the so-called collinear divergence). The singularity at the point $z = z_+$ is the infrared one. The divergence at $r = 1$ is nonphysical. It arises during the integration procedure due to the separation of the poles in the expression $(\chi_1\chi_2)^{-1}$. It is necessary to explicitly extract the collinear and infrared divergences in the above formula.

The integrand in the above expression can be written in the form that does not explicitly contain the infrared divergences if we add term (39) to it. For this, we use the transformations

$$\frac{G(z, r_i)}{1 - r_i}\ln\frac{\varphi_i(x, y)}{xy(\tau + z_+)} + \int_{r_-}^{r_+}\frac{dr}{|r - r_i|}\left[\frac{G(z, r)}{1 - r} - \frac{G(z, r_i)}{1 - r_i}\right] = P\int_{r_-}^{r_+}\frac{dr}{(1 - r)|r - r_i|}[G(z, r) - G(z, r_i)], \quad i = 1, 2, \tag{48}$$

where

$$\varphi_1(x, y) = (1 - xy)^2, \quad \varphi_2(x, y) = (1 - z_+)^2,$$

and the symbol P denotes the principal value of the integral. The total radiative correction (which is the sum of the contribution due to the hard-photon emission and the contribution due to the real soft-photon emission and virtual-photon contribution) to the part of the differential cross section caused by the tensor polarization of the target is written as

$$\frac{d\sigma}{dxdQ_B^2} = \frac{d\sigma_B}{dxdQ_B^2} + \delta^{\text{tot}}, \tag{49}$$

where

$$\delta^{\text{tot}} = \frac{\alpha}{2\pi}\left\{3L + 2(L - 1)\ln\frac{z_+^2}{(1 - xy)(1 - z_+)} - \ln^2\frac{1 - xy}{1 - z_+} - 4 - \frac{\pi^2}{3} - 2f\left[\frac{b}{(1 - xy)(1 - z_+)}\right]\right\}\frac{d\sigma_B}{dxdQ_B^2} + \frac{\alpha y}{xQ_B^4}\int_0^{z_+}\frac{dz}{z_+ - z}\left[\alpha_1^2 N_1\Sigma_1(z) + \alpha_2^2 N_2\Sigma_2(z) - \alpha^2\frac{8\tau}{Vy^2}\Sigma_1(z_+)\right] + \frac{\alpha y}{2xVQ_B^4} \tag{50}$$

$$\times\sum_{i=1}^{4}\sum_{m,n}R_{mn}\left\{L\int_0^{z_+}\frac{dz}{z_+ - z}[G_i^{mn}(z, r_1) - G_i^{mn}(z_+, 1) - \tilde{G}_i^{mn}(z, r_2) + \tilde{G}_i^{mn}(z_+, 1)] + \frac{Q_B^2}{\sqrt{y^2 + 4a}}\right.$$

$$\left.\times\int_0^{z_+}dz\int_{r_-}^{r_+}dr\frac{\alpha^2}{r^2}B_i[C_{0i}^{mn}(z, r) + i_1C_{1i}^{mn}(z, r) + i_2C_{2i}^{mn}(z, r)] + R_i^{mn}\right\}.$$

The term R_i^{mn} has different forms depending on the integration region of the variable r. In the regions $r_- \le r \le r_1$ and $r_2 \le r \le r_+$ (where $r \ne 1$, and therefore the divergence at the point $r = 1$ is absent), the function R_i^{mn} has the form

$$R_i^{mn} = \frac{1}{1 - xy}\int_0^{z_+}dz\int_{r_-}^{r_+}\frac{dr}{(1 - r)|r - r_1|}[G_i^{mn}(z, r) - G_i^{mn}(z, r_1)] + \frac{1}{1 - z_+} \tag{51}$$

$$\times\int_0^{z_+}dz\int_{r_-}^{r_+}\frac{dr}{(1 - r)|r - r_2|}[\tilde{G}_i^{mn}(z, r) - \tilde{G}_i^{mn}(z, r_2)].$$

In the region $r_1 < r < r_2$, we have

$$R_i^{mn} = \int_0^{z_+} dz \ln\frac{1-r_-}{r_+-1}\Bigg\{g_{i1}^{mn}(z,1) - f_{i1}^{mn}(z,1) + \frac{1}{z_+-z}$$
$$\times [g_{i0}^{mn}(z,1) - g_{i0}^{mn}(z,r_1) + f_{i0}^{mn}(z,1) - f_{i0}^{mn}(z,r_2)]\Bigg\}$$
$$+ \int_0^{z_+} dz \int_{r_-}^{r_+} \frac{dr}{1-r}\{g_{i1}^{mn}(z,r) - g_{i1}^{mn}(z,1)$$
$$- f_{i1}^{mn}(z,r) + f_{i1}^{mn}(z,1)$$
$$+ \frac{1}{1-xy}[F^{mn}(z,r) - F^{mn}(z,1)]$$
$$- \frac{1}{1-z_+}[\tilde{F}^{mn}(z,r) - \tilde{F}^{mn}(z,1)]\Bigg\}, \quad (52)$$

where we introduce the notation

$$G_i^{mn}(z,r) = g_{i0}^{mn}(z,r) + \Delta_1 g_{i1}^{mn}(z,r),$$
$$\tilde{G}_i^{mn}(z,r) = f_{i0}^{mn}(z,r) + \Delta_2 f_{i1}^{mn}(z,r),$$
$$F^{mn}(z,r) = \frac{1}{r-r_1}[g_{i0}^{mn}(z,r) - g_{i0}^{mn}(z,r_1)],$$
$$\tilde{F}^{mn}(z,r) = \frac{1}{r-r_2}[f_{i0}^{mn}(z,r) - f_{i0}^{mn}(z,r_2)], \quad (53)$$
$$\Delta_1 = (1-xy)r - a - b - z,$$
$$\Delta_2 = (1-y+xy)r + z - 1.$$

In obtaining the above formula, we use the relation

$$P\int_{r_-}^{r_+} \frac{dr}{1-r}\Psi(r) = \int_{r_-}^{r_+} \frac{dr}{1-r}[\Psi(r) - \Psi(1)]$$
$$+ \Psi(1)\ln\frac{1-r_-}{r_+-1}. \quad (54)$$

We finally consider the part of the integral I caused by the R_{ln} and R_{tn} components of the deuteron quadrupole polarization tensor. As stated above, these components do not contribute to the cross section treated in the Born approximation. If these terms are integrated over the whole region of the φ variable, then these integrals are equal to zero as well (because only one plane remains after such integration). We discuss this problem in more detail in Appendix B.

We note that the integration limits for the variable z in formula (50) are given somewhat schematically. This integral contains two contributions (we neglect here the contribution of the radiative tail from the quasielastic scattering). One of them is the so-called inelastic contribution; the integration region for it in the variables r and z is presented in Fig. 1 by the dashed triangle. The integration over z for this contribution must be carried out from

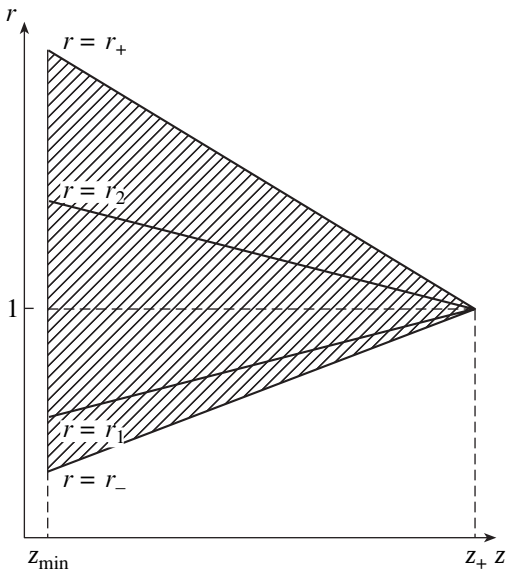


Fig. 1. The integration domain in r and z.

$$z_{\min} = \frac{M_{th}^2 - M^2}{V}$$

to z_+, where M_{th} is the inelastic threshold ($M_{th} = M + m_\pi$). The second contribution, related to the radiative tail of the elastic peak, is given by the interval

$$z = 0, \quad r_-(0) \le r \le r_+(0).$$

The contribution of the elastic radiative tail to the total radiative correction δ^{tot} (i.e., inclusion of radiative corrections to the elastic *ed* scattering) can be obtained from formula (30) by a simple substitution in the hadronic tensor,

$$B_i(q^2, x') \longrightarrow -\frac{1}{q^2}\delta(1-x')B_i^{(el)}, \quad i = 1, 2, 3, 4, \quad (55)$$

where $B_i^{(el)}$ are expressed in terms of the deuteron electromagnetic form factors as

$$B_1^{(el)} = \eta q^2 G_M^2,$$
$$B_2^{(el)} = -2\eta^2 q^2\left[G_M^2 + \frac{4G_Q}{1+\eta}\left(G_C + \frac{\eta}{3}G_Q + \eta G_M\right)\right], \quad (56)$$
$$B_3^{(el)} = 2\eta^2 q^2 G_M(G_M + 2G_Q),$$
$$B_4^{(el)} = -2\eta q^2(1+\eta)G_M^2, \quad \eta = -q^2/4M^2.$$

Here, G_C, G_M, and G_Q are the deuteron charge monopole, magnetic dipole, and quadrupole form factors, respectively. These form factors have the normalizations

$$G_C(0) = 1, \quad G_M(0) = (M/m_n)\mu_d,$$
$$G_Q(0) = M^2 Q_d,$$

where m_n is the nucleon mass and $\mu_d(Q_d)$ is the deuteron magnetic (quadrupole) moment, with the values

$$\mu_d = 0.857, \quad Q_d = 0.2859 \text{ fm}^2.$$

After substitution of $B_i^{(el)}$ in formula (30), we have to perform trivial integration over the z variable using the delta function

$$\delta(1 - x') = xyr\delta(z).$$

4. NUMERICAL ESTIMATE

We calculate the radiative corrections for the kinematical conditions of the HERMES experiment [10]. The energy of the positron beam is 27.6 GeV. The HERMES installation has provided the first direct measurement of the structure function b_1 in the kinematic range $0.002 < x < 0.85$ and $0.1 < Q^2 < 20$ GeV^2. A cylindrical target cell confines the polarized gas along the positron beam line, where a longitudinal magnetic field provides the quantization axis for the nuclear spin. The corresponding tensor atomic polarization is $T = 0.83$ (see Appendix C for the definition of this quantity).

The analysis of the experimental data was performed in the approximation $b_3 = b_4 = 0$. In the numerical estimate below, we also neglect these functions.

The deuteron spin-dependent structure function b_1 is extracted from the measured tensor asymmetry A_{zz} via the relation [10]

$$b_1 = -\frac{3}{2}A_{zz}\frac{(1+\gamma^2)F_2^d}{2x(1+R)}, \tag{57}$$

where the deuteron spin-independent structure function F_1^d is expressed in terms of the ratio

$$R = \frac{\sigma_L}{\sigma_T} = \frac{F_2^d(1 + 4M^2x^2/Q^2)}{2xF_1^d} - 1$$

(see [20]) and

$$\gamma^2 = \frac{4M^2x^2}{Q^2}$$

is a kinematic factor. Here, $\sigma_T(\sigma_L)$ is the cross section for the absorption of transversely (longitudinally) polarized virtual photons by the unpolarized target. The Born cross section of the deep inelastic scattering of the unpolarized electron beam by the unpolarized target has the form

$$\frac{d\sigma_B^{un}}{dxdQ_B^2} = \frac{4\pi\alpha^2}{xQ_B^4} \tag{58}$$
$$\times [(1 - y - xy\tau)F_2^d(x, Q^2) + xy^2F_1^d(x, Q^2)].$$

The structure functions $F_{1,2}^d$ are related to the structure functions $W_{1,2}$ (introduced in formula (7)) as

$$W_1 = 2F_1^d, \quad W_2 = 4(\tau/y)F_2^d.$$

The deuteron spin-independent structure function

$$F_2^d = \frac{F_2^p(1 + F_2^n/F_2^p)}{2}$$

is calculated using parametrizations for the proton structure functions F_2^p [21] and the ratio F_2^n/F_2^p [22]. The deuteron spin-dependent structure function b_2 is also extracted from the experiment using the Callan–Gross relation

$$b_2 = 2x\frac{1+R}{1+\gamma^2}b_1. \tag{59}$$

According to the preliminary results of the HERMES experiment, the tensor asymmetry can be parametrized as [23]

$$A_{zz} = -1.56 \times 10^{-2}(1 - 1.74x - 1.45\sqrt{x}). \tag{60}$$

The influence of the radiative correction on the spin-dependent part of the Born cross section is shown in Fig. 2 as a function of the variable x for various Q^2 values. Inclusion of the radiative correction shifts the zero value of b_1 and b_2 to the region of smaller x (see Figs. 2c and 2d). In the range of low x ($x \sim 10^{-3}$–10^{-2}), the value of the radiative correction changes from 10 to 30% compared with the Born contribution. This region is of the utmost importance for b_1 measurements. According to the theoretical predictions in [11–13], the structure function b_1 increases very rapidly in this region, and this fact was confirmed in the HERMES experiment [10].

From our estimate, we conclude that the radiative corrections to process (1) are not small, especially for

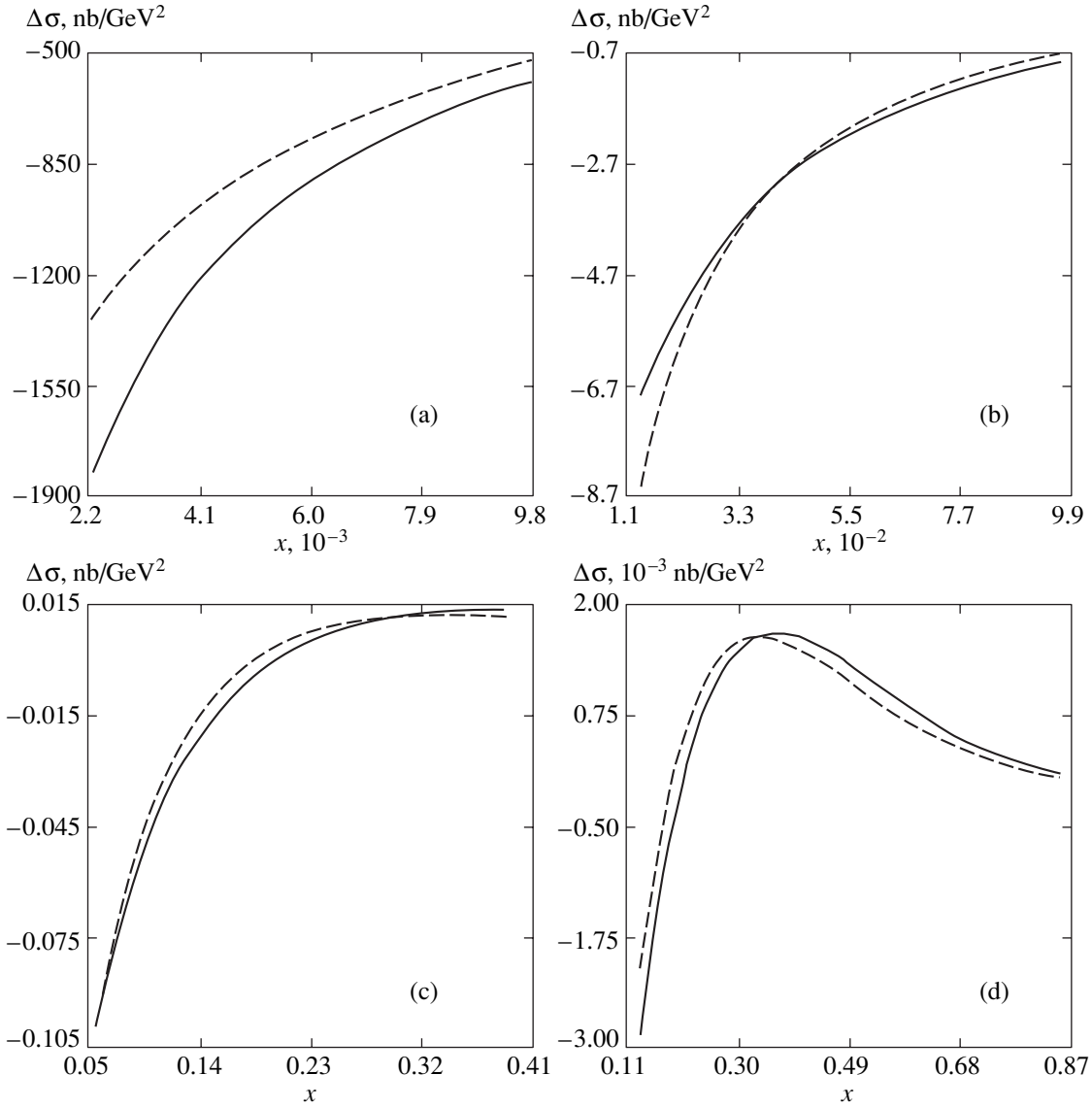


Fig. 2. The spin-dependent part of the cross section calculated for the kinematical conditions of the HERMES experiment [10]. The solid line is the Born approximation, the dotted line corresponds to the inclusion of the radiative corrections. The Q^2 values are as follows: (a) 0.1, (b) 1, (c) 4, and (d) 10 GeV^2.

the low-x region, and they have to be taken into account in data analysis.

ACKNOWLEDGMENTS

We wish to thank N.P. Merenkov for useful discussions and comments. We warmly acknowledge M. Contalbrigo for useful discussions on the HERA experimental conditions, as well as for sending us preliminary results on the A_{zz} parametrization.

APPENDIX A

In this Appendix, we present the formulas for the coefficients A_i^{mn}, B_i^{mn}, and C_{ji}^{mn} ($m, n = l, t, i = 1, 2, 3, 4, j = 0, 1, 2$) that determine the cross section of the hard-photon emission process (see formula (50)).

The coefficients determining the contribution proportional to the B_1 structure function are

$$A_1^{ll} = -\frac{n_1}{\tau}[(\bar{r}-\Delta_1)^2 - 2a(b+\Delta_1)],$$

$$B_1^{ll} = \frac{n_1}{\tau}\{[(2a-b)r + 1 + \Delta_2]^2 - ar(2+3ar)\},$$

$$C_{01}^{ll} = -\frac{VN}{\tau}\{(\bar{r}-\Delta_1)^2 + a[3a(1+r^2) - 2(b+\Delta_1)]\},$$

$$C_{11}^{ll} = -6N(c+2a), \quad C_{21}^{ll} = -6N\frac{\tau}{V},$$

$$A_1^{lt} = \frac{2n_1}{Md}Q_B^2(2b+\Delta_1)(\bar{r}-\Delta_1),$$

$$B_1^{lt} = \frac{2n_1}{Md}Q_B^2(\Delta_2-2br)[(a-b)r+1+\Delta_2],$$

$$C_{01}^{lt} = -4n_2Q_B^2$$
$$\times[a(1+r^2)(y+2a)-2b\bar{r}-\Delta_1(c+2a-2b)],$$

$$C_{11}^{lt} = -4n_2[(y+4a)(c+2a)-2a(\bar{r}+2b)],$$

$$C_{21}^{lt} = -8n_2\frac{\tau}{V}(y+2a),$$

$$A_1^{tt} = -\frac{n_1}{b}\frac{Q_B^2}{V}[b^2+(b+\Delta_1)^2], \tag{A.1}$$

$$B_1^{tt} = \frac{n_1}{b}\frac{Q_B^2}{V}(2b^2r^2-2br\Delta_2+\Delta_2^2),$$

$$C_{01}^{tt} = -\frac{NQ_B^2}{2b}$$
$$\times[(1+r^2)(y^2+4a-2ab)+(2b+\Delta_1)^2+\Delta_1^2],$$

$$C_{11}^{tt} = \frac{2N}{b}[b(1+y+2a-r)+(1+a)\Delta_1],$$

$$C_{21}^{tt} = -\frac{N}{d^2}[y^2+2a(2-b)].$$

The coefficients determining the contribution proportional to the B_2 structure function are

$$A_2^{ll} = \frac{n_3}{\tau}[b(1+r^2)+(1-r+ry)\Delta_1]$$
$$\times[(\bar{r}-\Delta_1)^2-2a(b+\Delta_1)],$$

$$B_2^{ll} = -\frac{n_3}{\tau}[b(1+r^2)-\Delta_2(\bar{r}-2a)]$$
$$\times\{[(2a-b)r+1+\Delta_2]^2-ar(2+3ar)\},$$

$$C_{02}^{ll} = -\frac{NV}{c}\{(7-3y)c^2$$
$$+3a(5-y+r)c+3a^2(3+r^2)-ar[5+3(a+b)^2]\},$$

$$C_{12}^{ll} = -3\tau\frac{N}{c}[4(a+c)-cy], \quad C_{22}^{ll} = -6N\frac{\tau^2}{cV},$$

$$A_2^{lt} = -2\frac{n_3Q_B^2}{Md}(2b+\Delta_1)(\bar{r}-\Delta_1)$$
$$\times[b(1+r^2)+(1-r+ry)\Delta_1],$$

$$B_2^{lt} = -2\frac{n_3Q_B^2}{Md}(\Delta_2-2br)[1+\Delta_2+(a-b)r]$$
$$\times[b(1+r^2)+(a+b-r)\Delta_2],$$

$$C_{02}^{lt} = \frac{NV^2}{Mdc}\{2ac[y\bar{r}+(3b+a)(1+r)-y-8a]$$
$$-c^2[2a+(2-y)(y+4a)]+2a[2a(b-a+r)$$
$$+(y+2a)(r-a(1+r^2)+r(a+b)^2)]\}, \tag{A.2}$$

$$C_{12}^{lt} = -4n_2\frac{\tau}{c}[2a(1-3r+2\bar{r}+4c)+cy(2+b-a)],$$

$$C_{22}^{lt} = -8n_2\frac{\tau^2}{cV}(y+2a),$$

$$A_2^{tt} = \frac{xy}{b}n_3Z_1[2b^2+\Delta_1(2b+\Delta_1)],$$

$$B_2^{tt} = -\frac{xy}{b}n_3Z_2[2b^2r^2+\Delta_2(\Delta_2-2br)],$$

$$C_{02}^{tt} = \frac{VN}{2bc}\{-2c^2[a+(1+a)(2-y)]$$
$$+c[(3-2y+a^2+b^2)(\bar{r}-2a)$$
$$+4(ab+b-a^2)+4r(a-b^2)]-2a[(r-a)^2+b^2]$$
$$+(1+2a-2b+a^2+b^2)[r-a(1+r^2)+(a+b)^2r]\},$$

$$C_{12}^{tt} = -\frac{VN}{2cd^2}$$
$$\times\{c[1+7a(1+a)-b(1+b)+(a+b)(a^2+b^2)]$$
$$+4a[(a-b)(1-r)+a^2+b^2-r]\},$$

$$C_{22}^{tt} = -\frac{N\tau}{cd^2}[y^2+2a(2-b)].$$

The coefficients determining the contribution proportional to the B_3 structure function are

$$A_3^{ll} = n_3\frac{c}{\tau}\{(a+\bar{r})[2Z_1+r\Delta_1(2a+r-\Delta_1)]$$
$$-\Delta_1[r^2(r-\Delta_1)+2(b+\Delta_1)+r(a+b)(a+r-\Delta_1)]\},$$

$$B_3^{ll} = -n_3\frac{c}{\tau}\{2Z_2[1+(2a-b)r+\Delta_2]$$
$$+3a\Delta_2[(b-a)r-1-\Delta_2]\},$$

$$C_{03}^{ll} = \frac{VN}{2}[c(6a-16+9y)+6a(y-3-r)],$$

$$C_{13}^{ll} = -3\tau N(2-y), \quad C_{23}^{ll} = 0,$$

$$A_3^{lt} = -n_3\frac{cQ_B^2}{Md}\{2Z_1(3b-a-r)+\Delta_1[4r(1+b^2+3ab)$$
$$-2a(1+r^2)+c(ar-3+5br)]\},$$
$$B_3^{lt} = n_3\frac{cQ_B^2}{Md}\{2Z_2[(3b-a)r-1]+\Delta_2[(1+ar)(a-6b)$$
$$-(a+3b)(r^2+\Delta_2)+r(b^2-1)+b+\Delta_2(3r-2b)]\},$$
$$C_{03}^{lt} = n_2V\{4a[2b\bar{r}-y^2+4(b^2-a)]$$
$$-c[3y(2-y)+8a(1+a+2b)]\}, \quad \text{(A.3)}$$
$$C_{13}^{lt} = -2N\frac{M}{d}(2-y)(y+2a), \quad C_{23}^{lt} = 0,$$
$$A_3^{tt} = n_3xy\frac{c}{b}$$
$$\times\{(\Delta_1-2b)[b(1+r^2)+(1-r+ry)\Delta_1]$$
$$+b\Delta_1[1+r(b-a+\Delta_1)]\},$$
$$B_3^{tt} = n_3xy\frac{c}{b}$$
$$\times\{(2br-\Delta_2)[b(1+r^2)+(1-r-y)\Delta_2]$$
$$+b\Delta_2[r(b-a+r)-\Delta_2]\},$$
$$C_{03}^{tt} = \frac{VN}{2b}\{3b-a-(a^2+b^2)(2+a+b)$$
$$+r[y^2+2y(2b-a)+2a(3-a)]$$
$$+c[y(1+y+3a)-4(1+a)-2ab]\},$$
$$C_{13}^{tt} = -\frac{VN}{2d^2}(2-y)[y^2+2a(2-b)], \quad C_{23}^{tt} = 0.$$

The coefficients determining the contribution proportional to the B_4 structure function are

$$A_4^{ll} = n_3\frac{c^2}{\tau}\{(b-a)(1+r^2)+\Delta_1[1+r(2a-b)]\},$$
$$B_4^{ll} = n_3\frac{c^2}{\tau}[(a-b)(1+r^2)+\Delta_2(a+\bar{r})],$$
$$C_{04}^{ll} = -\frac{cNV}{2}[1+3(b-a)], \quad C_{14}^{ll} = C_{24}^{ll} = 0,$$
$$A_4^{lt} = n_3\frac{c^2Q_B^2}{Md}\{2b(1+r^2)+\Delta_1[1-r(3b-a)]\}, \quad \text{(A.4)}$$
$$B_4^{lt} = n_3\frac{c^2Q_B^2}{Md}[-2b(1+r^2)+\Delta_2(\bar{r}-2b)],$$
$$C_{04}^{lt} = -c\frac{NV^2}{2Md}[1+4ab-(a-b)^2], \quad C_{14}^{lt} = C_{24}^{lt} = 0,$$
$$A_4^{tt} = -xyrc^2\Delta_1n_3, \quad B_4^{tt} = -xyc^2\Delta_2n_3,$$
$$C_{04}^{tt} = -\frac{cNV}{2}(y+2a), \quad C_{14}^{tt} = C_{24}^{tt} = 0.$$

We here use the notation

$$c = z+xyr, \quad \bar{r} = a-b+r, \quad n_1 = \frac{N}{2}\frac{1+r^2}{1-r}VQ_B^2,$$
$$n_2 = \frac{NV}{2Md}, \quad n_3 = \frac{N}{2c}\frac{V^2}{1-r},$$
$$d^2 = bQ_B^2, \quad \Delta_1 = (1-xy)r-a-b-z,$$
$$\Delta_2 = (1-y+xy)r+z-1, \quad N = \frac{4\tau}{Vc^2},$$
$$Z_1 = b(1+r^2)+\Delta_1(1-r+yr),$$
$$Z_2 = b(1+r^2)+\Delta_2(1-y-r).$$

APPENDIX B

In this Appendix, we consider the part of the integral I that is caused by the R_{ln} and R_{tn} components of the deuteron quadrupole polarization tensor (these components do not contribute to the differential cross section treated in the Born approximation). We define the integral caused by the R_{ln} component as

$$I_{ln} = \int\frac{d^3k}{2\pi\omega}\Sigma_{ln}(z,r,\varphi)R_{ln}, \quad \text{(B.1)}$$

with

$$\Sigma_{ln}(z,r,\varphi) = \frac{\alpha^2(q^2)}{Q_B^4}\frac{2VN}{Mr^2}nq$$
$$\times\left(\frac{P_{1ln}}{\chi_1}-\frac{P_{2ln}}{\chi_2}+U_{0ln}+U_{1ln}\chi_1\right),$$
$$P_{1ln} = \frac{V}{1-r}\left\{cgxy(1+r^2)B_1\right.$$
$$+g[c(1-r(1-y))+a(1+r^2)-4fr]B_2$$
$$+\left[2a-fr+\frac{1}{2}(3(1-r+yr)+2ar)\right]B_3$$
$$\left.+\frac{1}{2}c[1+(a-b)r]B_4\right\},$$

$$P_{2ln} = -\frac{V}{1-r}\Big\{-xy(1+r^2)(c+2ar)B_1 \qquad \text{(B.2)}$$

$$+\frac{2}{c}\Big[-ar(a(1+r^2)-4fr)+\frac{c^2}{2}(1-y-r)$$

$$+\frac{c}{2}(4fr-a(1+3r^2+2yr-2r))\Big]B_2$$

$$+\Big[r(f-2ar)-\frac{c}{2}(2a+3(r+y-1))\Big]B_3$$

$$-\frac{c}{2}(a-b+r)B_4\Big\},$$

$$U_{0ln} = 2g(cB_1+\tau B_2)+2\tau(2-y)(B_2+B_3),$$

$$U_{1ln} = \frac{4\tau}{V}\Big(B_1+\frac{\tau}{c}B_2\Big),$$

and

$$c = z+xyr, \quad f = 1+(1-y)^2,$$

$$g = 1+2a/c, \quad nq = S_\mu^{(n)}q_\mu.$$

The second integral, caused by the R_{tn} component, is defined as

$$I_{th} = \int\frac{d^3k}{2\pi\omega}\Sigma_{tn}(z,r,\varphi)R_{tn}, \qquad \text{(B.3)}$$

where the integrand is

$$\Sigma_{tn}(z,r,\varphi) = \frac{\alpha^2(q^2)}{Q_B^4}\frac{2VN}{dr^2(r-1)}nq$$

$$\times\Big(\frac{P_{1tn}}{\chi_1}-\frac{P_{2tn}}{\chi_2}+U_{0tn}+U_{1tn}\chi_1\Big),$$

$$P_{1tn} = Q_B^2\Big\{\bar f xy(1+r^2)B_1$$

$$-\frac{\bar f}{c}[a(1+r^2)-4r(f+4y)]B_2$$

$$+\bar f[1+r(y-1)](B_2+B_3)$$

$$+brc(B_3+B_4)+2br(1-y)B_3\Big\},$$

$$P_{2tn} = -Q_B^2\Big\{xy\bar g(1+r^2)B_1+\frac{\bar g}{c}[a(1+r^2)-4fr]B_2$$

$$+\bar g(r-1+y)(B_2+B_3) \qquad \text{(B.4)}$$

$$-bc(B_3+B_4)-2br(1-y)B_3\Big\},$$

$$U_{0tn} = (r-1)\Big[-2xy\bar f\Big(B_1+\frac{\tau}{c}B_2\Big)$$

$$+(2-y)(2a+y)(B_2+B_3)\Big],$$

$$U_{1tn} = \frac{1}{V}\Big\{(r-1)(2a+y)G_1-\bar f G_2$$

$$-\frac{y-2}{Mxy}[M(2a+y)-2\tau d](B_2+B_3)$$

$$-\frac{d}{Mxy}c-2a(r-2)]G_2\Big\},$$

$$G_1 = 3B_1+\frac{2\tau}{c}B_2+\frac{c}{2xyr}(B_2+2B_3+B_4),$$

$$G_2 = -B_1-\frac{c}{2xyr}(B_2+2B_3+B_4),$$

and

$$d^2 = bQ_B^2, \quad \bar f = b-a-z+r(1-xy),$$

$$\bar g = z-1+r(a-b+xy).$$

As before, we calculate the above integrals in the center-of-mass system of the hard photon and the undetected hadron system:

$$\mathbf{k}_1-\mathbf{k}_2+\mathbf{p} = 0.$$

The electron momenta $\mathbf{k}_1$ and $\mathbf{k}_2$ define the xz plane, the z axis is directed along the deuteron momentum $\mathbf{p}$. Then, the hard-photon momentum $\mathbf{k}$ is determined by the azimuthal (φ) and polar (θ) angles, and the phase space of the hard photon can be written as

$$\frac{d^3k}{2\pi\omega} = \frac{Q_B^2}{2\sqrt{y^2+4a}}\frac{d\varphi}{2\pi}dzdr, \qquad \text{(B.5)}$$

where ω is the hard-photon energy.

The quantity nq can be written in this coordinate system as $nq = \bar n\sin\varphi$, where $\bar n$ is a factor independent

of φ. Then, the integration over the φ variable in the region $(0, 2\pi)$ leads to the result

$$I_{ln} = I_{tn} = 0.$$

Therefore, the R_{ln} and R_{tn} components of the deuteron quadrupole polarization tensor do not contribute to the differential cross section of deep inelastic scattering of the unpolarized electron beam by the tensor polarized target. This is because only the scattered-electron variables are measured (this corresponds to the HERA experimental conditions, for example).

If the hard photon is detected, then I_{ln} and I_{tn} survive and the expressions for Σ_{ln} and Σ_{tn} have to be taken into account.

APPENDIX C

In this Appendix, we give some formulas describing the polarization state of the deuteron target in different cases. In the case of an arbitrary polarization of the target, it is described by the general spin-density matrix (defined by eight parameters in general), which in the coordinate representation has the form

$$\rho_{\mu\nu} = -\frac{1}{3}\left(g_{\mu\nu} - \frac{p_\mu p_\nu}{M^2}\right) - \frac{i}{2M}\varepsilon_{\mu\nu\lambda\rho}s_\lambda p_\rho + Q_{\mu\nu}, \quad (C.1)$$
$$Q_{\mu\nu} = Q_{\nu\mu}, \quad Q_{\mu\mu} = 0, \quad p_\mu Q_{\mu\nu} = 0,$$

where p_μ is the deuteron 4-momentum, and s_μ and $Q_{\mu\nu}$ are the deuteron polarization 4-vector and the deuteron quadrupole polarization tensor.

In the deuteron rest frame, the above formula is written as

$$\rho_{ij} = \frac{1}{3}\delta_{ij} + \frac{i}{2}\varepsilon_{ijk}s_k + Q_{ij}, \quad ij = x, y, z. \quad (C.2)$$

This spin-density matrix can be written in the helicity representation using the relation

$$\rho_{\lambda\lambda'} = \rho_{ij}e_i^{(\lambda)*}e_j^{(\lambda')}, \quad \lambda, \lambda' = +, -, 0, \quad (C.3)$$

where $e_i^{(\lambda)}$ are the deuteron spin functions that have the deuteron spin projection λ on the quantization axis (the z axis). They are

$$e^{(\pm)} = \mp\frac{1}{\sqrt{2}}(1, \pm i, 0), \quad e^{(0)} = (0, 0, 1). \quad (C.4)$$

The elements of the spin-density matrix in the helicity representation are related to those in the coordinate representation by

$$\rho_{\pm\pm} = \frac{1}{3} \mp \frac{1}{2}s_z - \frac{1}{2}Q_{zz}, \quad \rho_{00} = \frac{1}{3} + Q_{zz},$$
$$\rho_{+-} = -\frac{1}{2}(Q_{xx} - Q_{yy}) + iQ_{xy},$$
$$\rho_{+0} = -\frac{1}{2\sqrt{2}}(s_x - is_y) - \frac{1}{\sqrt{2}}(Q_{xz} - iQ_{yz}), \quad (C.5)$$
$$\rho_{-0} = -\frac{1}{2\sqrt{2}}(s_x + is_y) + \frac{1}{\sqrt{2}}(Q_{xz} + iQ_{yz}),$$
$$\rho_{\lambda\lambda'} = (\rho_{\lambda'\lambda})^*.$$

To obtain these relations, we use that $Q_{xx} + Q_{yy} + Q_{zz} = 0$.

The polarized deuteron target described by the population numbers n_+, n_-, and n_0 is often used in spin experiments. Here, n_+, n_-, and n_0 are the fractions of atoms with the respective nuclear spin projection on the quantization axis $m = +1$, $m = -1$, and $m = 0$. If the spin-density matrix is normalized to 1, i.e.,

$$\mathrm{Sp}\rho = 1,$$

then we have

$$n_+ + n_- + n_0 = 1.$$

Thus, the polarization state of the deuteron target is defined in this case by two parameters: the so-called V (vector) and T (tensor) polarizations,

$$V = n_+ - n_-, \quad T = 1 - 3n_0. \quad (C.6)$$

Using the definitions of the quantities $n_{\pm,0}$,

$$n_\pm = \rho_{ij}e_j^{(\pm)*}e_j^{(\pm)}, \quad n_0 = \rho_{ij}e_j^{(0)*}e_j^{(0)}, \quad (C.7)$$

we have the following relation between V and T and the parameters of the spin-density matrix in the coordinate representation (in the case where the quantization axis is directed along the z axis):

$$n_0 = \frac{1}{3} + Q_{zz}, \quad n_\pm = \frac{1}{3} \mp \frac{1}{2}s_z - \frac{1}{2}Q_{zz}, \quad (C.8)$$

or

$$T = -3Q_{zz}, \quad V = -s_z. \quad (C.9)$$

REFERENCES

1. D. Ryckbosch, hep-ex/0311021.
2. D. Adams, B. Adeva, E. Arik, *et al.* (Spin Muon Collab.), Phys. Lett. B **357**, 248 (1995); P. L. Anthony, R. G. Arnold, H. R. Band, *et al.*, Phys. Rev. D **54**, 6620 (1996).
3. S. Kumano, Preprint MKPH-T-93-03 (Univ. Mainz, Germany, 1993).

4. P. Hoodbhoy, R. L. Jaffe, and A. Manohar, Nucl. Phys. B **312**, 571 (1989).
5. F. E. Close and S. Kumano, Phys. Rev. D **42**, 2377 (1990).
6. L. Mankiewicz, Phys. Rev. D **40**, 255 (1989).
7. H. Khan and P. Hoodbhoy, Phys. Lett. B **298**, 181 (1993).
8. H. Khan and P. Hoodbhoy, Phys. Rev. C **44**, 1219 (1991).
9. S. Kumano, hep-ph/0012341.
10. M. Contalbrigo, hep-ex/0211014.
11. N. N. Nikolaev and W. Schafer, Phys. Lett. B **398**, 245 (1997).
12. J. Edelmann, G. Piller, and W. Weise, Phys. Rev. C **57**, 3392 (1998).
13. K. Bora and R. L. Jaffe, Phys. Rev. D **57**, 6906 (1998).
14. I. V. Akushevich and N. M. Shumeiko, J. Phys. G: Nucl. Part. Phys. **20**, 513 (1994).
15. S. Y. Choi, T. Lee, and H. S. Song, Phys. Rev. D **40**, 2477 (1989).
16. G. I. Gakh and N. P. Merenkov, Pis'ma Zh. Éksp. Teor. Fiz. **73**, 659 (2001) [JETP Lett. **73**, 579 (2001)].
17. É. A. Kuraev, N. P. Merenkov, and V. S. Fadin, Yad. Fiz. **45**, 782 (1987) [Sov. J. Nucl. Phys. **45**, 486 (1987)].
18. T. V. Kukhto and N. P. Shumeiko, Nucl. Phys. B **219**, 412 (1983).
19. É. A. Kuraev, N. P. Merenkov, and V. S. Fadin, Yad. Fiz. **47**, 1593 (1988) [Sov. J. Nucl. Phys. **47**, 1009 (1988)].
20. L. W. Whitlow *et al.*, Phys. Lett. B **250**, 193 (1990).
21. H. Abramowicz and A. Levy, hep-ph/9712415.
22. P. Amaudruz *et al.* (NMC Collab.), Nucl. Phys. B **371**, 3 (1992).
23. M. Contalbrigo, private communication.

Journal of Experimental and Theoretical Physics, Vol. 99, No. 5, 2004, pp. 915–933.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 5, 2004, pp. 1051–1072.

ATOMS, SPECTRA, RADIATION

Resonances in Electron Scattering by Molecules

S. A. Pozdneev

Lebedev Physical Institute, Russian Academy of Sciences, Moscow, 199991 Russia
e-mail: pozdneev@sci.lebedev.ru

Received October 21, 2003

Abstract—The main features of resonance scattering of electrons by molecules are described and resonances are determined on the basis of the theory of collisions in a two-body system, as well as resonances emerging as a result of collisions in a few-body system. Regularities in the emergence of such resonances and their characteristics are analyzed. The results of calculations of these resonant processes occurring during collisions of electrons with diatomic molecules, made on the basis of the quantum theory of scattering in a few-body system, are presented. The results of calculating the cross sections of resonant processes of electron collisions with molecules are compared with the available experimental data and with the results of calculations based on other approximations. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Resonances bringing a variety of features to physical phenomena exist in any field from molecular physics to elementary particle physics. The concept of resonance is one of the fundamental concepts in quantum physics. We can attach a broad physical meaning to the term resonance, including stable levels, and imply their effect on scattering processes [1, 2]. Resonances play a special role in the physics of irreversible processes. In this case, in accordance with the Poincaré theorem [3], resonances are responsible for nonintegrability of most dynamic systems.

A theoretical explanation of resonances and their parameters can be constructed on the basis of forces of interaction between particles, which are treated as elementary particles in such processes. For example, resonant processes in atomic physics are determined by the forces of interaction between electrons and nuclei, while resonances in nuclear physics are determined by forces acting between nucleons.

A resonance in scattering is any peak on the experimental curve describing the scattering cross section. The resonance is characterized by the moment, parity, spin, lifetime, etc. Collisions of electrons with molecules often result in the formation of metastable negative molecular ions, which are also traditionally referred to as molecular resonances [4–11]. In this case, since atoms move slowly as compared to electrons, the electron + molecule system can be regarded as a quasimolecule whose electron shell at each instant corresponds to a quasistationary state of such a quasimolecule. This is in accordance with the well-known adiabatic approximation in quantum mechanics. In this approximation, various electron transitions (excitation, ionization, charge transfer) are hampered for collisions of electrons, atoms, or ions with molecules under ordinary conditions. The necessary condition for such a charge transfer [5, 6] is $\Delta E \Delta\tau \sim \hbar$, where ΔE is the change in the quasimolecule energy and $\Delta\tau$ is the collision time. Thus, for slow collisions, when the value of $\Delta\tau$ is large, transitions can occur only if ΔE is small; i.e., two states of the quasimolecule before and after the collision must be close and such a process can also be treated as a resonant process. Such treatment of a resonance reveals the relation between equilibrium and dynamics, on the one hand, and the physics of dissipative processes, on the other hand [12].

The importance of resonant processes is determined by the fact that all practical applications of experimental studies are based on resonances since it is resonant processes that are characterized by large cross sections or long lifetimes as compared to nonresonant processes and play an important role in low-temperature plasmas (resonant processes determine the emergence and disappearance of excited and charged particles; i.e., they determine optical and electrical properties of a plasma), in controlled thermonuclear synthesis, mu-catalysis, and so on [4–12].

Proceeding from the theory of collisions in a two-body system in which the target molecule is regarded as a force center, the following type of resonances can be distinguished [4–11].

1. A shape resonance appears in the case when the incident electron is trapped to a quasi-stationary level separated from the level in the continuum by a centrifugal barrier formed by a combination of attractive and repulsive fields of the target molecule. This type of resonance appears only when the electron possesses an angular momentum relative to the target molecule. In the case of low-energy *s* scattering ($l = 0$), electron cannot be trapped and form resonance is absent.

2. A vibrationally excited resonance appears when the incident electron excites vibrations of the target

molecule and is temporally bound. In this case, the kinetic energy of the electron is directly transformed into the vibrational energy of motion of the nuclei of the negative molecular ion; thus, this type of resonance is associated with violation of the Born–Oppenheimer principle. The lifetimes of such resonant states are extremely long (especially for polyatomic molecules) and attain tens of microseconds.

3. An electron-excited resonance is formed when the projectile electron excites the electron system of the target molecule and also becomes temporally bound. In this case, the detachment of an electron is impossible as long as the molecule remains in the excited state. Nevertheless, an electron may still be detached if closed and open channels are coupled.

Theoretical description of such resonances appearing as a result of formation of negative metastable ions is presented in [4–11] on the basis of the theory of scattering in a two-body system. In these works, resonances are defined as complex poles of the scattering matrix continued to the nonphysical energy sheet or as poles of an analytic continuation of the Green function.

Collisions between electrons and molecules occurring without the formation of intermediate complexes as well as collision processes at thermal energies of incident electrons, in which a nonmonotonic energy dependence of scattering cross section is also observed, remain unstudied. In the latter case, the application of standard techniques for calculating cross sections is unjustified in view of violation of the Born–Oppenheimer approximation [4–6]. The application of the theory of collisions in a two-body system for calculating such processes encounters considerable difficulties since the system considered here is essentially a many-particle system [13, 14].

For this reason, we will describe resonant processes occurring during collisions of an electron with molecules by using a more consistent approach based on the quantum theory of scattering in a few-particle system [13, 14]. The main approximation in this case is that the interaction of the projectile electron with the electrons and nuclei of the target molecule is replaced by the interaction of the electrons with the atoms of the molecule, the atoms being treated as force centers. Thus, a complex many-particle system consisting of the electron and the nuclei is replaced by a system of few interacting bodies, which can be described with the help of Faddeev equations [13]. Naturally, this approximation imposes certain constraints on the energy of the projectile electron: this energy should not be higher than that the ionization energy of the atoms constituting the molecule. However, it is precisely this energy range that is interesting in connection with the presence of resonance peaks in the effective cross sections of collisions of electrons with various molecules [4–11].

In such a formalism, a resonance in a three-particle system is determined by two-particle resonances under certain conditions [1, 13, 14]. Thus, the reason for the emergence of three-particle resonances is the existence of resonant states in paired subsystems. This not very popular point of view is due to the fact that such a coupling does not exist always and cannot be determined explicitly even when it is present. This was demonstrated for the first time in nuclear physics and in elementary particle physics where the interaction between particles leading to the appearance of resonances is determined by the exchange between the particles of the same resonances; thus, resonances produce themselves [1, 2].

In atomic physics, coupling between resonances is observed for a large number of phenomena (such as scattering of electrons by molecules, coupling between clusters in biopolymer molecules, and in Bose condensate) [4–7, 14, 15]. In this type of coupling, two-particle resonances lead to a series of three-particle resonance. A peculiar feature of this phenomenon is that the stronger the two-particle resonance, the larger the number of three-particle resonances produced by it. Experiments show [1, 14, 15] that such resonant states in many-particle systems lead to anomalously high rates of chemical reactions, dynamic coupling of noninteracting particles, etc. [14–16]. The importance of studying such states is directly associated with determining the binding energy of a system of N bodies using information on subsystems of this many-particle system, i.e., the construction of dependences $E_N = f(E_{N-1}, E_{N-2},\ldots)$ and the determination of the conditions for the formation of a coupled many-particle system provided that some subsystems are not coupled [16].

The physical foundation of the effect considered here is presented in [1], where the following aspects are revealed.

1. The effect of two-particle resonances on the spectrum of a three-particle system is clearly manifested; i.e., a two-particle resonance can radically reconstruct the discrete spectrum of three particles. However, not every two-particle resonant state can reconstruct the spectrum of three particles, but only the state whose size $r_{\text{res}} \propto (2m_{ij}|e_0|)^{-1/2}$ is much larger than the range r_0 of its action (e_0 is the binding energy and m_{ij} is the reduced mass of a pair of particles. Such a resonance can only be an s resonance ($l = 0$) since such resonant states strongly differ in size from other types of resonant states. For $e \longrightarrow 0$, size $r_{\text{res}} \longrightarrow \infty$. The size of a resonant state is manifested in the scattering of particles in the form of a large scattering length a, which is equal to the size of this resonant state for small e_0. Analyzing resonant states from the standpoint of their size, we can observe that all these states sharply differ from the resonance considered above. For example, the state occupied by the system in a partial wave with $l \neq 0$ has a size on the order of the range of forces due to the centrifugal barrier; a compound resonance is not large either. Thus, a two-particle s level with a small binding energy occupies an exceptional position among resonant states as regards its size.

2. Three-particle levels are stable and their number is proportional to $\ln(|a|/r)$. It can be proved [1, 13–15] that the interaction responsible for the emergence of these levels has the form $U \propto A/R^2$, where

$$R^2 = \frac{2}{3}(r_1^2 + r_2^2 + r_3^2),$$

r_i is the distance between a pair of particle, and is operative in the interval (r_0, a) (Fig. 1). In the general case, the constant A of this interaction is a function of quantum numbers of the three-particle state, angular momentum, parity, and symmetry relative to the transposition of the particles. The value of A is estimated in [1, 13–15]. The strongest attraction should be observed for the orbital angular momentum $L = 0$ for three particles since centrifugal forces are absent in this case. The symmetry of this state must be maximal; otherwise, the wavefunction has nodes and the coupling becomes weaker.

3. Centrifugal forces suppress the effect.

4. Such states possess the maximal symmetry.

5. Triple forces do not influence on the effect.

6. The addition of a particle to the three-particle system suppresses the effect.

7. The particle charge has no influence on the effect, which is manifested less clearly in this case.

8. For particles with spins, the effect is also pronounced less clearly.

It should be noted that such peculiar states of three particles are independent of the specific form of the potential (i.e., independent of the forces of interaction between particles) and are universal in the sense that these states reflect only the fact of existence of a resonance. Thus, irrespective of the form of pair forces between the particles, if it leads to a low-energy two-particle *s* resonance, this automatically leads to the formation of a family of three-particle resonances. Consequently, the reason for the emergence of a three-particle level lies in the production of long-range interaction between three particles by a two-particle resonance with a large spatial size.

Thus, the number of resonant states in a three-particle system is determined only by specific properties of paired subsystems.

The masses of the particles have the strongest influence on the effect. The following three characteristic regimes can be singled out: the mode of identical particles, the mode of a heavy center, and the molecular mode [1, 13–15].

The heavy-center mode takes place when the masses of two particles are of the same order m_l, while the mass m_h of the third particle is much larger. The pair of light particle has no energy level and these particles do not interact with each other, but interact with the heavy particle through the attracting potential. In this case, if the mass of the third particle is infinitely large, we are dealing with the case of a pair of particles in a force center;

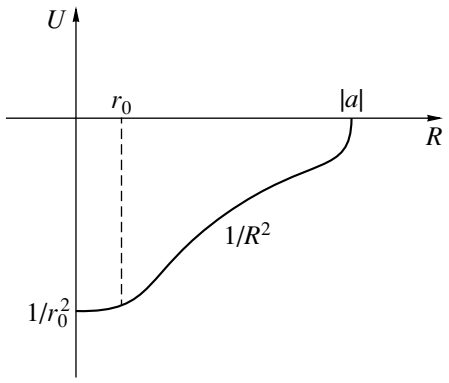


Fig. 1. Effective potential responsible for resonances in a three-particle system.

naturally, three-particle levels do not emerge in such a system. In this case, the heavy particle does not respond to the motion of the noninteracting particles moving independently from each other in the field of the stationary heavy particle. Consequently, in this limit, the binding energy of the three particles is the additive sum of the binding energies of two-particle systems. However, for a finite mass of the heavy particle, the motion of all the three particles is correlated, so that the center of mass of the system remains at rest. In this case, the heavy particle responds to a change in the position of other particles whose motion becomes correlated in spite of the absence of a direct interaction between them. Thus, dynamic correlation in the motion of coupled particles can be treated as a sort of attraction. It should be noted that such a dynamic attraction also appears in the case when repulsive force act between the particles coupled in this way. In this case, dynamic attraction compensates mutual repulsion and leads to stabilization of the system. This can be clearly seen, for example, for the ion of positronium $e^+e^-e^-$ [14–16]. In this case, for any finite mass of a heavy center, the number of levels is

$$N \propto \frac{m_l}{m_h}\ln\frac{1}{e_0 m_l r_0^2}.$$

A special feature of this mode is that extremely shallow levels in paired subsystem are required for the existence of three-particle levels in contrast to the molecular mode, where the requirements imposed on paired levels are much less stringent and more realistic.

In the molecular mode, when a light particle has shallow levels in the interaction with the heavy particles, the number of levels is

$$N \propto \sqrt{\frac{m_l}{m_h}}\ln\frac{1}{|e_0| m_l r_0^2},$$

and the potential of the interaction produced by the

light particle has the form

$$V \propto \frac{-0.32}{m_l r_{hh}^2},$$

which is precisely the energy of the molecular energy level. A simple example of this mode is a system consisting of an electron and two neutral atoms. A molecule formed in this way differs from a conventional molecule in that its nuclei vibrate in region R whose size is determined by the energy e_0 of the shallow paired level; in addition to vibrational levels, this system also has a rotational spectrum. Thus, two-particle levels in this mode lead to the formation of a series of not only vibrational, but also rotational levels [1, 13–15].

It should be noted that such peculiar resonance states are manifested in a wide range of conditions and form a stable phenomenon which can be reliably identified and confirmed experimentally.

2. BASIC EQUATIONS AND RESULTS OF CALCULATIONS

We will analyze these peculiar resonant states quantitatively in the case of the molecular mode using the Faddeev integral equations [13]. In the given approximation (three particles, viz., two atoms and an electron), these equations are formulated for three parts into which the total wavefunction of the three-body system splits,

$$\Psi = \sum_{i=1}^{3} \Psi_i.$$

Each part corresponds to possible divisions of the system of three particles into noninteracting subgroups. In the momentum space, in the case of scattering of particle 1 from the coupled pair (2, 3), these equations have the form [13, 14]

$$\begin{gathered}\Psi_i = \Phi_i \delta_{i1} - G_0(Z) T_i (\Psi_j + \Psi_k), \\ i, j, k = 1, 2, 3; 3, 1, 2; 2, 1, 3.\end{gathered} \tag{1}$$

Here, Φ_1 describes the initial state of the three-body system: free motion of particle 1 and the bound state of pair (2, 3); $G_0(Z) = (H_0 - Z)^{-1}$, $Z = E + i0$, where H_0 is the operator of free motion of the three particles; E is the total energy of the three-body system, which is equal to the sum of the kinetic energy of projectile 1 and the binding energy of pair (2, 3); and T_i is a paired T matrix that can be unambiguously defined in terms of the paired interaction potential V_i with the help of the Lippmann–Schwinger equations

$$\begin{gathered}T_i = V_i + V_i G_i T_i, \quad C_i = (h_i - Z_i)^{-1}, \\ h_i = \Delta_i + V_i.\end{gathered} \tag{2}$$

To describe the motion of three particles in the center-of-mass system, we use the generally accepted Jacobi coordinates. It should be borne in mind that we must use as integration variables in Eq. (1) a certain system of variables which is found to be most convenient. For example, in the integral corresponding to the expression $G_0 T_1 \Psi_2$, it is more convenient to take $\mathbf{k}_2$ and $\mathbf{p}_2$ as integration variables. In this case, variables $\mathbf{k}_1$ and $\mathbf{p}_1$ determining the kernel of operator T_1 should be expressed in terms of variables $\mathbf{k}_2$ and $\mathbf{p}_2$. Sometimes, it is more convenient to use variables $\mathbf{p}_1$ and $\mathbf{p}_2$ in the same situation.

Paired T matrices $t_i(\mathbf{k}_i, \mathbf{k}_i'; Z)$ appearing in the kernels of the equations have singularities in variable Z: the poles corresponding to the discrete spectrum of paired subsystems and a cut along the positive part of the real axis generated by the spectrum of the two-body problem. The explicit form of these singularities gives the spectral representation of matrix T. The poles of the T matrix corresponding to the discrete spectrum generate singularities in the wavefunction components Ψ_i; separating these components, we obtain the representation

$$\begin{gathered}\Psi_i(\mathbf{k}_i, \mathbf{p}_i; \mathbf{p}_i^0) = \varphi(\mathbf{k}_i)\delta(\mathbf{p}_i - \mathbf{p}_i^0) \\ - \frac{B_i(\mathbf{k}_i, \mathbf{p}_i; \mathbf{p}_i^0; Z)}{p_i^2/2n_i + k_i^2/2m_{jk} - Z},\end{gathered} \tag{3}$$

where

$$\begin{gathered}B_i(\mathbf{k}_i, \mathbf{p}_i; \mathbf{p}_i^0; Z) \\ = -\sum_{j=1}^{3}\left(Q_j(\mathbf{k}_i, \mathbf{p}_i; \mathbf{p}_i^0; Z) - \frac{\varphi_j(\mathbf{p}_j) R_{ji}(\mathbf{k}_j; \mathbf{p}_i^0; Z)}{p_j^2/2n_j - \kappa_j - Z}\right),\end{gathered}$$

and Q_j and R_{ji} are smooth functions of their variables. Such a division of singularities appears automatically in the numerical solution of integral equations. To define functions Q_j and R_{ji} unambiguously, we can proceed as follows. We substitute Ψ_i in form (3) into initial equations (1) and equate the coefficients of identical singularities. This gives the equations for these functions which can be used for expressing explicitly all main characteristics of the three-body problem: wave function, elements of the S matrix, as well as the amplitudes and cross sections of all processes occurring in the three-body system. Thus, the cross section of the elastic scattering process has the form

$$\frac{d\sigma_{11}}{d\Theta} = (2\pi)^4 n_1 |R_{11}|^2,$$

the cross section of rearrangement processes is given by

$$\frac{d\sigma_{1i}}{d\Theta} = \frac{(2\pi)^4 n_1 p_f |R_{1i}|^2}{p_1^0},$$

and the cross section of the process of decay into three

free particles has the form

$$\frac{d^2\sigma_{1\to 3}}{d\Theta dp} = \frac{(2\pi)^4 n_i p_f |B_{0i}|^2}{p_1^0},$$

where

$$p_f^2 = 2n_i\left(\frac{p_i^{02}}{2n_i} - \kappa_1^2 - \kappa_i^2\right).$$

The main advantage of the Faddeev equations (1) is that

(i) the solution of this equation gives simultaneously the amplitudes and cross sections of all processes occurring in the three-particle system;

(ii) the accuracy in determining the bound state from the solution of the Faddeev equations is much higher than the accuracy obtained by solving the Schrödinger equations (this peculiarity is associated with the fact that Eqs. (1) were formulated for the wavefunction components and, hence, take into account possible asymptotic forms of the three-particle system);

(iii) these equations make it possible to carry out a correct (from the standpoint of mathematics) analysis of scattering processes, in which all three free particles are in the initial state [12, 13]; this is impossible in all approaches proposed earlier [5–11]:

$$1 + 2 + 3$$

$$\longrightarrow \begin{cases} 1 + (2, 3) \text{ (elastic scattering processes)}, \\ 1 + (2, 3)^* \text{ (excitation processes)}, \\ \left.\begin{array}{l} 3 + (1, 2)^* \\ 2 + (1, 3)^* \end{array}\right\} \text{(rearrangment processes with excitation)}, \\ 1 + 2 + 3 \text{ (ionization processes)}. \end{cases}$$

In this case, we have the following representation for the wavefunction [13–15]:

$$\Psi_0(\mathbf{k}, \mathbf{p}; \mathbf{k}^0, \mathbf{p}^0) = \delta(\mathbf{k} - \mathbf{k}^0)\delta(\mathbf{p} - \mathbf{p}^0)$$

$$- \frac{\sum_{i,j} M_{i,j}\left(\mathbf{k}, \mathbf{p}; \mathbf{k}^0, \mathbf{p}^0; \frac{k^{02}}{2m} + \frac{p^{02}}{2n} + i0\right)}{\frac{p^2}{2n} + \frac{k^2}{2m} - \frac{k^{02}}{2m} + \frac{p^{02}}{2n} + i0},$$

where functions $M_{i,j}$ satisfy the following system of equations:

$$M_{i,j}(Z) = \delta_{i,j} T_i(Z) + T_i(Z) G_0(Z) \sum_{k \neq i} M_{k,j}(Z).$$

For cross sections of these processes, we obtain the following expressions [13–15]:

$$S_{00}(\mathbf{k}, \mathbf{p}; \mathbf{k}', \mathbf{p}') = \delta(\mathbf{k} - \mathbf{k}')\delta(\mathbf{p} - \mathbf{p}')$$

$$- 2\pi i \delta\left(\frac{p^2}{2n} + \frac{k^2}{2m} - \frac{p'^2}{2n} - \frac{k'^2}{2m}\right)$$

$$\times \sum_{i,j} M_{i,j}\left(\mathbf{k}, \mathbf{p}; \mathbf{k}', \mathbf{p}'; \frac{k'^2}{2m} + \frac{p'^2}{2n} + i0\right)$$

corresponds to processes in which three free particles are in the initial and final states;

$$S_{0s_i}(\mathbf{k}, \mathbf{p}; \mathbf{p}_i') = 2\pi i \delta\left(\frac{p^2}{2n} + \frac{k^2}{2m} + \kappa_{s_i}^2 - \frac{p_i'^2}{2n_i}\right)$$

$$\times \sum_k Q_{k,i}^{s_i}\left(\mathbf{k}, \mathbf{p}; \mathbf{p}_i'; -\kappa_{s_i}^2 + \frac{p_i'^2}{2n_i} - i0\right)$$

$$+ \sum_{s_k} \psi_{s_k}(\mathbf{k}_k) R_{k,i}^{s_k s_i}\left(\mathbf{p}; \mathbf{p}_i'; -\kappa_{s_i}^2 + \frac{p_i'^2}{2n_i} + i0\right),$$

$$S_{s_i 0}(\mathbf{p}_i, \mathbf{k}', \mathbf{p}') = 2\pi i \delta\left(-\kappa_{s_i}^2 + \frac{p_i'^2}{2n_i} - \frac{p'^2}{2n} - \frac{k'^2}{2m}\right)$$

$$\times \sum_j \tilde{Q}_{j,i}^{s_i}\left(\mathbf{p}_i; \mathbf{k}', \mathbf{p}'; \frac{k'^2}{2m} + \frac{p'^2}{2n} + i0\right)$$

$$+ \sum_{s_j} \psi_{s_j}(\mathbf{k}_j') R_{ij}^{s_i s_j}\left(\mathbf{p}_i; \mathbf{p}_i'; \frac{k'^2}{2m} + \frac{p'^2}{2n} + i0\right)$$

correspond to processes in which a coupled pair of particles s_j is present in the initial or the final state. The equations for functions Q_{ij}, $\tilde{Q}_{ij}$, and R_{ij} are analogous to the equations for $M_{i,j}$ and are given in [13–15].

It should be noted that potentials do not appear explicitly in integral equations (1); these equations contain a more general characteristic, viz., T matrices, which are connected with the potentials of the Lippmann-Schwinger equations (2). Consequently, although potentials are formally used in the given method, we essentially model T matrices, which are constructed on the basis of the Bateman method [13, 14] suitable for any local potential. This method considerably simplifies numerical solution of the system of integral equations (1) and sometimes even leads an analytic solution [13–17].

Integral equations (1) possess good properties (from the mathematical standpoint)) such as the Fredholm property and unambiguous solvability only under certain conditions imposed on two-particle data [13]:

(i) paired potentials $V_i(\mathbf{k}, \mathbf{k}')$, which are nonlocal in the general case, are smooth functions of $\mathbf{k}$, $\mathbf{k}'$ and satisfy the condition

$$|V_i(\mathbf{k}, \mathbf{k}')| \le (1 - |\mathbf{k} - \mathbf{k}'|)^{1-\epsilon}, \quad \epsilon > 0;$$

(ii) point $Z = 0$ is not a singular point for Eqs. (2); i.e., all three scattering lengths in pair channels are finite;

(iii) the positive two-particle spectrum is continuous. This condition is essential for nonlocal potentials since positive eigenvalues may appear only in this case, and this condition is satisfied for virtually all physical processes.

Coulomb potentials and hard-core potentials do not satisfy the first condition: Coulomb potentials lead to a singularity of the type $|\mathbf{k} - \mathbf{k}'|^{-2}$ in T matrices, while hard-core potentials result in a slow decrease in the T matrix for large momenta. When the second condition is violated, the Fredholm property of Eqs. (1) is lost for $Z = 0$, which leads to the above-mentioned Efimov's effect (emergence of an infinitely large discrete spectrum in a three-body system under certain conditions). A similar situation emerges in the case of scattering of electrons from diatomic molecules, for which this effect was experimentally observed for the first time. The three-body approximation considered here reproduces these experimental results in a quite natural way.

It should be emphasized once again that the given approximation appears quite reasonable for values of the incident electron energy lower than the electron excitation energy of the molecule.

As the initial data in such a formulation of the problem, we use pair interaction potentials, masses, and energies of colliding particles. For potentials of pair interaction of electrons with atoms of the molecule, we used potentials of the form

$$V(r) = \frac{\lambda \exp(-\beta r)}{r}, \tag{4}$$

whose parameters were determined on the basis of the electron binding energy at a negative ion, scattering lengths, and effective radius. Allowance for spin (in the case of homonuclear molecules) was made as follows. For the scattering length, we used the quantity [5, 6, 14–17]

$$\frac{1}{a} = \frac{1}{a_1} = \frac{1}{a_2} = \frac{1}{4}\left(\frac{3}{a_t} + \frac{1}{a_s}\right),$$

where a_t and a_s are the triplet and singlet scattering lengths, respectively.

Pair potentials of interaction between atoms in molecules was simulated by the Morse potentials

$$V(r) = D(1 - \exp(-\alpha(r - r_0))), \tag{5}$$

whose parameters were determined on the basis of spectroscopic data [18].

Numerical solution of integral equations (1) involves considerable difficulties because the kernels of integral equations (1) contain the following singularities:

(i) Branching points of the square-root type, which correspond to the thresholds of two-particle processes.

(ii) Poles determining the bound states of two-particle subsystems in pair T matrices.

(iii) Logarithmic singularities. The position of these singularities and their form depend on the form of pair potentials of interaction, the masses of colliding particles, and the total energy of the three-body system.

Consequently, various methods of numerical solution of Eqs. (1) are based on:

(i) Bypassing the singularities by deforming the integration contour to the complex plane; this is possible only when pair potentials are defined analytically and do not introduce additional singularities in the kernels of Eqs. (1), which is the case with the potential of the rectangular well or hard core type.

(ii) Using the method of integration along the real axis; in this case, the solution should be approximated in the vicinity of these singularities either by summing the Born series with the help of Padé approximants, or using interpolation polynomials for approximating the solution, or using the moments methods for this purpose.

(iii) Application of various variational principles, collocation methods, methods of artificial separation of singularities, and so on.

All methods mentioned above have the same disadvantage: they can be used only for a specific three-particle system. These methods mainly involve calculation of bound states or states of scattering of three identical particles.

Here, we propose a quite universal method for solving system of equations (1) for calculating bound states as well as scattering states in systems with arbitrary masses, which interact via arbitrary pair short-lived potentials that can also be defined numerically. In the method proposed here, the domain of an unknown function is divided into a number of intervals on each of which the function is approximated with the help of corresponding interpolation polynomials. The method for solving system of equations (1) is a modification of the standard method for solving integral equations, in which the integral on the right-hand side is replaced with the help of a quadratures formula for solving Eq. (1). As a result, we arrive at a system of algebraic equations for values of the sought function at the nodes of the quadratures formula. In the proposed method, the

domain of the sought function is divided into a number of segments, on each of which the function is determined with the help of interpolation polynomials reproducing the correct behavior of the function in the vicinity of the above singularities, after which integration is carried out using quadratures formulas. A package of application programs was used for realization of the proposed numerical method for solving the system of integral equations (1) [13–15].

We will consider the results of calculation of these resonant processes using as an example the calculation of cross sections for the simplest reaction of dissociative attachment of electrons to hydrogen molecules. These results are shown in Fig. 2 together with the latest experimental data [4, 19–21] and the results of calculations based on other approximations [5–11]. These results confirm the existence in this system of the resonant states considered above. It should be noted here that experimental results of observation of three-particle resonant states were presented for the first time by Poincaré [3] for the dissociative attachment of an electron to hydrogen molecules. However, in view of the electron energy distribution within about 0.1 eV in the beam, only nonmonotonic character of the energy dependence of the dissociative attachment cross section was revealed in these experiments; this nonmonotonicity was confirmed in theoretical calculations published more than 20 years ago [14, 17]. The oscillatory structure of the dissociative attachment cross sections was confirmed only recently in experiments [19], in which a special technique was used for energy stabilization of the electron beam (~meV).

To estimate the influence of particle charges on the effect, we consider the scattering of electrons from hydrogen halide molecules. Since the electron affinity to the hydrogen atom is much smaller than to a halogen atom [18], a hydrogen halide molecule can be proposed as a system consisting of a proton and a negative halogen ion. Thus, in the approach proposed here, the main approximation is that the interaction of the projectile electron with the nuclei of the target molecule is replaced by the interaction of the incident electron with the proton and the negative halogen ion. The complex many-particle problem of calculating the electron scattering cross section at diatomic molecules is reduced to the problem of collision in a three-body system, which can be solved using the method of quantum scattering problem in a few-particle system. Naturally, this approximation is valid for energies of the incident electron lower than the electron excitation energy of the molecule.

Computational difficulties encountered in calculation of cross sections in the given approximation are mainly associated with the long-range Coulomb interaction potentials between a projectile electron, a proton, and a negative halogen ion. It was mentioned above that in this case the integral Faddeev equations cannot be

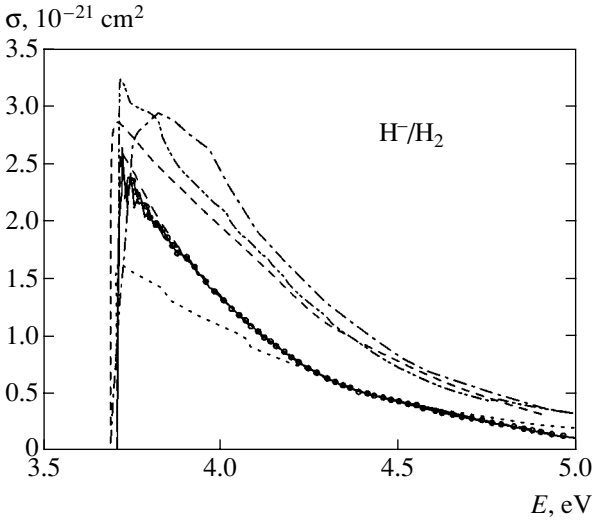


Fig. 2. Dependences of the cross section of electron dissociative attachment to hydrogen molecules on the energy of projectile electrons: the solid curve corresponds to experimental data [19]; the results of calculations performed in [11], [7, 10], [10, 26], [21], and [20] are represented by the fine-dash, large-dash, dot-and-dash, dotted, and double dot-and-dash curves, respectively; our results of calculation are presented by circles.

applied directly; either these equations should be modified or the differential formulation of the Faddeev equations in the coordinate state should be used [13–15]. It should be noted that in the general case of scattering of an electron from halogen molecules in the given approximation,

$$e + (H^+G^-)$$

$$\longrightarrow \begin{cases} e + (H^+G^-) \ \text{(elastic scattering processes)}, \\ e + (H^+G^-)^* \ \text{(excitation processes)}, \\ \left.\begin{matrix} H^+ + (e, G^-)^* \\ G^- + (e, H^+)^* \end{matrix}\right\} \ \text{(rearrangment processes with excitation)}, \\ \left.\begin{matrix} e + H^+ + G^- \\ e + e + H^+ + G \\ e + H + G \end{matrix}\right\} \ \text{(ionization processes)}, \end{cases}$$

the Faddeev equations for four mutually interacting bodies (two electrons, a halogen atom, and a proton) should be used. However, for some processes such as dissociative attachment reactions, we can confine our analysis to equations for three pairwise interacting bodies.

In our case, to calculate dissociative attachment of an electron to hydrogen halide molecules, we apply the

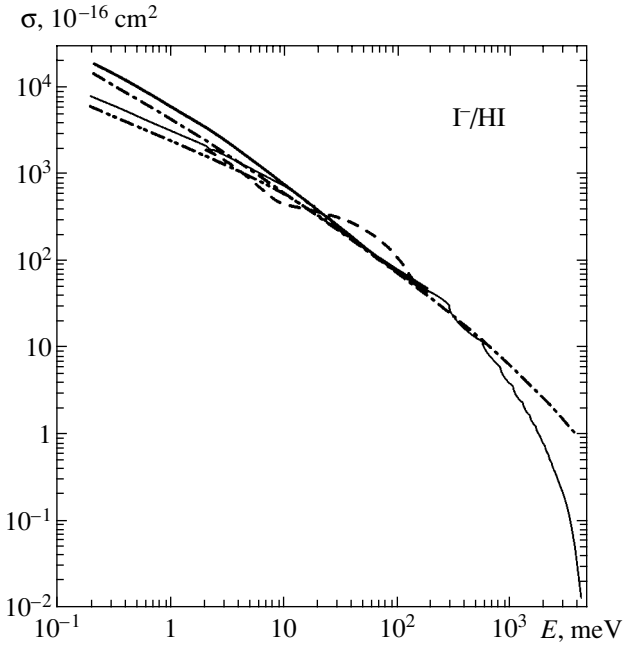


Fig. 3. Dissociative attachment of electrons to HI molecules: the results of calculations performed in [19], [9], and [21] are represented by the dashed, fine solid, and bold solid curves, respectively; the results of calculations in the semiclassical approximation [21] are depicted by the dash-and-double dot curve; our results are shown by the dot-and-dash curve.

Faddeev equation for three charged particles in the coordinate space, which have the form [13–15]

$$(-\Delta_{x_i} - \Delta_{y_i} + V_i(x_i) - E)\Psi_i = -V_i \sum_{j \neq i} \Psi_j, \qquad (6)$$

where

$$V_i = \frac{n_i}{x_i} + V_{st}(x_i), \quad n_i = \frac{q_k q_j}{\sqrt{2m_{kj}}},$$

$$\mathbf{x}_i = \sqrt{\frac{2m_j m_k}{m_j + m_k}}(\mathbf{r}_j - \mathbf{r}_k),$$

$$\mathbf{y}_i = \sqrt{\frac{2m_i(m_j + m_k)}{m_i + m_j + m_k}}\mathbf{r}_i - \frac{m_j \mathbf{r}_j + m_k \mathbf{r}_k}{m_j + m_k},$$

and the coordinates are connected via the relations

$$\mathbf{x}_i = c_{ij}\mathbf{x}_j + s_{ij}\mathbf{y}_j, \quad \mathbf{y}_i = -s_{ij}\mathbf{x}_j + c_{ij}\mathbf{y}_j,$$

$$s_{ij}^2 = \frac{m_k \sum_k m_k}{(m_i + m_j)(m_j + m_k)}, \quad s_{ij}^2 + c_{ij}^2 = 1,$$

$V_{st}(x_i)$ being pair short-range interaction potentials defined by (4) and (5). The relation between the momentum and coordinate representations is defined by the Fourier transformation,

$$\Psi(\mathbf{k}_i, \mathbf{p}_i) = (2\pi)^{-3} \int \exp[-i(\mathbf{k}_i \cdot \mathbf{x}_i + \mathbf{p}_i \cdot \mathbf{y}_i)]\Psi(\mathbf{x}_i, \mathbf{y}_i) d\mathbf{x}_i d\mathbf{y}_i.$$

To obtain a unique solution of integrodifferential equations in the coordinate space, we must add the boundary conditions, which have the form [13–15]

$$\Psi_i(\mathbf{x}_i, \mathbf{y}_i)_{x_i, y_i \to 0} \longrightarrow 0, \qquad (7)$$

$$\Psi_i(\mathbf{x}_i, \mathbf{y}_i)_{\rho = \sqrt{x^2 + y^2} \to \infty} \longrightarrow \phi_i(x_i)\exp(i\mathbf{k}_i \cdot \mathbf{y}_i - iw_i^0) + \sum_j A_{ij}(\hat{\mathbf{y}}_j \cdot \hat{\mathbf{k}}_i)\phi_i(x_j)\frac{\exp(i\sqrt{E_j}|\mathbf{y}_j| + iw_{ij})}{|y_j|} + A_{0i}(\hat{X}, \hat{k}_i)\frac{\exp(i\sqrt{E}|X| + iw_0)}{|X|^{5/2}}, \qquad (8)$$

where

$$w_i^0 = \frac{n_i}{2|\mathbf{k}_i|}\ln(|\hat{\mathbf{k}}_i||\hat{\mathbf{x}}_i| - (\mathbf{k}_i \cdot \mathbf{x}_i)),$$

$$w_{ij} = \sum_{k \neq j} \frac{n_k}{2|s_{jk}|\sqrt{E_k}}\ln(2\sqrt{E_k})|\mathbf{y}_k|,$$

$$w_0 = -\frac{|\mathbf{X}|}{2\sqrt{E}}\sum_i \frac{n_i}{|\mathbf{x}_i|}\ln(2\sqrt{E})|\mathbf{X}|,$$

$$n_i = \frac{kq_i q_j}{\sqrt{2m_{ij}}}, \quad E_k = E - \kappa_j.$$

A large number of various numerical methods have been developed on the basis of approximation of components Ψ_i by bicubic Hermite splines, quintet basis splines, etc. However, an effective, reliable, and universal algorithm of the numerical solution of Eqs. (6) with boundary conditions (7) and (8) in the coordinate space has not been developed for the following reasons. First, an algorithm of numerical solution for processes with three free particles in the initial and final states does not exist in view of rather complex boundary conditions. Second, point-by-point convergence of the result to the exact solution upon a decrease in the mesh size cannot be proved analytically in any of the known numerical methods based on finite different approximation.

Consequently, the application of the mesh method in the polar coordinate system [14] for numerically solving the system of coupled integrodifferential equations (6) in partial derivatives with boundary conditions (7) and (8) appears as most justified since analytic solutions also exist in this case for some potentials determining the resonant states under investigation [13, 14, 22]. This makes it possible to monitor the accuracy of the solutions obtained by the numerical method.

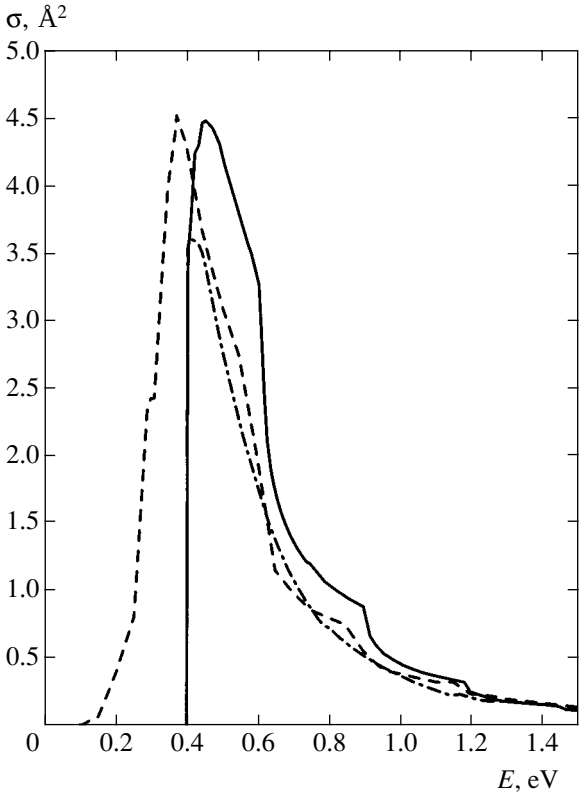


Fig. 4. Dissociative attachment of electrons to HBr molecules: the experimental results [9, 10] are shown by the dashed curve; the results of calculations performed in [26] are represented by the solid curve; our results are shown by the dot-and-dash curve.

Figures 3–6 show the results of calculating the cross sections of electron dissociative attachment to hydrogen halide molecules and their isotope-substituted modifications in the ground state and in excited vibration-rotation states. Also shown are experimental results [4, 7, 9, 11, 19, 21] and the results of calculations based on other approximations [4–11], which demonstrate suppression of oscillations in the scattering cross sections.

Let us consider the effect of these peculiar resonances on the rates of chemical reactions, which appears interesting for explaining electron transport in proteins (physically, this transport is one of the main functionally important processes in a cell [15, 23]). Knowing this transport mechanism, it would be possible to explain how a transition is made from structurally disorganized chemical transformations (e.g., in solutions) to coordinated subsequent stages typical of biological systems [15]. It should be noted that no new interactions are used in nature and the process is organized due to an appropriate choice of molecular structures and the corresponding well-known interactions. Thus, identification and analysis of these interactions will make it possible to understand physical mechanisms of processes occurring in proteins and other molecular structures such as DNA and RNA.

To study the processes described above, we first consider a simple system of two identical coupled particles tunnelling through a potential barrier by using a

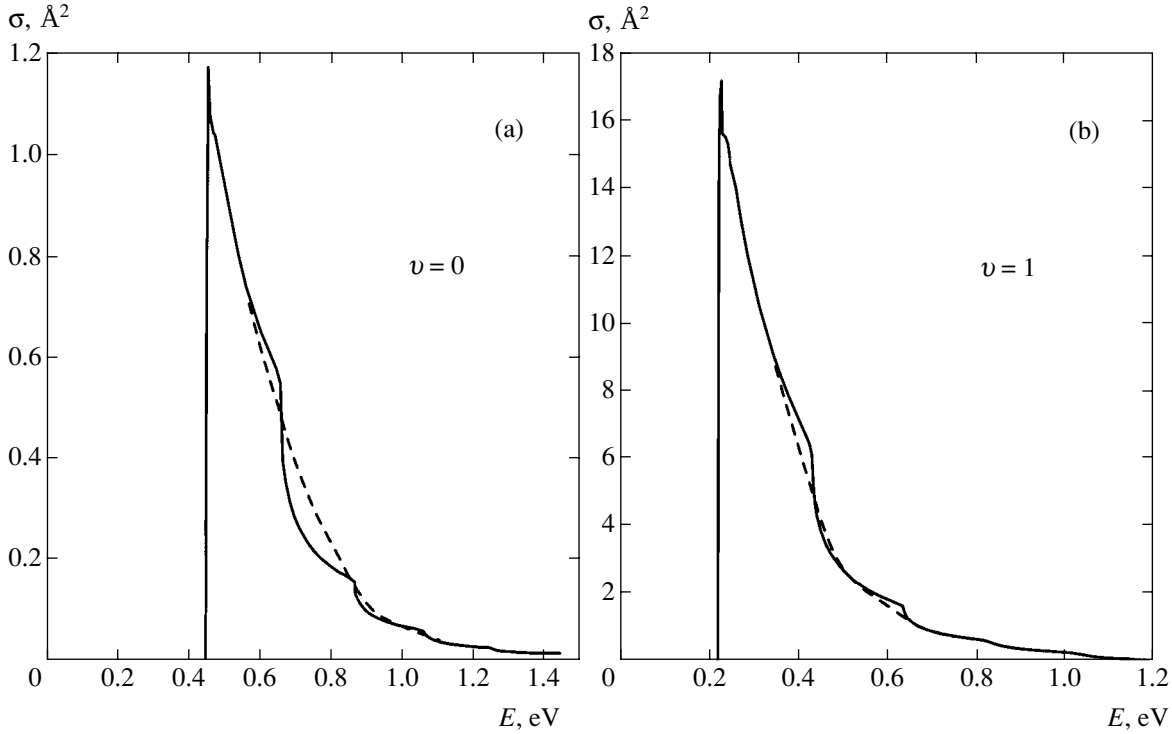


Fig. 5. Dissociative attachment of electrons to DBr molecules that were initially (a) in the ground state ($v = 0$) and (b) in the excited vibrational state ($v = 1$). Our results and the results of calculations performed in [26] are shown by the dashed and solid curves, respectively.

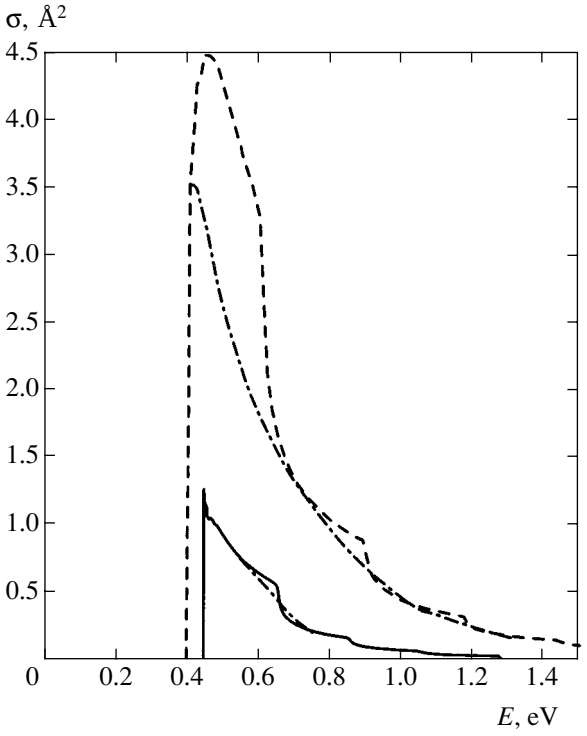


Fig. 6. Isotopic effect in the reactions of dissociative attachment of electrons to HBr and DBr molecules. The solid and dashed curves correspond to the results of calculations performed using the nonlocal theory [26]; our results are shown by the dot-and-dash curve.

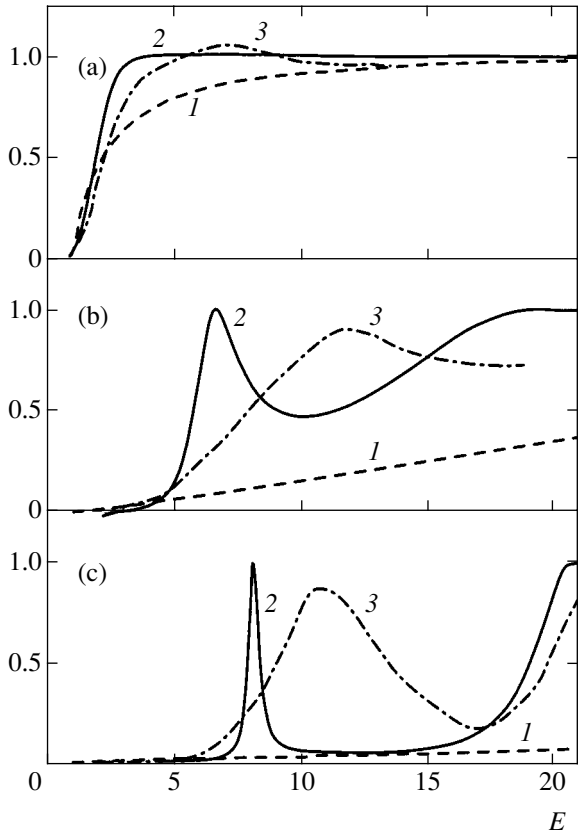


Fig. 7. Tunneling probabilities for a structureless (*1*) and structured (*2*, *3*) particles as functions of the energy of the projectile particle for the harmonic potential (*2*) and the Morse potential (*3*) for various characteristics of the barrier a = 1 (a), 6 (b), and 11 (c). Energy E is given in units of the corresponding potential.

mathematically correct quantum theory of scattering in a few-body system [13–15].

It should be noted that tunnelling of particles (including structured ones) is usually considered on the basis of well-known theories [24] and results are automatically extended to many-particle systems (especially in applied studied, e.g., in biology [23]). Most results in these applications are associated precisely with analysis of tunnelling through various potential barriers of multicomponent structured complexes. Quite often, such results do not correspond to the initial problem and do not reproduce experimental data. It should be noted that if the barrier size is much larger than the characteristic size of a complex, the difference from the structureless case is insignificant. If the size of the complex is commensurate with the barrier width, mechanisms appear [15, 25] leading to anomalous transparency of the barrier (analogously to the Ramsauer effect [4, 5, 7]).

The physical reason for the barrier transparency is associated with the possibility of formation of a barrier resonance since the potential energy of the system may have a local minimum ensuring the metastable state of the complex; to this end, the interaction of all particles of the complex with the barrier is required.

To demonstrate this effect, we use the quantum scattering theory to consider the tunnelling of a pair of identical particles $m_1 = m_2 = m$ coupled through various types of interaction (harmonic oscillator, Morse and Gauss potentials) through a potential barrier which was simulated by a potential of the form [25]

$$V(x^{\pm}y) = \frac{a\exp((x^{\pm}y)^2/2b)}{\sqrt{2\pi b}},$$

where a and b are the barrier height and width. The results of our calculations are shown in Fig. 7 together with the results of calculations [25] based on Eqs. (1). It can be seen from the figure that, for a barrier height much larger than the characteristic size of the incident complex, its transmission probability differs insignificantly from the transmission probability in the case of structureless particles. If the size of the complex is commensurate with the spatial size of the barrier, mechanisms appear which lead to a substantial transparency of the barrier and even to its total transparency in some cases.

In the simplest case of a complex consisting of two particles, the physical pattern of the barrier transparency is determined by the formation of a resonant state

upon the passage of only one of the particles through the barrier. Thus, two particles are on different sides of the barrier (i.e., this resonant state is preserved until the other particle passes through the barrier). The barrier width is determined by the lifetime of this resonance.

It should be noted here the penetrability symmetry breaking for structured particles may serve as a possible mechanism explaining different penetrabilities of biomembranes in opposite directions (osmosis). Another explanation of penetrability of a biomembrane is given in [23] on the basis of the assumption that an isothermal phase transition of the melting/crystallization type occurs in a monolayer of the membrane, but in 2D and not in 3D system. This makes it possible to interpret some peculiarities of the blood circulatory system, which could not been explained earlier.

Analogous results are considered in [24]; however, factors determining the penetrability of potential barriers for structured particles are completely different.

Figure 7 shows that the tunnelling probability in some cases may attain unity, which can be explained by interference suppression of the reflected wave (this phenomenon is widely used for blooming of optical systems). As the number of interacting particles increases, the effect of enhancement of the barrier penetrability may substantially increase. Thus, under certain conditions, coupled clusters not only surmount obstacles more easily, but can also be themselves transparent to other particles (this is often encountered in biological systems).

The most astonishing fact is that this mechanism of potential barrier transparency for structural particles was confirmed in experiments [19–25].

These features are commonly observed in real systems, e.g., in the simplest chemical reactions induced by electrons:

$$e + AB(v_1, J_1)$$
$$\longrightarrow \begin{cases} e + \mathrm{AB}(v_1, J_1) \ \text{(elastic scattering processes)}, \\ e + \mathrm{AB}(v_2, J_2) \ \text{(vibration-rotation excitation)}, \\ \left.\begin{matrix} e\mathrm{A} + \mathrm{B} \\ \mathrm{A} + e\mathrm{B} \end{matrix}\right\} \ \text{(dissociative attachment of electron to molecule)}, \\ e + \mathrm{A} + \mathrm{B} \ \text{(dissociation of molecule)}. \end{cases}$$

In the approximation of the quantum theory of scattering in a few-body system, it is possible to reproduce the experimental data [9–11, 19–26] on the simplest chemical reactions occurring during the interaction of electrons with diatomic molecules in the ground state as well as in excited vibration-rotation states. These results are presented in Figs. 8–11; all calculations in this case are performed in the above-mentioned approximation, in which the interaction of an electron with nuclei and electrons of the target molecule was replaced by the interaction of the electron with each atom as a whole (the atom was treated as a force center). The same figures also show the results of calculations performed in the resonance model approximation with nonlocal potentials based on the quantum theory of scattering in a two-body system [26].

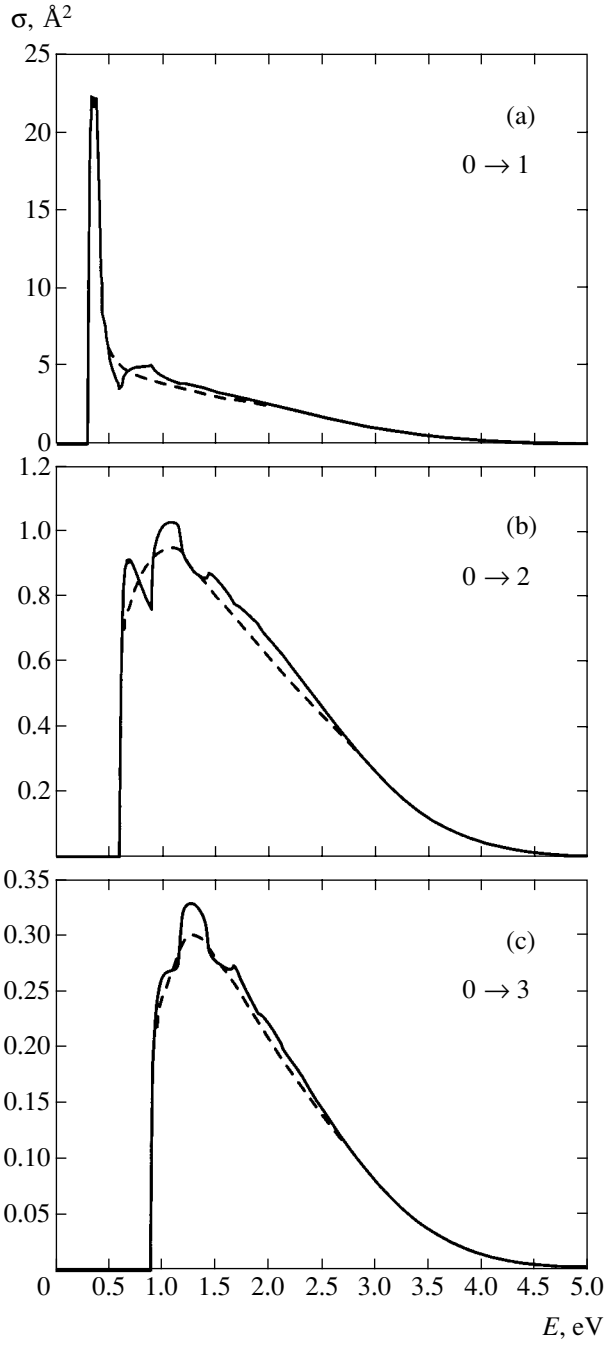


Fig. 8. Vibrational excitation of hydrogen halide (HBr) molecules by electrons, $e + \mathrm{HBr}(v = 0) \longrightarrow e + \mathrm{HBr}(v = 1$ (a), 2 (b), and 3 (c)). The dashed curves describe our results and solid curves are the results of calculations performed in [26].

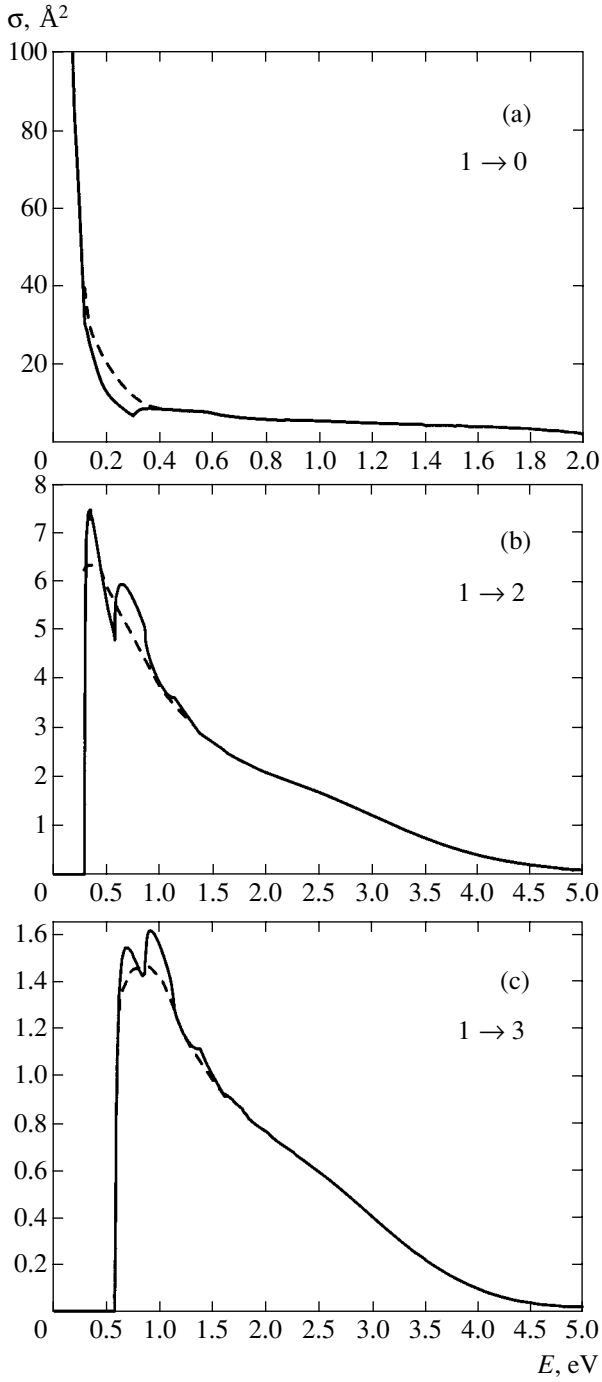


Fig. 9. Vibrational excitation of hydrogen halide (HBr) molecules by electrons, e + HBr(v = 1) ⟶ e + HBr(v = 0 (a), 2 (b), and 3 (c)). The dashed curves describe our results and solid curves are the results of calculations performed in [26].

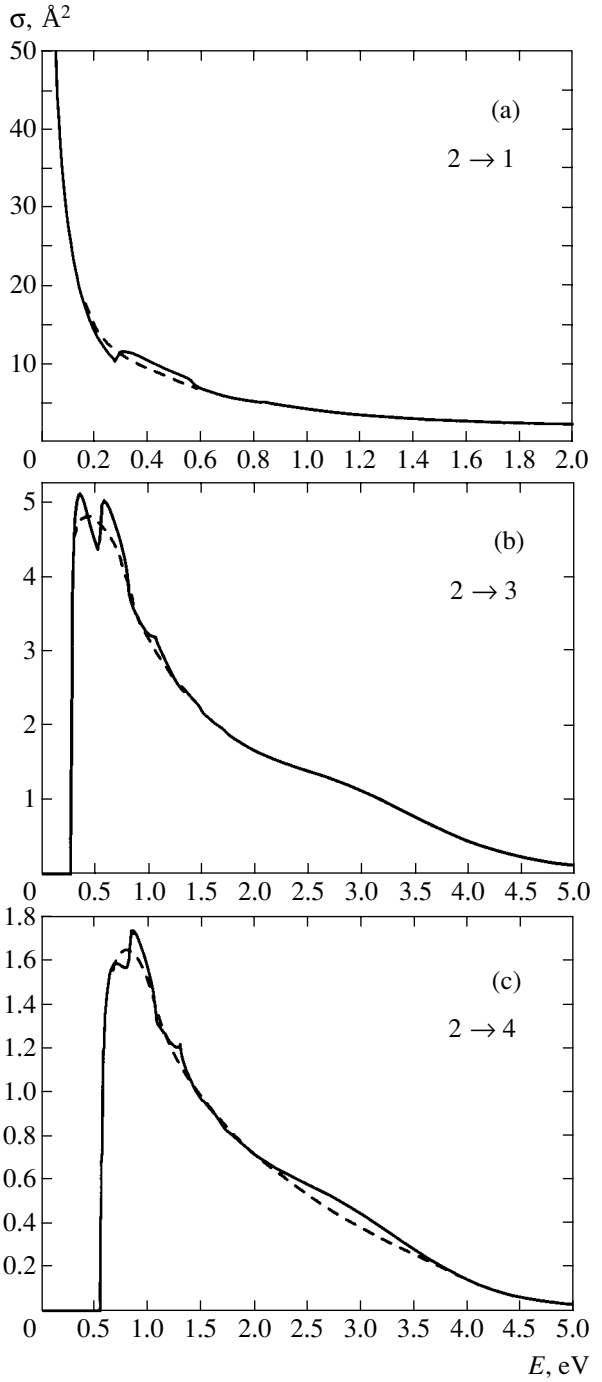


Fig. 10. Vibrational excitation of hydrogen halide (HBr) molecules by electrons, e + HBr(v = 2) ⟶ e + HBr(v = 1 (a), 3 (b), and 4 (c)). The dashed curves describe our results and solid curves are the results of calculations performed in [26].

It should be noted in this connection that the application of nonlocal separable potentials opens wide prospects for detailed reproduction of experimental results [14, 15, 17]. For example, using the separable potential

$$V(k, k') = \lambda g(k) g(k')$$

with form factors of the type

$$g_{NL}(k) = k^L \sum_{i=0}^{N} \gamma_i k^{2i} \left(\prod_{i=1}^{L+N} (1 + \beta_i k^2) \right)^{-1},$$

we can reproduce the experimental data on scattering cross sections with a preset accuracy both in the approx-

imation of the two-body problem [10, 16] and in the approximation of the few-body problem [14, 15, 17, 27].

However, the main difficulty encountered in the application of these potentials is associated with the choice of parameters λ, γ, and β. For this reason, we use here the simplest pair potentials (4), (5) whose parameters can be chosen on the basis of real spectroscopic data [18] rather than empirical data, as was done in [26].

It should be noted once again that this approximation appears reasonable for energies of the projectile electron lower than the electron excitation energy of the molecule. Otherwise, it is necessary to use the many-particle approximation instead of the three-particle approximation since the contributions from the dissociation channels become significant (this is demonstrated in Figs. 8, 10, and 11). Consequently, we can speak of the agreement with the experimental results only on the average in view of the initial model of the process as well as the simplest pair potentials simulating the interaction of an electron with atoms [14, 15, 17].

A comparison of the results of our calculations with the available experimental data [4–11, 19–23] shows that simulation of the electron interaction with each atom of the molecule based on Eqs. (1), which correspond to the multiple scattering pattern [6, 14, 17], makes it possible to attain satisfactory agreement with experiment (coincidence of the orders of magnitude of cross sections, including isotopic effects and threshold singularities [14–17]).

The well-known theoretical methods for studying resonant processes occurring during electron collisions with molecules [4–11] (the boomerang method, the R matrix method, the method of time evolution of the wavefunction, the Feschbach operator method, etc.) are based on interpreting this process as a multistage process. The first stage involves the electron capture by a molecule and the formation of a negative molecular ion. The second stage is the decay (evolution) of this state to various states of the decay products: a negative ion and a neutral or excited atom, two neutral or excited atoms and an electron (dissociation of the molecule), and an excited molecule and an electron (excitation of the molecule by electron impact). The basis of this formalism i.e., the formation of an intermediate state of a negative molecular ion) is not always substantiated from the physical point of view. For example, in the case of dissociative attachment of an electron to a hydrogen molecule, the lifetime of this complex is comparable to the electron mean free time, during which it covers a distance equal to the diameter of the hydrogen molecule. An analogous situation also emerges for the reaction [14, 27]

$$O(^3P) + CS(X^1\Sigma^+) \longrightarrow CO(X^1\Sigma^+) + S(^3P),$$

in which a considerable fraction of the translational energy (in accordance with the momentum limit $E_v/E_t \approx 0.88$ [27]) is transformed into the vibrational energy of the CO molecule. Such a reaction also occurs without the formation of an intermediate complex.

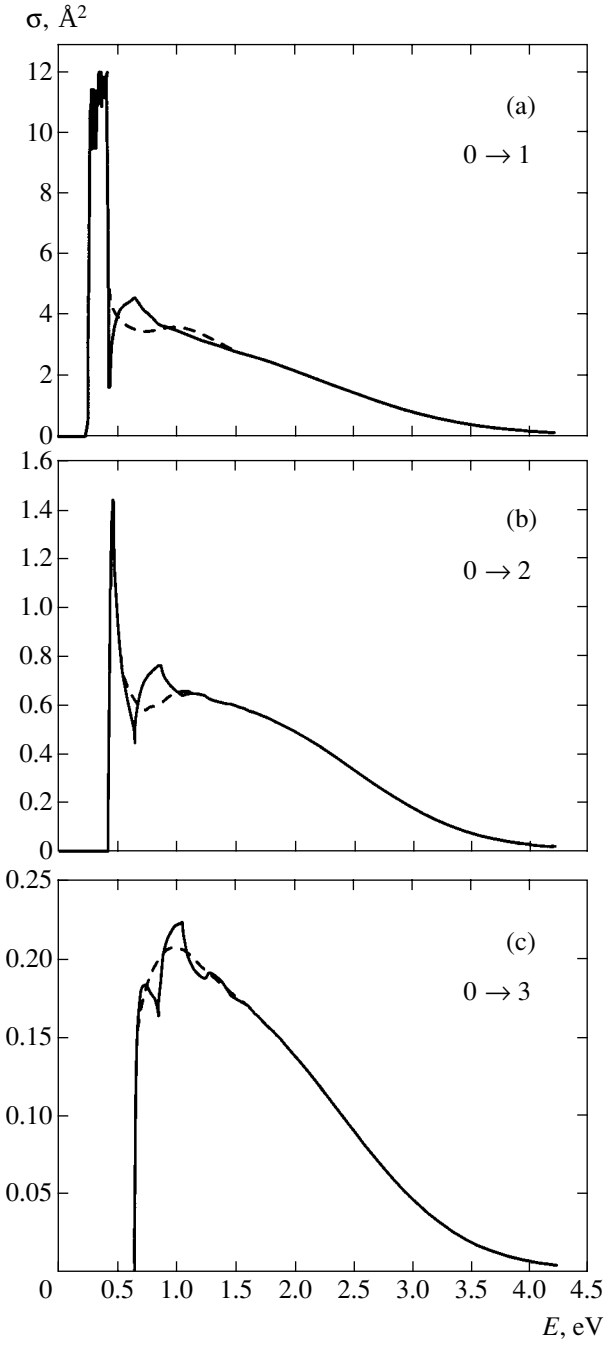


Fig. 11. Vibrational excitation of hydrogen halide (DBr) molecules by electrons, e + DBr(v = 0) ⟶ e + DBr(v = 1 (a), 2 (b), and 3 (c)). The dashed curves describe our results and solid curves are the results of calculations performed in [26].

Naturally, many reactions exist in which a long-lived intermediate complex is formed in the course of the reaction (see [5–11, 14, 15, 17] for details). However, for some processes like those described above, a preliminary analysis of experimental data for a given

processes required for analyzing various collisions; the absence of such an analysis often leads to erroneous interpretation of experimental data.

Thus, a class of processes existing in atomic, chemical, and biological physics can be referred to as direct processes in analogy with nuclear physics. The main feature of these processes is that no intermediate long-lived complex is formed in the course of scattering.

Consequently, the most adequate methods for interpreting such direct processes and reaction occurring with the formation of an intermediate complex have been those proposed by Faddeev, Yakubovskiĭ, and Merkur'ev [13], who developed a quantum theory of scattering in few-body systems without model assumptions concerning the formation of an intermediate complex during a collision. This method can be applied for describing direct processes as well as processes occurring with the formation of intermediate long-lived states. Thus, we can state that quantum transparency effects for various barriers and peculiarities of chemical reaction mechanisms described above can take place in various branches of physics, chemistry, and biology and can be interpreted in the framework of nonrelativistic quantum mechanics with the help of the formalism proposed in [13–15].

This is especially important for molecular biology, in which a consistent and mathematically correct explanation of fermentation reactions has not been obtained as yet [23]. The contemporary description of these reactions based on the assumption that a part of the free energy liberated as a result of a reaction is used for accelerating catalysis, i.e., penetration through a barrier (recuperation of energy), does not permit one to quantitatively analyze the reaction energy. For this reason, it is extremely difficult to experimentally confirm or reject the proposed model.

On the contrary, the above substantiation of the transparency of potential barriers for structural complexes with a size commensurate with the barrier width provides an explanation for such reactions based on the well-known physical principles in the framework of ordinary quantum theory for a few-body system.

Let us consider the geometrical (spatial) characteristics of the above-mentioned peculiar resonant states. Since it is quite difficult to study these characteristics experimentally in the case of electron collisions with molecules, we will consider systems that are accessible for experimental studies, viz., clusters of molecules of inert gases [28].

It should be noted that these molecular clusters consisting of atoms of helium, lithium, and a number of inert gases attract attention of both theoreticians [29] and experimentalists [28] primarily in the context of applied studies such as superfluidity, superconductivity, Bose condensation, chemistry and physics of clusters, and laser physics (the possibility of developing He_2^+ molecular laser), as well as due to the possibility of observing unusual quantum phenomena such as the Efimov effect in real systems.

However, a direct theoretical analysis of even the simplest of the above systems, viz., He_3 consisting of three helium nuclei and six electrons, is an extremely complicated problem.

To analyze the He_3 system, we consider the cluster approximation in which this system is replaced by a simpler system consisting of three force centers (helium atoms). The validity of this approximation for calculations of bound states is obvious since the difference between the binding energy of the system and the ionization energy of the atom is several orders of magnitude. It is well known that helium atoms are bosons; consequently, the problem boils down to analysis of three pairwise identical neutral spinless particles. To solve this problem, we propose mathematically correct model-free methods in the theory of scattering in the three-body system [13–15].

It should be emphasized that virtual levels in paired subsystems in the case of complex many-particle systems do not lead to the emergence of resonant states in a many-particle system [1]. This, however, does not mean that this effect is absent in these systems since it can be due to many-particle and not two-particle virtual states.

For this reason, we will consider the interpretation of a number of peculiar properties of He_3, Ar_3, Kr_3, Ne_3, Xe_3, Li_3, and Rn_3 systems precisely on the basis of the three-particle approximation. It should be noted that a large number of theoretical and experimental methods exist for studying clusters consisting of atoms of helium and a number of inert gases. Most methods are intended for studying bound states; however, scattering states [28–31], which are most informative for confirming the existence of peculiar resonant states, were practically ignored.

It was stated by a number of authors [30] that the main difficulties in studying the He_3 system are associated with its low binding energy (~1 mK), an unusually large size of the excited state (~150 Å), and a strong repulsion at small distances. However, the results obtained in [15, 31], where an analogous three-particle approximation was used for calculating the He_3 system, differ from the statements made in [30].

For this reason, it would be also interesting to verify the conclusions drawn in [30] on the basis of the three-particle approximation with the short-range pair potentials used in [32]. The main purposes of this investigation are

(i) determining the number of possible resonant states;

(ii) clarifying the influence of pair interaction potentials on the characteristics of these states;

(iii) estimating the effect of repulsion at short distances, which can be approximated by a hard core in the

model for the boundary conditions [13–15] imposed on the characteristics of these peculiar states.

Thus, the theoretical analysis of the He_3 system is reduced to solving equations in the quantum theory of scattering in a three-body system, which makes it possible to use the well-known methods [13–15]. In contrast to [30], where resonances in a three-particle system were studied using the Faddeev equations on the basis of analytic continuation of the scattering matrix to the range of complex energy values, we use here direct numerical solution without an analytic continuation.

In this case, after the separation of angular variables, the Faddeev equations (6) in the coordinate space for the He_3 system in the three-particle approximation with pair short-range potentials [32] have the form [13–15]

$$[H_{\lambda,l}-z]\Psi_{aL}(x,y) = -V(x)\Bigg(\Psi_{aL}(x,y) + \sum_{a'}\int_{-1}^{+1}\Psi_{a'L}(x',y')h^{L}_{aa'}(x',y',\eta)d\eta\Bigg), \quad (9)$$

where

$$H_{\lambda,l} = -\frac{\partial^2}{\partial x^2}-\frac{\partial^2}{\partial y^2}+\frac{l(l+1)}{x^2}+\frac{\lambda(\lambda+1)}{y^2},$$

$$z = E+i0, \quad L = l+\lambda, \quad a = (l,\lambda).$$

For calculations with a hard core in the model of boundary conditions, the right-hand side is equal to zero for $x < c$, where c is the core size. To obtain an unambiguous solution to the equations, we must preset boundary conditions (7), (8),

$$\Psi_{aL}(x,y)|_{x=0} = 0, \quad \Psi_{aL}(x,y)|_{x=0} = 0, \quad (10)$$

which assume the following form in the boundary-condition model:

$$\Psi_{aL}(c,y)+\sum_{a'}\int_{-1}^{1}\Psi_{a'L}(x',y')h^{L}_{aa'}(x',y')\eta\,d\eta = 0,$$

$$x' = \left(\frac{x^2}{4}+\frac{3y^2}{4}-\frac{\sqrt{3}xy\eta}{2}\right)^{1/2},$$

$$y' = \left(\frac{3x^2}{4}+\frac{y^2}{4}-\frac{\sqrt{3}xy\eta}{2}\right)^{1/2}.$$

For $\rho \longrightarrow \infty$, the boundary conditions in the case of short-range pair potentials can be written in the form [13]

$$\Psi_{aL} \xrightarrow[\rho\to\infty]{} a_{aL,\nu}\sum_{\nu}\psi_{l,\nu}(x)H_{\nu}(\sqrt{E-E_{2,l,\nu}}) + A_{aL}(\theta)\frac{\exp(i\sqrt{E}\rho)+i\pi L/2}{\sqrt{\rho}}, \quad (11)$$

where $\psi_{l,\nu}(x)$ are the partial components of the wave functions of paired subsystems with binding energy $\epsilon_{l,\nu}$; $\rho = \sqrt{x^2+y^2}$, $\theta = \arctan(y/x)$; $a_{aL,\nu}$ and $A_{aL}(\theta)$ are the scattering amplitudes of processes with two or three particles, respectively, in the final state; and $H_\nu(x)$ are the Hankel spherical functions.

In calculations of bound states, the wavefunctions decrease quite rapidly at infinity; consequently, at a large distance $x = R_x$, $y = R_y$, the asymptotic boundary conditions can be replaced by the conditions

$$\frac{\partial_x\Psi_{aL}|_{x=R_x}}{\Psi_{aL}|_{x=R_x}} = i\sqrt{\epsilon_\nu},$$

$$\frac{\partial_y\Psi_{aL}|_{y=R_y}}{\Psi_{aL}|_{y=R_y}} = i\sqrt{\epsilon_\nu-E}.$$

For the He_3 system in the three-particle approximation with angular momentum $L = 0$, we have

$$H_{\lambda,l} = H_{0,l} = -\frac{\partial^2}{\partial x^2}-\frac{\partial^2}{\partial y^2}+l(l+1)\left(\frac{1}{x^2}+\frac{1}{y^2}\right),$$

where partial components l assume even values; $l = 0, 2, 4, \ldots$; and the expression for functions $h^{L}_{aa'}(x,y,\eta)$ is given in [13–15].

The asymptotic behavior of the components of Eqs. (9) for scattering processes with short-range potentials can be described by the function [13–15]

$$\Psi_l(x,y;z) = \delta_{l0}\psi_d(x)[\sin(y\sqrt{z-\epsilon_d}) + \exp(iy\sqrt{z-\epsilon_d})[a_0(z)+o(y^{-1/2})]] + \frac{\exp(i\sqrt{z}\rho)}{\sqrt{\rho}}[A_l(z,\theta)+o(\rho^{-1/2})], \quad (12)$$

where $a_0(z)$, $z = E + i0$ is the elastic scattering amplitude for $E > \epsilon_d$, and $A_l(E,\theta)$ is the decay amplitude for $E > 0$. We also assume that the helium molecule $^4He_2$ has only one bound state with binding energy $\epsilon_d < 0$ and with a corresponding wavefunction $\psi_d(x)$.

For processes of scattering, the scattering matrix for $z = E + i0$, $E > \epsilon_d$, the scattering phases and lengths in the s state can be expressed with the help of the following formulas:

$$S_0(z) = 1+2ia_0(z),$$

$$\delta_0(p) = \frac{1}{2}\mathrm{Im}\ln S_0(\epsilon_d+p^2+i0), \quad p>0,$$

$$L_{sl} = -\frac{\sqrt{3}}{2}\lim_{p\to 0}\frac{a_0(p)}{p}.$$

Table 1. Binding energy, coefficient of clusterization, scattering length, mean radius, and the square of mean radius for various molecules

Potential	E_{He_3}, mK	$\|f_c\|^2$	L_{sl}, Å	$\langle r_{He_3}\rangle$, Å	$\langle r^2_{He_3}\rangle^{1/2}$, Å
He_3 molecule					
HFDHE2	–0.1171	0.2094	140	5.65	6.46
HFD-B	–0.1330	0.2717	137	5.48	6.23
HFD-ID	–0.1061	0.1555	139	5.80	6.64
LM2M1	–0.1247	0.2412	132	5.57	6.35
LM2M2	–0.1264	0.2479	131	5.55	6.32
TTYPT	–0.1264	0.2487	130	5.56	6.33
He_3 molecule, boundary-condition model					
HFDHE2	–0.1170	0.2095	138	5.65	6.46
HFD-B	–0.1329	0.2717	135	5.48	6.23
HFD-ID	–0.10612	0.1555	134	5.80	6.64
LM2M1	–0.12465	0.2412	130	5.57	6.35
LM2M2	–0.12641	0.2479	131	5.55	6.32
TTYPT	–0.12640	0.2487	131	5.56	6.33
He_3^* molecule					
HFDHE2	–1.6653	0.9077	134	55.26	66.25
HFD-B	–2.7430	0.9432	135	48.33	57.89
HFD-ID	–1.0612	0.8537	140	62.75	75.38
LM2M1	–2.1550	0.9283	129	51.53	61.74
LM2M2	–2.2713	0.9319	131	50.79	60.85
TTYPT	–2.2806	0.9323	131	50.76	60.81
He_3^* molecule, boundary-condition model					
HFDHE2	–1.6765	0.9078	135	56.22	67.11
HFD-B	–2.7458	0.9439	135	48.31	58.00
HFD-ID	–1.1061	0.8597	136	62.87	76.13
LM2M1	–2.2585	0.9323	132	52.41	62.04
LM2M2	–2.2801	0.9319	131	50.79	61.05
TTYPT	–2.2885	0.9339	131	51.23	60.89

Table 2. Binding energies of inert gas molecules calculated by using the HFD-B potential

Energy, 10^{-6} at. units	Ne_2	Ar_2	Kr_2	Xe_2	Rn_2
E_{thr}	178	394	619	854	9268
E_{exp}	135	446	629	874	–

To solve the system of equations (9) with boundary conditions (10), (12) numerically, we used the standard method described in detail in [13–15, 17]. For pair interaction potentials, we used potentials HFDHE2, HFD-B, HFDID, LM2M1, LM2M2, and TTYPT with appropriate parameters [32], which reproduce in detail the main parameters of the corresponding molecules [18].

The results of calculation of the energy of bound states in systems He_3 and He_3^* with and without taking into account the hard core are given in Tables 1–3.

Interpretation of the geometric characteristic of the He_3 molecule in both ground and excited states was given in [31]. Using the methods developed in these papers, let us consider the geometric characteristics of Ne_3 and Ne_3^* molecules which are of considerable interest in the context of investigations into Bose condensation, superconductivity, and superfluidity. The results of calculation of the density function defined as [31]

$$\rho(\mathbf{r}_1) = \int |F(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)|^2 d\mathbf{r}_2 d\mathbf{r}_3,$$

where

$$F(\mathrm{r}_1, \mathbf{r}_2, \mathbf{r}_3) = \frac{1}{2\pi xy}\left\{\Psi(x, y, z') + xy\left[\frac{\Psi(x^+, y^+, z'^+)}{x^+ y^+} + \frac{\Psi(x^-, y^-, z'^-)}{x^- y^-}\right]\right\},$$

$$z' = \frac{\mathbf{x}\cdot\mathbf{y}}{xy},$$

$$x^\pm = \left(\frac{x^2}{4} + \frac{3y^2}{4} \mp \frac{\sqrt{3}xyz'}{2}\right)^{1/2},$$

$$y^\pm = \left(\frac{3x^2}{4} + \frac{y^2}{4} \mp \frac{\sqrt{3}xyz'}{2}\right)^{1/2},$$

are presented in Fig. 12. This function has the form

$$\rho(r) = \frac{\sqrt{3}}{4\pi^2 r^2}\int |F(x, r\sqrt{3}, z')|^2 dx dz'.$$

A sufficiently clear representation of the geometric characteristics of the molecules of inert gases is provided by plotting this function in the coordinates r_l, r_a, where

$$r_l = rz',\ r_a = \frac{z'}{|z'|} r(1 - z'^2)^{1/2}.$$

Note that, for excited states of the inert gas molecules, as well as for the molecules of He_3^* [31] and Ne_3^* (Fig. 12), this function has two peaks, which corresponds to a linear structure. This implies that the third particle in the excited state is located with a high probability between two other particles (as if this state corresponded to two combined paired subsystems). It is precisely this configuration that corresponds to the conditions for the emergence of the Efimov effect in a three-particle system. This conclusion is confirmed by calculations of the clusterization coefficient defined by the formula [31]

$$f_c = \int \Psi(x, y, z')\phi_2(x)az'dx.$$

The results of such calculations are given in Table 1. It can be seen that two-particle states dominate in the excited state He_3^*, while their role in the ground state is insignificant. In the ground state, system He_3 forms a nearly equilateral triangle, while in the excited state, one of the atoms is at a large distance from the other two atoms. Other excited states can be obtained by the similitude method [1, 13–15].

An analogous structure is formed in the calculation of the systems Ne_3, Ar_3, Kr_3, Xe_3, and Rn_3 using the three-particle approximation. The results of calculation of these systems in the given approximation with the HFD-B potential and the parameters borrowed from [32] are presented in Tables 2 and 3.

In calculations based on the boundary-condition model, the value of core c was chosen so that even a slight change in this quantity did not affect the binding energy of paired subsystems. In our calculations, c = 1.5 Å, the value of binding energy for the helium molecule was 1.69 mK, and the value of r_0 was 100 Å. A detailed description of the numerical method for solving system of equations (9) with asymptotic boundary conditions (11), and Eqs. (12) is given in [13–15].

It should be noted that, according to our calculations, the size of the ground state of the He_3 system is smaller than the size of the He_2 molecule. However, the size of the excited state He_3^* of the three particle system is much larger than that of the two-particle system He_2. The experimental data [28] confirm this statement. Thus, in the given approximation, the results of calculations indicate that peculiar resonant states can exist in the He_3 system, the number of such states being not more than two.

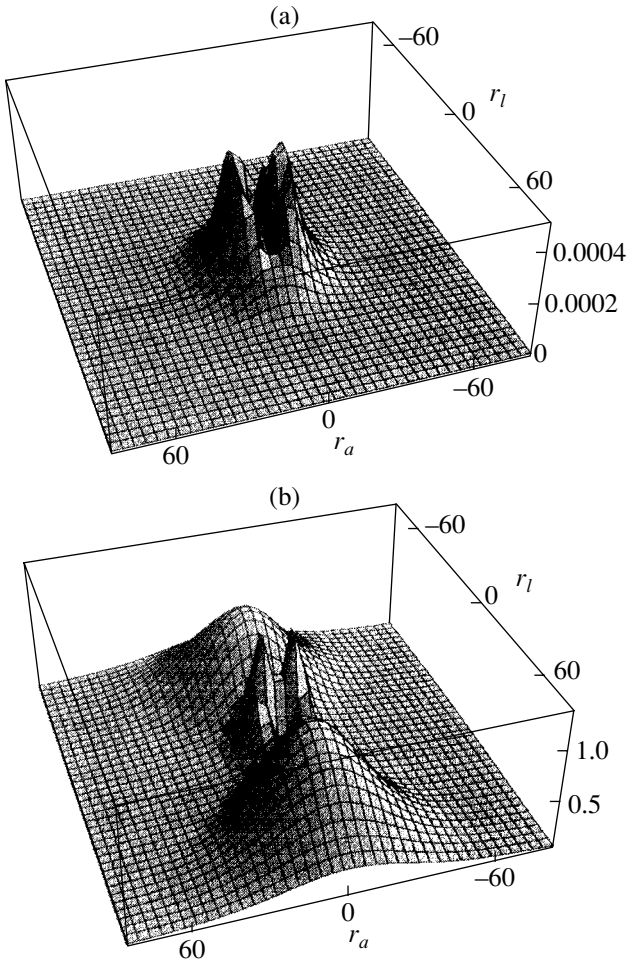


Fig. 12. The density functions for Ne_3^* molecule in the (a) ground and (b) excited states.

To study the scattering processes occurring during the collision of an atom with a helium molecule and to determine the role of pair interaction potentials, we calculated the amplitudes of elastic scattering and decay as well as phase shifts with and without taking into account the hard core. The results of these calculations are depicted in Fig. 13. The results are almost independent of the form of pair interaction potentials and on whether or not the hard core was taken into account both for bound states and for scattering state.

Table 3. Binding energies of the ground state and the first excited state of the inert gas molecule trimers calculated by using the HFD-B potential

	Ne_3	Ne_3^*	Ar_3	Ar_3^*	Kr_3	Kr_3^*	Xe_3	Xe_3^*	Rn_3	Rn_3^*
Energy, 10^{-6} at. units	398	330	1278	1215	1885	1811	2509	2438	30875	30801

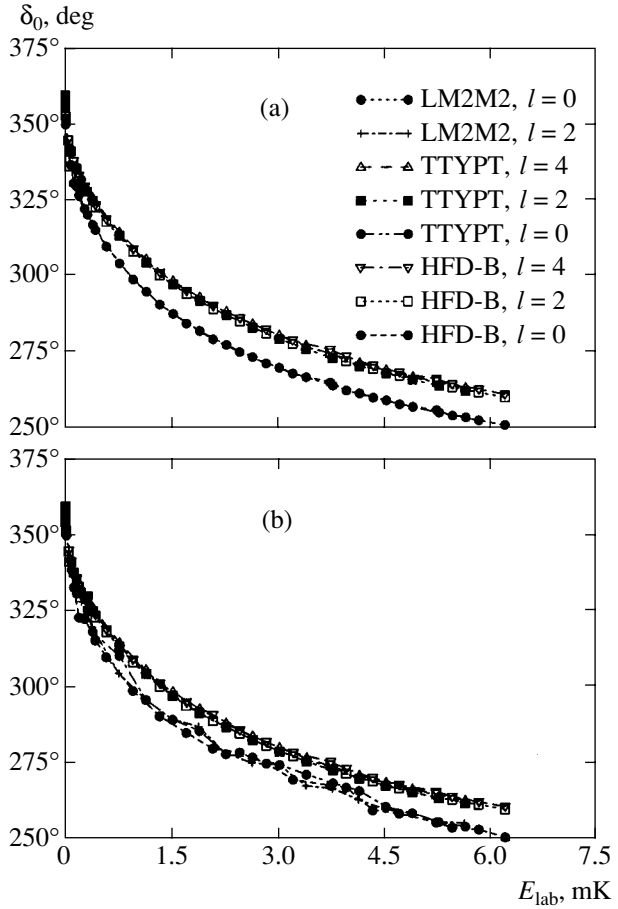


Fig. 13. Dependence of phase shifts on energy $\delta_0(E_{lab})$, $E_{lab} = (3/2)(E + |\epsilon_d|)$ for collisions between the helium atom and molecule for potentials HFD-B, LM2M2, TTYPT [32] calculated (a) without and (b) with taking into account the solid core.

Thus, it can be concluded that the form of pair interaction potentials and allowance for a hard core in the boundary-condition model in the given approximation does not substantially affect the results of calculations.

ACKNOWLEDGMENTS

This study was supported by the Taiwan Academy of Sciences (grant no NSC 85-212-M-007-009), Scientific Research Foundation of the China People's Republic (grant no. NSF 19734030), Joint Research Foundation of Israel and the USA, and the Russian Foundation for Basic Research (project nos. 98-02-17266, 01-02-16075, 04-02-08241 OFI-A, and 05-02-16506).

REFERENCES

1. V. Efimov, *Influence of Resonance in Pair Forces on the Spectrum of Three-Particle Levels* (Mosk. Inzh.–Fiz. Inst., Moscow, 1973) [in Russian]; Nucl. Phys. A **362**, 45 (1981); **378**, 581 (1982); Phys. Rev. C **47**, 1876 (1993); S. A. Vugal'ter and G. M. Zhislin, Dokl. Akad. Nauk SSSR **267**, 784 (1982).
2. T. Andras, *Resonances in Few-Body Systems* (Springer, New York, 2002); V. I. Kukulin, K. M. Krasnopolsky, and J. Horacek, *Theory of Resonances. Principles and Applications* (Kluver, Dordrecht, 1989).
3. A. Poincaré, *Selected Works* (Nauka, Moscow, 1971–1972), Vols. 1–2.
4. G. J. Schultz, Rev. Mod. Phys. **45**, 423 (1973).
5. G. F. Drukarev, *Collisions of Electrons with Atoms and Molecules* (Nauka, Moscow, 1978; Plenum, New York, 1987).
6. Yu. N. Demkov and V. N. Ostrovskiĭ, *Zero-Range Potentials and Their Applications in Atomic Physics* (Leningr. Gos. Univ., Leningrad, 1975; Plenum, New York, 1988).
7. L. G. Christophorou, *Electron Molecule Interaction and Their Application* (Academic, New York, 1984).
8. H. S. W. Massey, *Negative Ions* (Cambridge Univ. Press, Cambridge, 1976; Mir, Moscow, 1979).
9. A. Herzenberg, *Electron–Molecular Collision* (Plenum, New York, 1984).
10. W. Domcke, Phys. Rep. **208**, 98 (1991); A. K. Kazanskiĭ and I. I. Fabrikant, Usp. Fiz. Nauk **143**, 602 (1984) [Sov. Phys. Usp. **27**, 607 (1984)].
11. A. Chutjian, A. Garscadden, and J. M. Wadehra, Phys. Rep. **264**, 393 (1996).
12. I. Prigogine and I. Strengers, *Quantum, Chaos, Time. To a Solution of the Time Paradox* (Editorial URSS, Moscow, 2003).
13. S. P. Merkur'ev and L. D. Faddeev, *Quantum Theory of Scattering for Several Particle Systems* (Nauka, Moscow, 1985) [in Russian]; D. R. Yafaev, *Mathematical Theory of Scattering* (S.-Peterb. Gos. Univ., St. Petersburg, 1994) [in Russian]; V. B. Belyaev, *Lectures on the Theory of Small-Particle Systems* (Énergoizdat, Moscow, 1986) [in Russian].
14. S. A. Pozdneev, *Application of Quantum Theory of Scattering to Three-Body System for Calculating Various Processes in Nuclear, Atomic, and Molecular Physics* (Yanus-K, Moscow, 2001) [in Russian].
15. S. A. Pozdneev, Kratk. Soobshch. Fiz., No. 6, 88 (1987); No. 1, 3 (2003); No. 5, 32 (2003); Phys. Lett. B **125B**, 335 (1983); J. Russ. Laser Res. **22**, 175 (2001); in *Book of Abstracts of XXI ICPEAC* (Stockholm, Sweeden, 2003), p. 41; in *Book of Abstracts of Heron Island International Conference on Reactive Intermediates and Unusual Molecules* (Univ. of Queensland, Brisbane, Australia, 2003), p. 123.
16. T. K. Rebane, Yad. Fiz. **61**, 61 (1998) [Phys. At. Nucl. **61**, 56 (1998)].
17. S. A. Pozdneev, Zh. Éksp. Teor. Fiz. **77**, 38 (1979) [Sov. Phys. JETP **50**, 18 (1979)]; Zh. Éksp. Teor. Fiz. **117**, 35 (2000) [JETP **90**, 30 (2000)]; J. Phys. B: At. Mol. Phys. **16**, 867 (1983); S. Pozdneev and G. F. Drukarev, J. Phys. B: At. Mol. Phys. **13**, 2611 (1980).
18. K. R. Huber and G. Herzberg, *Molecular Spectra and Molecular Structure*, Vol. 4: *Constants of Diatomic Molecules* (Van Nostrand Reinhold, New York, 1979; Mir, Moscow, 1981).
19. H. Drexel, G. Senn, T. Fiegele, *et al.*, J. Phys. B: At. Mol. Opt. Phys. **34**, 1415 (2001).

20. A. P. Hickman, Phys. Rev. A **43**, 3495 (1991).

21. G. A. Gallup, Y. Xu, and I. I. Fabricant, Phys. Rev. A **57**, 2596 (1998); A. K. Kazansky, Y. Xu, and I. I. Fabricant, Phys. Rev. A **63**, 014703 (2001); I. I. Fabricant and H. Hotop, Phys. Rev. A **63**, 022706 (2001).

22. V. V. Pupyshev, Yad. Fiz. **66**, 64 (2003) [Phys. At. Nucl. **66**, 61 (2003)]; J. Phys. A: Math. Gen. **36**, L13 (2003).

23. A. V. Finkel'shteĭn and O. B. Ptitsin, *Physics of Protein: Course of Lectures* (Knizhnyĭ Dom Universitet, Moscow, 2002) [in Russian]; D. S. Chernavskiĭ and N. M. Chernavskaya, *Protein-Machine. Biological Macromolecular Constructions* (Mosk. Gos. Univ., Moscow, 1999) [in Russian].

24. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 3: *Quantum Mechanics: Non-Relativistic Theory*, 4th ed. (Nauka, Moscow, 1989; Oxford Univ. Press, Oxford, 1980); B. N. Zakhar'ev and V. M. Chabanov, *Controllable Quantum Mechanics. A New Status of the Theory of the Approach to Inverse Problem* (Inst. Komp'yut. Issled., Moscow, 2002) [in Russian].

25. F. M. Pen'kov, Zh. Éksp. Teor. Fiz. **118**, 806 (2000) [JETP **91**, 698 (2000)]; Phys. Rev. A **62**, 044701 (2000); A. K. Motovilov and F. M. Pen'kov, *Selected Problems of Theoretical Physics and Astrophysics* (Ob. Inst. Yad. Issled., Dubna, 2003) [in Russian].

26. Y. Horacek and W. Domke, Phys. Rev. A **53**, 2262 (1996).

27. S. A. Pozdneev, Khim. Vys. Énerg. **18**, 280 (1984).

28. F. Lon, C. F. Giese, and W. R. Gentry, J. Chem. Phys. **104**, 1151 (1996); W. Schollkopf and J. P. Toennies, J. Chem. Phys. **104**, 1155 (1996); Phys. Rev. Lett. **85**, 2284 (2000); M. V. Rama Krishna and K. V. Whaley, Phys. Rev. Lett. **64**, 1126 (1990); G. C. Hegerfeldt and T. Kohler, Phys. Rev. Lett. **84**, 3215 (2000).

29. D. V. Fedorov and A. S. Jensen, Phys. Rev. C **50**, 2372 (1994); A. Cobis, Phys. Lett. B **424**, 1 (1998); J. Yuan and C. D. Lin, J. Phys. B: At. Mol. Opt. Phys. **31**, L637 (1998); T. Gonzalez-Lezana, J. Rubayo-Soneira, S. Miret-Aztes, *et al.*, J. Chem. Phys. **110**, 9000 (1999); Phys. Rev. Lett. **82**, 1648 (1999); T. Frederico, L. Tomio, A. Delfino, *et al.*, Phys. Rev. A **60**, R9 (1999); Y. Hahn, Phys. Rev. A **60**, 2139 (1999); E. Nielsen, D. V. Fedorov, and A. S. Jensen, Phys. Rev. Lett. **82**, 2844 (1999); P. F. Bedaque, H. W. Hammer, and U. Van-Kolek, Phys. Rev. Lett. **82**, 463 (1999); R. A. Ionescu and C. Nategan, Europhys. Lett. **45**, 269 (1999).

30. A. K. Motovilov, E. A. Kalganova, and S. A. Sofianos, J. Phys. B: At. Mol. Opt. Phys. **31**, 1279 (1998); J. Chem. Phys. **275**, 168 (1997); Phys. Rev. A **56**, R1686 (1997).

31. V. Rudnev and S. Yakovlev, Chem. Phys. Lett. **22**, 97 (2000); Phys. At. Nucl. **63**, 61 (2000); **63**, 77 (2000); **63**, 271 (2000); **63**, 278 (2000); **63**, 402 (2000); **63**, 409 (2000); **63**, 830 (2000).

32. A. R. Janzen and R. A. Aziz, J. Chem. Phys. **79**, 4330 (1979); **94**, 8047 (1991); **103**, 9626 (1995); Mol. Phys. **61**, 1487 (1987); K. T. Tang, J. P. Toenries, and C. L. Yiu, Phys. Rev. Lett. **74**, 1546 (1995); S. Pozdneev, in *Proceedings of Laser Chemistry, Biophysics and Biomedicine ICONO'95* (1996), p. 92; Proc. SPIE, 96 (1996); Kratk. Soobshch. Fiz., Nos. 1–2, 61 (1997); W. J. Meath, D. J. Margoliash, B. L. Jhanwar, *et al.*, *Intermolecular Interactions: From Diatomics to Biopolymers* (Wiley–Interscience, London, 1978).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 99, No. 5, 2004, pp. 934–941.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 5, 2004, pp. 1073–1082.

ATOMS, SPECTRA, RADIATION

Picosecond Stimulated Raman Scattering in Crystals

T. T. Basiev, P. G. Zverev, A. Ya. Karasik, V. V. Osiko, A. A. Sobol', and D. S. Chunaev

Scientific Center of Laser Materials and Technologies, Prokhorov Institute of General Physics, Moscow, 119991 Russia
e-mail: karasik@lst.gpi.ru

Received March 17, 2004

Abstract—The comparative values of the peak and integral cross sections of spontaneous Raman scattering and the optical dephasing time of molecular vibrations were determined for several oxide crystals by spontaneous Raman spectroscopy. The spectral, time, and energy parameters of stimulated Raman scattering (SRS) were measured for ten crystals using picosecond YLF : Nd laser pumping with a radiation wavelength of 1047 nm. An analysis of the experimental dependence of the threshold energy of pumping SRS on the integral and peak cross sections of spontaneous Raman scattering showed that the SRS gain increment explicitly depended on the integral cross section and was independent of the peak cross section of spontaneous Raman scattering as the ratio between the pumping pulse width (11 ps) and the time of optical dephasing of molecular vibrations changed from 0.42 to 9.3. The gain coefficients of steady-state stimulated Raman scattering under threshold stimulated Raman scattering conditions were determined for all the crystals studied on the basis of the measured threshold SRS pumping energies, the duration and width of the spectrum of pulses, the nonlinear interaction length, the intensity of pumping, and the theoretical dependences that relate the steady-state and transient SRS gain increments. The steady-state SRS gain coefficients obtained in this work fitted well a linear dependence on the peak cross sections of spontaneous Raman scattering, which substantiated the correctness of our analysis and measurements. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Stimulated Raman scattering (SRS) under picosecond laser pumping conditions had been extensively studied in the late 1960s–early 1970s [1]. This was caused by the observation of new (compared with nanosecond pumping) features of the process related in particular to manifestations of the transient character or violation of coherence of SRS and the influence of dispersion and self-phase modulation on the time, spectral, and spatial radiation parameters. The theory of transient SRS (see reviews [1, 2]) explained well some of the experimental results obtained in gases and liquids [3] and, more recently, in quartz optical fibers under the conditions of a large nonlinear interaction length between radiation and media [4]. Picosecond SRS was also studied in well-known calcite and lithium niobate crystals, more recently, in popular $KGd(WO_4)_2$ (KGW) crystals [5], and others. In recent years, interest in nanosecond and picosecond SRS in crystals has increased because crystals offer much promise for use in SRS frequency converters of laser radiation and for the advancement into the infrared spectral range.

Currently, several new promising crystals active in SRS have come into view [6–10], and the search for new ones continues. Although the general laws governing SRS in liquid, gaseous, and solid media should be similar, there are certain differences related, for instance, to different phase relaxation rates and inhomogeneity of the spectrum of vibrational resonances [1], dispersion [2–4], the mechanism and dynamics of self-phase modulation [11, 12], and the mechanical and thermophysical properties of media.

As the selection of the object of study and the modern experimental methods and laser techniques allow us to move from the micro- to femtosecond time range in studying SRS [13–15], of special interest is, in our view, correspondence between experimental data and theoretical concepts of transition from the steady-state to transient SRS mode. As is well known, there is no strict time boundaries that separate steady-state and transient SRS, which impedes using one or another theory and a comparison of experimental results. If the response of a system to an exciting light pulse of width τ_p is linear, the nonstationarity condition for the system with the phase relaxation time T_2 is comparatively clear, namely, $\tau_p < T_2$, but if the response is nonlinear, this condition can be substantially different. This is caused by the transformation of the time envelope of the pulse and its spectrum in a nonlinear process.

If spontaneous intensity I_{S0} noise increases exponentially under undepleted pumping conditions, the intensity of the amplified wave at a Stokes frequency shifted with respect to the pumping frequency increases as $I_S = I_{S0}\exp G$ for both steady-state and transient SRS [2]. For steady-state SRS, when the pumping pulse width τ_p is much larger than the phase relaxation time of Raman scattering of the active vibrational mode, the $G_0 = gI_pL$ gain increment is linearly proportional to the intensity of pumping I_p, the length L of the

interaction of light with the medium, and the SRS gain coefficient g. As a result of SRS amplification, the spectrum of the Stokes pulse narrows. At G_0 varying from single digits to hundreds, this can substantially transform the transient system response condition $G_0T_2 \gg \tau_p$ [1]. On the other hand, under transient SRS conditions, the Stokes pulse during amplification can narrow with respect to the pumping pulse [2, 3], which complicates an analysis of the nonstationarity conditions. The theory developed in [2], which combines analytic and numerical methods, was used to determine the relation between steady-state and transient SRS gain increments over wide ranges of their values, in particular, for the τ_p/T_2 ratios that ranged from 10 to 0.01.

In the limit of large gain coefficients, the relation between steady-state and transient gain increments has the form [2]

$$G_\tau = \sqrt{4G_0\frac{\tau_p}{T_2}}, \quad (1)$$

where $T_2 = 1/\pi c\Delta\nu_R$ and $\Delta\nu_R$ is the width of the spontaneous Raman scattering spectrum. Equation (1) can easily be transformed into

$$G_\tau = \left(\frac{4\pi\lambda_s^2\lambda_p\sigma_{\text{int}}}{hcn_s^2}I_pL\right)^{1/2}, \quad (2)$$

where λ_p and λ_s are the wavelengths of pumping and Stokes scattering shifted to the longer waves, n_s is the refractive index at the Stokes frequency, σ_{int} is the integral cross section of spontaneous Raman scattering,

$$\sigma_{\text{int}} = \int\sigma(\nu)d\nu \approx \sigma_p\Delta_\nu,$$

and $\nu = 1/\lambda$ is the radiation frequency. The intensity of pumping radiation is $I_p = E_p/\tau_pS_{\text{eff}}$, where E_p is the pumping energy and S_{eff} is the effective cross section area of the pumping beam. The steady-state SRS gain coefficient g is related to the peak cross section σ_p of spontaneous Raman scattering as [1]

$$g(\nu) = \frac{\sigma_p\lambda_s^2\lambda_p}{c^2hn_s^2}. \quad (3)$$

Note that, as $\sigma_p(\sigma_{\text{int}}) \sim \lambda_s^{-4}$ [1], the steady-state Raman gain coefficient g (accordingly, the G_0 gain increment too) is linearly proportional to the radiation frequency. (A quadratic frequency dependence of the SRS gain coefficient was observed in several experiments, probably because of a contribution of resonance electronic transitions in the ultraviolet spectral range to the nonlinear process [6].) It follows in particular from (2) that, in conformity with theory [2], at large amplifications and if dispersion is ignored, the transient gain increment G_τ is proportional to the square root of the integral spontaneous Raman scattering cross section σ_{int} and independent of the width of the radiation spectrum, the width of the pumping pulse τ_p, and the time of phonon optical dephasing T_2.

In this work, we studied SRS in many crystals with molecular vibration optical dephasing times T_2 varying within more than an order of magnitude. We were therefore able to analyze the characteristics of the transition from steady-state to transient SRS and estimate the most important parameters of the nonlinear process at a constant picosecond laser pumping.

Spontaneous Raman spectroscopy techniques allowed us to obtain comparative peak and integral SRS cross sections that determine the steady-state and transient SRS gain increments in various crystals. For ten oxide crystals, we measured the spectral, time, and energy SRS parameters under pumping with a YLF : Nd laser (pulse width about 11 ps, radiation wavelength 1047 nm). An analysis of the experimental dependence of the threshold SRS pumping energy on the integral and peak cross sections of spontaneous Raman scattering led us to conclude that, in agreement with the concept of transient SRS, the SRS gain increment explicitly depended on the integral cross sections and was independent of the spontaneous Raman scattering peak cross section over the range of changes in the τ_p/T_2 ratio between the pumping pulse width τ_p = 11 ns and the time T_2 of optical dephasing of molecular vibrations from 0.42 to 9.3. We used the theoretical dependences relating the steady-state and transient SRS gain increments calculated in [2] for various τ_p/T_2 ratios to estimate the steady-state SRS gain increments G_0 under the threshold SRS conditions for all the crystals studied in this work. Next, we determined nonlinear interaction lengths L and pumping intensities I_p to calculate the steady-state SRS gain coefficients $g = G_0/I_pL$ from the measured threshold SRS pumping energies and the duration and width of the spectrum of pulses. This allowed us to find the dependence of the gain coefficient g on the spontaneous Raman scattering peak cross section σ_p for all crystals. We showed that the experimental steady-state SRS gain coefficient g satisfactorily obeyed the linear dependence $g = f(\sigma_p)$ [Eq. (3)], which in turn substantiated the correctness of our analysis and measurements.

2. OPTICAL DEPHASING TIME AND SPONTANEOUS RAMAN SCATTERING PEAK AND INTEGRAL CROSS SECTIONS IN OXIDE CRYSTALS

The crystals studied in this work are listed in Table 1.

SRS in crystals typeset in boldface was studied for the first time. Detailed spectroscopic studies of the series of alkaline-earth metal (Ca, Sr, and Ba) and lead (Pb) tungstates and molybdates were performed in [16, 17]. Table 1 contains the parameters of the most intense Raman-active vibrations observed in spontaneous

Table 1. Spectroscopic parameters of SRS-active modes of the crystals studied at 300 K

Crystal	Raman mode frequency ν_R, cm^{-1}	Spontaneous Raman scattering line width $\Delta\nu_R$, cm^{-1}	Vibrational phase relaxation time $T_2 = 1/\pi c\Delta\nu_R$, ps	Refractive index n	Excitation polarization	Integral cross section* of spontaneous Raman scattering σ_{int}, arb. unit	Peak cross section* of spontaneous Raman scattering σ_p, arb. units
$Ba(NO_3)_2$	1047	0.4	26.5	1.55	$\parallel C_4$	21	63
$BaMoO_4$	892	1.85	5.73	1.98	$\parallel C_4$	55	62
$SrMoO_4$	887	2.5	4.24	1.98	$\parallel C_4$	63	51
$SrWO_4$	921	2.7	3.93	2.13	$\parallel C_4$	50	41
$KGd(WO_4)_2$ (KGW)	901	5.4	1.96	1.9	$\perp C_2$	50	35
$CaMoO_4$	879	5.5	1.93	1.93	$\parallel C_4$	65	34
$CaWO_4$	911	6.9**	1.54	1.98	$\parallel C_4$	52	18
TeO_2	648	9	1.18	2.21	$\perp C_4$	150	35
$PbWO_4$	904.7	4.1	2.59	2.15	$\perp C_4$	171	97
$PbMoO_4$	871	6.0**	1.77	2.25	$\perp C_4$	390	120

* Peak and integral cross sections of spontaneous Raman scattering are given with respect to the cross sections of the Raman-active mode (1300 cm^{-1}) of the diamond single crystal.

Raman scattering spectra and assigned to totally symmetrical ν_1 vibrations of the $[WO_4]^{2-}$ and $[MoO_4]^{2-}$ tetrahedral complexes. These complexes are the scheelite structure fragments of metal tungstate and molybdate crystals under consideration. It was shown earlier that the type of the cation (Ca, Sr, Ba, Pb) substantially influenced the width $\Delta\nu_1$ of the spontaneous Raman scattering line of metal tungstates; this width changed from 6.9 cm^{-1} in $CaWO_4$ to 1.6 cm^{-1} in $BaWO_4$. A similar trend was observed for metal molybdates. It was shown that the widening of the ν_1 line depended on the mass of the cation and the distance between the $[WO_4]^{2-}$ or $[MoO_4]^{2-}$ complexes in the scheelite crystal structure. The $BaWO_4$ and $BaMoO_4$ crystals are characterized by a large mass of the cation and the largest unit cell size among the scheelites, which explains why the spontaneous Raman scattering spectra of these crystals contain the narrowest ν_1 lines. The integral cross section of the ν_1 line remained virtually unchanged in the series of alkaline-earth metal (Ca, Sr, and Ba) tungstates and molybdates. At the same time, the peak cross sections for this line, which are inversely proportional to the width of the Raman-active vibration, are 3–4 times larger for $BaWO_4$ and $BaMoO_4$ than for $CaWO_4$ and $CaMoO_4$.

In this work, we for the first time measured the peak and integral spontaneous Raman scattering cross sections for oriented $SrWO_4$, $PbWO_4$, and $PbMoO_4$ crystals. Table 1 shows that the σ_{int} value for $SrWO_4$ is virtually equal to that for $CaWO_4$ and $BaMoO_4$. These results substantiate our suggestion that the kind of the cation weakly influences the integral cross section σ_{int} for the ν_1 line of totally symmetrical vibrations in the series of alkaline-earth metal (Ca, Sr, and Ba) tungstates. Note that the width of the ν_1 line in $SrWO_4$ is larger than in $BaWO_4$ but smaller than in $CaWO_4$. The peak cross section of this line σ_p in $SrWO_4$ is smaller by 30% than in $BaWO_4$ but larger almost twofold than in $CaWO_4$ [17].

An important special feature of the spontaneous Raman scattering spectra of $PbWO_4$ and $PbMoO_4$ crystals is the dependence of the intensity of the ν_1 line on the geometry of scattering (*hh* and *vh* in Fig. 1), different from that observed in the polarized spontaneous Raman scattering spectra of alkaline-earth metal tungstates and molybdates. The intensity of this line in Ca, Sr, and Ba tungstates and molybdates is highest when the directions of liner exciting light polarization and the C_4 crystallographic axis coincide (**E** $\parallel C_4$, *hh*) (Fig. 1a, Table 1). The intensity ratio I_{hh}/I_{vh} (*vh* corresponds to **E** $\perp C_4$) for these crystals is larger than one, I_{hh}/I_{vh} = 1.2–1.25. For the $PbWO_4$ and $PbMoO_4$ crystals, this ratio is smaller than one and equals 0.67 and 0.53, respectively. This effect is most pronounced for the $PbMoO_4$ crystal, where the intensity of the ν_1 line in the spontaneous Raman scattering spectrum is abnormally high for the *vh* geometry and exceeds the intensity for the *hh* geometry by a factor of 1.9 (Fig. 1b).

Certain special features of the vibrational spectra of $PbWO_4$ and $PbMoO_4$ were also observed in [16, 17],

where the energies of vibrational levels in a series of scheelites were analyzed. The essential difference of the energy diagrams of these crystals from those of the other scheelites was explained by a comparatively large degree of covalence of the bond between Pb and O and the large mass of the Pb cation. This in particular manifested itself by the lowest energy of totally symmetrical stretching vibrations ν_1 for crystalline and molten $PbWO_4$ and $PbMoO_4$ among the scheelites.

Earlier, in the absence of data on the integral intensities of Raman scattering by $PbWO_4$ and $PbMoO_4$, it was assumed that the peak cross section of the ν_1 line of spontaneous Raman scattering is low because of its comparatively large width. Correct measurements of all the parameters for all geometries and crystal orientations showed that the width of the ν_1 line in the spontaneous Raman scattering spectra of $PbWO_4$ and $PbMoO_4$ was close to the line widths for KGW, $CaWO_4$, and $CaMoO_4$. The anomalously high integral cross section values for the ν_1 lines of lead tungstate and molybdate, however, also predetermine their largest peak cross section values in the series of scheelites (Table 1), which offers promise for using these crystals not only in picosecond but also in nanosecond SRS.

Along with the refractive indexes of all crystals, Table 1 contains dephasing times determined from the spontaneous Raman scattering spectrum width $T_2 = 1/\pi c\Delta\nu_R$ on the assumption of homogeneous broadening and the Lorentz shape of the spectrum. As follows from our analysis, the spontaneous Raman scattering spectra of the crystals labeled by asterisks (**) are inhomogeneously broadened, which makes the corresponding estimates of T_2 not unambiguous [1].

Table 1 shows that, for pumping pulses of width τ_p = 10 ps, the formal condition of transient SRS, namely, $G_0T_2 \gg \tau_p$, is certainly satisfied at the conventional threshold value $G_0 = 25$ for the $Ba(NO_3)_2$ crystal with the largest dephasing time $T_2 = 26.5$ ps. At the same time, for the TeO_2 crystal, for which $T_2 = 1.18$ ps is the lowest, the G_0T_2 product is comparable with $\tau_p = 10$ ps under threshold conditions, and the SRS process can be close to steady-state. We stress that these considerations are not strict, and only a comparison of experimental and theoretical data allows the questions put above to be answered.

3. EXPERIMENTAL DATA ON STIMULATED RAMAN SCATTERING AND THEIR CORRESPONDENCE TO THEORETICAL CONCEPTS

In conformity with the above considerations, we selected a $YLiF_4$ crystal laser as the source for pumping SRS. The laser operated in the regime of passive mode-locking and generated at 1047 nm emitting a 200–300 ns train of 11 ps pulses with the total train energy up to 1.5 mJ (Fig. 2). The optical scheme of the experimental unit for studying SRS was similar to that used by us in [10, 18]. Radiation of the laser operating in the TEM_{00} mode was attenuated by neutral light filters and focused by a spherical long-focus (f = 70 cm) mirror onto the sample under study. Laser and scattered radiation passed through a focusing lens and a dichroic filter that cut off pumping radiation and fell onto calibrated fast-response germanium photodiodes.

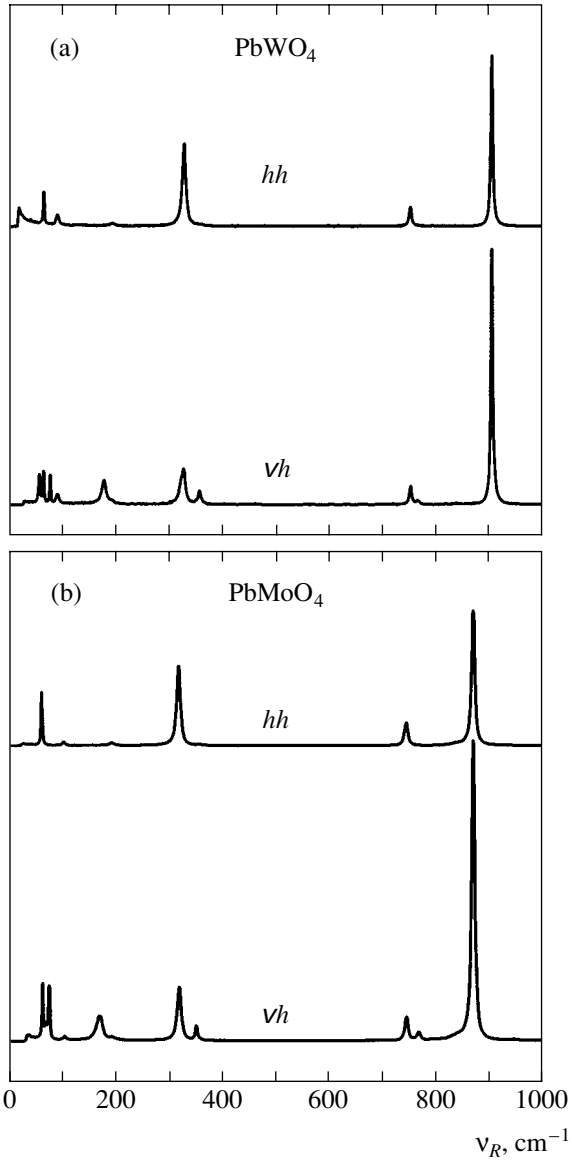


Fig. 1. Spontaneous Raman scattering spectra of (a) $PbWO_4$ and (b) $PbMoO_4$ crystals at linear exciting light polarizations collinear (**E** ∥ C_4, *hh*) and orthogonal (**E** ⊥ C_4, *vh*) with respect to the C_4 axes of these crystals.

During one flash, the energy of separate pulse train and scattered radiation picosecond pulses was measured with a digital oscilloscope (Fig. 2), which allowed us to determine the dependence of the efficiency of SRS on the energy of pumping pulses. The

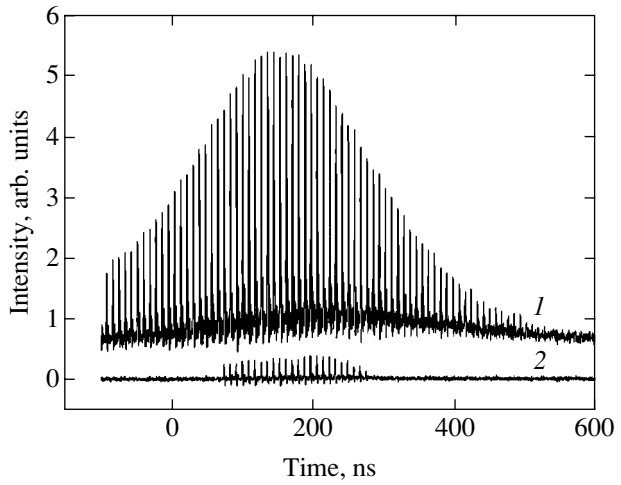


Fig. 2. (*1*) Pumping laser and (*2*) SRS radiation oscillograms for the $PbMoO_4$ crystal.

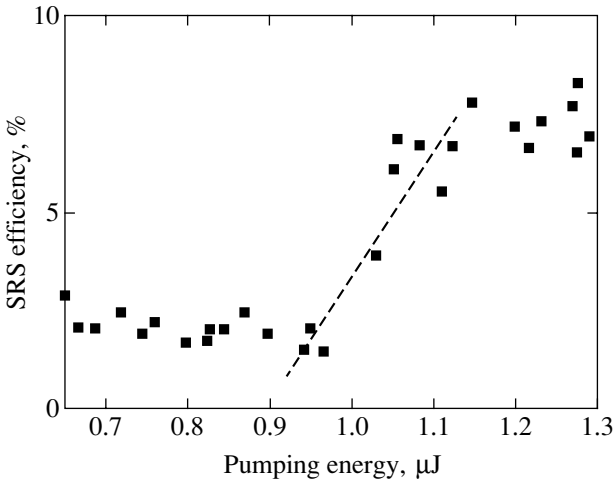


Fig. 3. SRS efficiency as a function of pumping energy for the $PbMoO_4$ crystal.

time parameters of the pulses of scattered and excited radiation were measured with a time resolution of 2–3 ps on an IMACON-501 electron-optical camera interfaced with a reading device on a silicon CCD matrix. The spectral composition of pumping and scattered radiation was analyzed with a PGS-2 grating polychromator with the use of the CCD matrix and a computer. The width of the pumping laser radiation spectrum measured at the second harmonic frequency $\Delta\nu_p$ was 2.26 cm^{-1} ($\Delta\lambda_p$ = 0.062 nm). For the measured width of pulses τ_p = 11 ps at the wavelengths λ = 1047 and 523.5 nm, we had $\tau_p\Delta\nu_p$ = 0.74, which was evidence that pumping laser pulses were not fully limited spectrally [1].

Table 2. Normalized threshold pumping energies $E_p^{thr}L$ and steady-state SRS gain coefficients in crystals

Crystal	L, mm	$E_p^{thr}L$, μJ cm	τ_p/T_2	G_0	g, cm/GW
$Ba(NO_3)_2$	40	24	0.42	300	14.2
$BaMoO_4$	18	13.0	1.92	93	8.1
$SrMoO_4$	36	15.8	2.6	78	5.6
$KGd(WO_4)_2$	36	10.4	5.6	47	5.1
$SrWO_4$	43	18.1	2.8	75	4.7
$CaMoO_4$	28	12.04	5.7	45	4.3
$CaWO_4$	39	21.5	7.14	42	2.2
TeO_2	10	4.3	9.3	37	9.7
$PbWO_4$	30	5.4	4.25	52	10.9
$PbMoO_4$	30	2.6	6.2	40	17.5

The oscillograms of laser pumping radiation and SRS in the $PbMoO_4$ crystal are shown by way of example in Fig. 2. The threshold of the appearance of SRS is clearly seen. The ratio between the energies of SRS and pumping pulse trains gives the dependence of the efficiency of SRS on the pumping energy (Fig. 3). We see that an increase in the efficiency of SRS corresponding to undepleted pumping conditions transforms into saturation with cascade generation of the second Stokes component as the pumping energy increases. Cascade generation of three to four SRS Stokes components with radiation frequencies sequentially decreasing by ν_R (Table 1) with respect to ν_p and of several anti-Stokes components with similar frequency shifts caused by four-photon mixing [1] is easily attainable experimentally. The dependence shown in Fig. 3 was obtained with pumping of comparatively low intensities, when the efficiency of generation of the higher Stokes and anti-Stokes components was negligibly low. We extrapolated the dependences like that shown in Fig. 3 to zero SRS efficiency to determine the threshold SRS pumping energy for all crystals and normalized it by the crystal length ($E_p^{thr}L$) (see Table 2). As follows from the preceding section, spontaneous RS cross sections in the crystals under study can depend on the direction of radiation polarization with respect to the crystallographic axes of the crystals. We measured SRS at pumping linear polarization directed parallel or normally to the crystallographic axes that corresponded to maximum spontaneous RS cross section values.

3.1. An Analysis of Experimental Data

In this section, we analyze the experimental data in terms of the relation between phase relaxation times obtained by linear spectroscopy methods, spontaneous Raman scattering integral and peak cross sections, and

energy, time, and spectral SRS parameters. This relation, determined by (1)–(3), was discussed in the Introduction.

As follows from the measured energy characteristics of SRS, the threshold SRS pumping energies ($E_p^{\text{thr}} L$) of various crystals differ almost by a factor of ten (Table 2). The maximum SRS threshold was obtained for the $Ba(NO_3)_2$ crystal with the smallest integral cross section of spontaneous Raman scattering and the largest vibration dephasing time T_2 = 26.5 ps (Table 1). These features make SRS transient to the greatest degree. SRS thresholds decrease for the group of Ca and Sr tungstate crystals, whose T_2 times are smaller compared with barium nitrate but are substantially different and whose integral cross sections are virtually equal.

SRS thresholds decrease for the group of Ca, Sr, and Ba molybdates, which have somewhat larger integral cross sections compared with Ca and Sr tungstates and different dephasing times T_2. A substantial decrease in the SRS threshold is observed for lead tungstate and molybdate, which are characterized by the largest integral cross sections and minimum dephasing times T_2. For all these groups of crystals, SRS thresholds explicitly depend on the integral cross section of spontaneous Raman scattering and are virtually independent of the peak cross sections or phase relaxation times T_2 in conformity with the theory of transient SRS [2] [see (2)].

This conclusion is visually illustrated by the dependence of the SRS threshold on the integral cross section of spontaneous Raman scattering shown in Fig. 4. According to this figure, the SRS threshold monotonically decreases as the integral cross section of spontaneous Raman scattering increases. If the threshold gain increment value is set at G_τ = 25 for all crystals in Eq. (2) for transient SRS, then (2) gives the hyperbolic dependence $E_p^{\text{thr}} L = f(n^2/\sigma_{\text{int}})$. We do not analyze the linear dependence of this function on the Stokes radiation frequency, whose values are close for the crystals under study (Table 1), but take into account the sharper quadratic dependence on the refractive index [Eq. (2)]. The dependence of the experimental normalized threshold SRS energies on n^2/σ_{int} is shown in Fig. 5 for the crystals studied. The hyperbolic dependence is given by a solid line; it shows that, in agreement with theory and Eq. (2), precisely the integral cross section of spontaneous Raman scattering is responsible for transient SRS amplification. Note that the experimental values for the $PbMoO_4$, $PbWO_4$, and TeO_2 crystals with the shortest times T_2, which make the SRS process close to steady-state, also satisfactorily fit the hyperbolic dependence of the SRS threshold on the integral cross section.

3.2. SRS Amplification Factors

Next, we checked the relation between transient and steady-state SRS and estimated the gain coefficients g of steady-state SRS on the basis of the theory described in [2]. For this purpose, we used the dependences that related the gain increments of transient and steady-state SRS $G_\tau = f(G_0)$ calculated by Carman *et al.* for Gaussian pulses and various ratios between the pumping pulse width and the phonon dephasing time τ_p/T_2, which changed from 10 to 0.01 for the samples under study (Table 1). We assumed that the gain increments of all the crystals were equal, G_τ = 25 [1], and used the results obtained in [2] to determine the steady-state SRS gain increment G_0 for each of them (Table 2). The transient and steady-state SRS gain increments differed most strongly (by a factor of 12) for the $Ba(NO_3)_2$ crys-

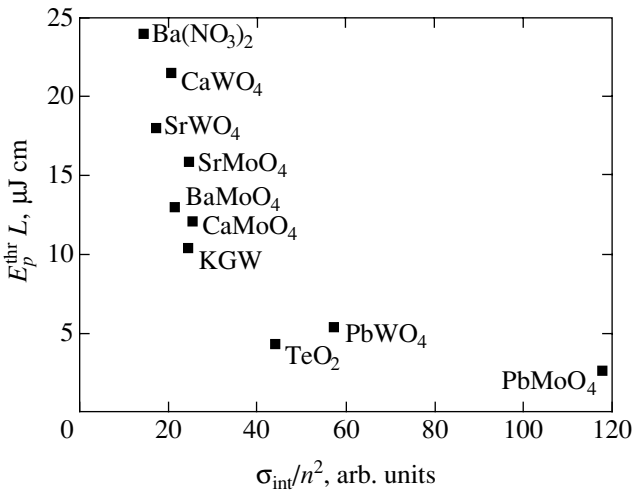


Fig. 4. Threshold energy of SRS pumping normalized by the nonlinear interaction length as a function of the integral cross section of spontaneous Raman scattering.

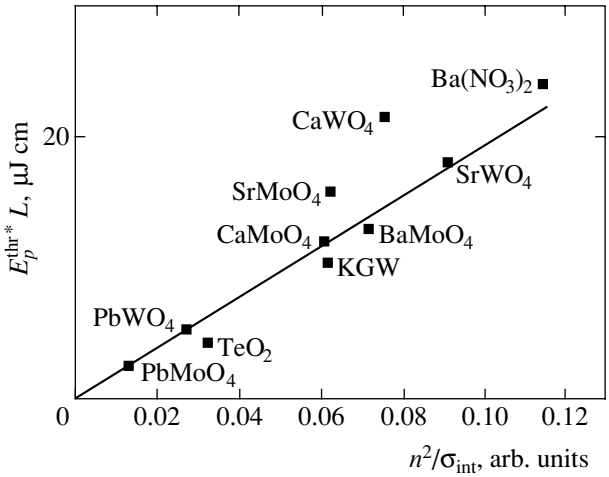


Fig. 5. Experimental normalized threshold energy of SRS pumping as a function of the reciprocal integral cross section of spontaneous Raman scattering n^2/σ_{int}. The solid line qualitatively describes dependence (2).

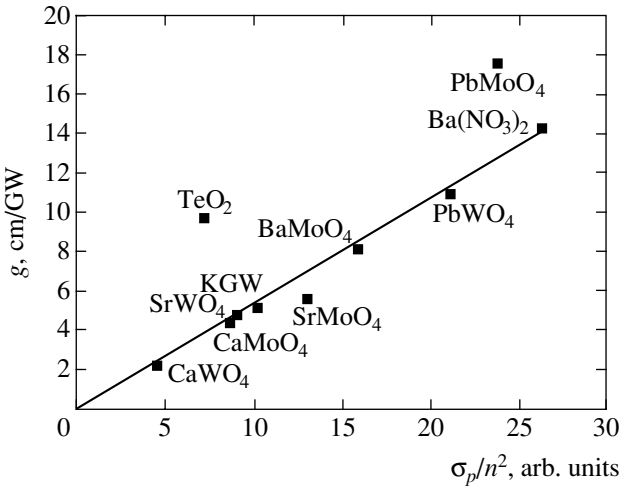


Fig. 6. Steady-state SRS gain coefficient g for all crystals as a function of the peak cross section of spontaneous Raman scattering σ_p/n^2 normalized by the refractive index n according to (3).

tal with the smallest $\tau_p/T_2 = 0.42$ ratio (sharply transient SRS). A decrease in the difference between the steady-state and transient gain increments as the τ_p/T_2 ratio increases to values of about 4–9 is indicative of the approach to a steady-state SRS mode in the $PbMoO_4$, $PbWO_4$, and TeO_2 crystals. The G_0 increments found this way can be used to calculate SRS gain factors g by the formula $G_0 = gI_pL$, the nonlinear interaction length L and the pumping intensity I_p being known.

The nonlinear interaction length is as a rule determined as twice the length along which the beam cross section increases twofold because of diffraction. For the Gaussian beams that we are dealing with, this length is given by the equation

$$L = \frac{4\pi w_0^2 n}{\lambda}, \quad (4)$$

where $2w_0$ is the diameter at the beam waist at $I = I_0/e$ and n is the refractive index. The laser beam profile measured using the CCD chamber was close to Gaussian with a diameter of $2w_0 = 80$ μm. The length L found for $n = 2$ was approximately 5 cm. As follows from Table 2, the lengths of our crystals are smaller than $L = 5$ cm and therefore determine the nonlinear interaction length. The pumping intensity I_p is given by

$$I_p = \frac{E_p}{S_{\text{eff}}\tau_p}. \quad (5)$$

Here, E_p and τ_p are the measured pulse energy and width and S_{eff} is the effective area of the beam cross section determined by the overlap integral of pumping and Stokes wave modes; for Gaussian beams, $S_{\text{eff}} = 2\pi w_0^2$ [4]. Note that, in our calculations for near-threshold conditions, we ignore possible transformation of the spatial Stokes beam profile caused by SRS amplification or self-focusing.

The $E_p^{\text{thr}}L$, G_0, and $\tau_p = 11$ ps values, measured and listed in Table 2, were used to calculate the SRS gain coefficients g for all the crystals. These values are shown in Fig. 6 as depending on the peak cross section of spontaneous Raman scattering normalized by refractive indexes according to (3). On the whole, most of the data well fit the linear dependence of gain coefficient g of steady-state SRS on the peak cross section of spontaneous Raman scattering, which is evidence of the correctness of measurements and their correspondence to (3). The data on the $PbMoO_4$ and TeO_2 crystals, which are characterized by large integral cross sections of spontaneous Raman scattering and the shortest dephasing times T_2, fall outside the linear dependence shown in Fig. 6. As has been mentioned above, possible reasons for this may be related to inaccuracy of determining T_2 from the spontaneous Raman scattering spectra. In contrast to the other crystals, TeO_2 is characterized by two-photon absorption even under low-intensity pumping. This effect manifests itself by visible glow of the laser-excited channel and can distort the results of SRS gain coefficient measurements.

The values obtained in this work can be compared with the literature estimates of SRS gain coefficients for steady-state SRS excited in crystals by nanosecond laser pulses at 1060 nm. For instance, $g = 11$ cm/GW was obtained for the $Ba(NO_3)_2$ crystal in [19]. For the KGW crystal, the g value varies from 6 [20] to 4 [21] cm/GW, which is close to our data. The value obtained for $SrWO_4$ in [22], $g = 5$ cm/GW, virtually coincides with that given in Table 2. On the whole, the g values estimated by us as described above are in satisfactory agreement with the results obtained by independent methods.

3.3. The Time and Spectral Properties of SRS Pulses

We measured SRS pulse widths and spectra for two "extreme" crystals with the longest, $T_2 = 26.5$ ps, and shortest, $T_2 = 1.77$ ps, optical dephasing times, that is, for $Ba(NO_3)_2$ and $PbMoO_4$. The Stokes pulse width for $Ba(NO_3)_2$ decreased relative to τ_p in satisfactory agreement with theory [2] from 11 to 7 ps, the spectrum width being $\Delta\nu_S = 3$ cm^{-1}. As has been mentioned above, the properties of the $PbMoO_4$ crystal are somewhat anomalous, which also follows from the time and spectral parameters of SRS. For this crystal, the width of Stokes pulses $\tau_s = 10.5$ ps at the fundamental and second harmonic frequencies is close to the pumping pulse width, and the spectrum width is $\Delta\nu_S = 3.76$ cm^{-1}. At the same time, according to [2], the Stokes pulse for $PbMoO_4$ should be more than twofold narrower (for a Gaussian envelope). The product $\tau_s\Delta\nu_S = 1.18$ for $PbMoO_4$ corresponds to a substantial deviation from a

spectrally limited pulse, whereas this product for $Ba(NO_3)_2$ is 0.63, which is much closer to a spectrally limited pulse. Note that the self-phase modulation-induced insignificant chirp of the pumping pulse frequency of our laser had to be transferred from the pumping frequency to the SRS Stokes wave frequency [12, 13] and cause phase modulation and spectral broadening of Stokes radiation. The reasons for the discrepancy between theory and experiment for $PbMoO_4$ can be related to the complex spectrum of vibrational resonances in the medium.

As has been shown in [4], in an optical fiber, that is, in a medium with inhomogeneously broadened vibrational resonances (fused quartz), the SRS pulse under picosecond pumping conditions slightly narrows, as in the $PbMoO_4$ crystal considered above, and has a femtosecond time structure caused by the large width of the spontaneous Raman scattering spectrum. The $PbMoO_4$ crystal may have similar characteristics, and the Stokes pulse can have a substructure of width about 2 ps (irresolvable under our experimental conditions) determined by its spectrum. This substructure may prevent it from narrowing.

4. CONCLUSIONS

We comparatively studied SRS of a large number of oxide crystals, for the first time for some of them, under picosecond pumping conditions. Measurements of the energy, spectral, and time SRS parameters and their comparison with the spectroscopic parameters of spontaneous Raman scattering determined by us substantiated the laws governing transient SRS. Measurements of the threshold SRS pumping energies for a large number of crystals showed that the gain increment of transient SRS was independent of the dephasing time or the spectral width of Raman-active vibrations but was determined by the integral cross section of Raman scattering. The theory of transient SRS developed in [2] was used to show that the theory on the whole correctly described our data on transient SRS in oxide crystals and the relation between the transient and steady-state SRS gain increments. A comparatively simple experimental technique for picosecond excitation of SRS and a comparison of the measured SRS parameters with theoretical dependences allowed us to estimate the steady-state SRS gain coefficients in crystals and show that they were related to the peak cross sections of spontaneous Raman scattering determined by linear spectroscopy methods.

ACKNOWLEDGMENTS

This work was financially supported by the Russian Foundation for Basic Research (project nos. 04-02-17004a and 03-02-17309a) and MNTTs (grant EOARD no. 2022P).

REFERENCES

1. *Ultrashort Light Pulses*, Ed. by S. Shapiro (Springer, Heidelberg, 1977; Mir, Moscow, 1980).
2. R. L. Carman, F. Shimizu, N. Bloembergen, *et al.*, Phys. Rev. A **2**, 60 (1970); S. A. Akhmanov, K. N. Drobovich, A. P. Sukhorukov, *et al.*, Zh. Éksp. Teor. Fiz. **59**, 485 (1970) [Sov. Phys. JETP **32**, 266 (1971)].
3. R. L. Carman and M. E. Mack, Phys. Rev. A **5**, 341 (1972).
4. E. M. Dianov, L. M. Ivanov, A. Ya. Karasik, *et al.*, Zh. Éksp. Teor. Fiz. **91**, 2031 (1986) [Sov. Phys. JETP **64**, 1205 (1986)].
5. K. Andryunas, Yu. Vishchakas, V. Kabelka, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **42**, 333 (1985) [JETP Lett. **42**, 410 (1985)].
6. T. T. Basiev, Usp. Fiz. Nauk **169**, 1149 (1999) [Phys. Usp. **42**, 1051 (1999)].
7. A. A. Kaminskii, S. N. Bagaev, K. Ueda, *et al.*, Laser Phys. **11**, 1142 (2001).
8. P. Cerny, P. G. Zverev, H. Jelinkova, *et al.*, Opt. Commun. **177**, 397 (2000).
9. P. G. Zverev, T. T. Basiev, A. A. Sobol', *et al.*, Kvantovaya Élektron. (Moscow) **30**, 55 (2000).
10. P. G. Zverev, A. Ya. Karasik, T. T. Basiev, *et al.*, Kvantovaya Élektron. (Moscow) **33**, 331 (2003).
11. V. N. Lugovoĭ, Zh. Éksp. Teor. Fiz. **71**, 1307 (1976) [Sov. Phys. JETP **44**, 683 (1976)].
12. E. M. Dianov, A. Ya. Karasik, P. V. Mamyshev, *et al.*, Zh. Éksp. Teor. Fiz. **89**, 781 (1985) [Sov. Phys. JETP **62**, 448 (1985)].
13. E. M. Dianov, A. Ya. Karasik, P. V. Mamyshev, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **41**, 242 (1985) [JETP Lett. **41**, 294 (1985)].
14. A. B. Grudinin, E. M. Dianov, D. V. Korobkin, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **45**, 211 (1987) [JETP Lett. **45**, 260 (1987)].
15. G. P. Agrawal, *Nonlinear Fiber Optics* (Academic, Boston, 1994; Mir, Moscow, 1996).
16. T. T. Basiev, A. A. Sobol, P. G. Zverev, *et al.*, Appl. Opt. **38**, 594 (1999).
17. T. T. Basiev, A. A. Sobol, Yu. K. Voronko, *et al.*, Opt. Mater. **15**, 205 (2000).
18. T. T. Basiev, N. A. Es'kov, A. Ya. Karasik, *et al.*, Opt. Lett. **17**, 201 (1992).
19. T. T. Basiev, P. G. Zverev, V. N. Voĭtsekhovskiĭ, *et al.*, Kvantovaya Élektron. (Moscow) **14**, 2452 (1987).
20. I. V. Mochalov, Opt. Zh. **62** (11), 4 (1995) [J. Opt. Technol. **62**, 746 (1995)].
21. T. T. Basiev, V. V. Osiko, A. M. Prokhorov, *et al.*, in *Solid-State Mid-Infrared Laser Sources*, Ed. by I. T. Sorokina and K. L. Vodopyanov (Springer, Berlin, 2003); Top. Appl. Phys. **89**, 351 (2003).
22. P. G. Zverev, T. T. Basiev, L. I. Ivleva, *et al.*, in *Advanced Solid-State Lasers*, Ed. by M. E. Fermann and L. R. Marshall (Optical Society of America, Washington, DC, 2002), OSA Trends in Optics and Photonics, Vol. 68, p. 70.

Translated by V. Sipachev

Journal of Experimental and Theoretical Physics, Vol. 99, No. 5, 2004, pp. 942–946.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 5, 2004, pp. 1083–1088.

ATOMS, SPECTRA, RADIATION

Optical Rectification Effect in Nanostructured Carbon Films

G. M. Mikheev[a,*], R. G. Zonov[a], A. N. Obraztsov[b], and Yu. P. Svirko[c]

[a] *Institute of Applied Mechanics, Ural Division, Russian Academy of Sciences, Izhevsk, Udmurtia, 426067 Russia*

[b] *Department of Physics, Moscow State University, Moscow, 119992 Russia*

[c] *Department of Physics, Joensuu University, Joensuu, Finland*

**e-mail: gmmikheev@udmnet.ru*

Received March 19, 2004

Abstract—Electrically conducting nanostructured carbon films obtained by chemical vapor deposition and composed of nanodimensional graphite crystals exhibit the effect of optical rectification on exposure to nanosecond pulsed laser radiation. Experiments show that the amplitude and polarity of the pulsed voltage strongly depend on the angle of incidence and polarization of the laser radiation and on the spatial orientation of a carbon film with electrodes relative to the laser beam. Under the optimum conditions corresponding to maximum amplitude of the response signal, the factor of conversion of the laser pulse power into electric voltage was about 500 and 650 mV/MW at a laser wavelength of 1064 and 532 nm, respectively. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The phenomenon of optical rectification originally reported by Bass *et al.* [1] offers an example of nonlinear interaction between high-power laser radiation and an optical medium. The effect is manifested when a sufficiently powerful laser pulse passes through a nonlinear optical crystal and consists in that a pulse of electric polarization is induced in the crystal, the shape of this pulse repeating the laser pulse envelope. If electrodes are applied to opposite faces of such a nonlinear optical crystal, the laser-induced polarization will lead to the appearance of a potential difference between these electrodes [2, 3]. In noncentrosymmetric media, the optical rectification effect is related to the second term in the expansion of the electric polarization vector into series with respect to the electric field strength—that is, to the second-order nonlinear susceptibility, which is also responsible for the second harmonic generation. Although both phenomena were discovered almost simultaneously and experimentally observed in a large number of noncentrosymmetric crystals (see, e.g., review [3]), the optical rectification effect has received much less attention than the second harmonic generation. The situation changed with the appearance of lasers capable of generating femtosecond pulses, which allowed optical rectification to be used for the generation of terahertz radiation in ZnTe and some other noncentrosymmetric semiconductor crystals [4]. Quite recently, the generation of terahertz radiation was observed in epitaxial films of centrosymmetric YBCO crystals, where the optical rectification effect is due to a quadrupole (rather than dipole) mechanism of optical nonlinearity [5]. Since the quadrupole nonlinearity increases with the dimensions of molecules, we may expect that the optical rectification effect related to this mechanism can be observed in carbon nanotubes and other nanostructured materials (as a rule, possessing the center of inversion). The existing theoretical descriptions of the second harmonic generation and optical rectification in carbon nanotubes refer to the effects observed in the presence of a constant electric field [6].

This paper presents the results of our experimental investigation of the optical rectification effect in carbon films composed of nanodimensional graphite crystals.

2. EXPERIMENTAL

The experiments were performed with carbon films obtained by chemical vapor deposition (CVD) from a methane–hydrogen mixture according to our standard technology described elsewhere [7, 8]. The films were deposited onto 0.5-mm-thick silicon substrates with linear dimensions 25×25 mm. The main structural elements in these carbon films are crystallites of an irregular shape comprising several (typically, five to fifty) parallel, well-ordered graphite layers. The thickness of these crystallites ranges from 2 to 20 nm, while their lateral dimensions vary within 1–3 μm. In all crystallites, the atomic layers exhibit predominant orientation in the direction of normal to the substrate surface, with a maximum deviation not exceeding ±20° [9]. The distance between crystallites is about 0.5–1 μm. The carbon films possessed clearly pronounced porous nanocrystalline structure. The average thickness of nanostructured carbon films studied was 3–4 μm.

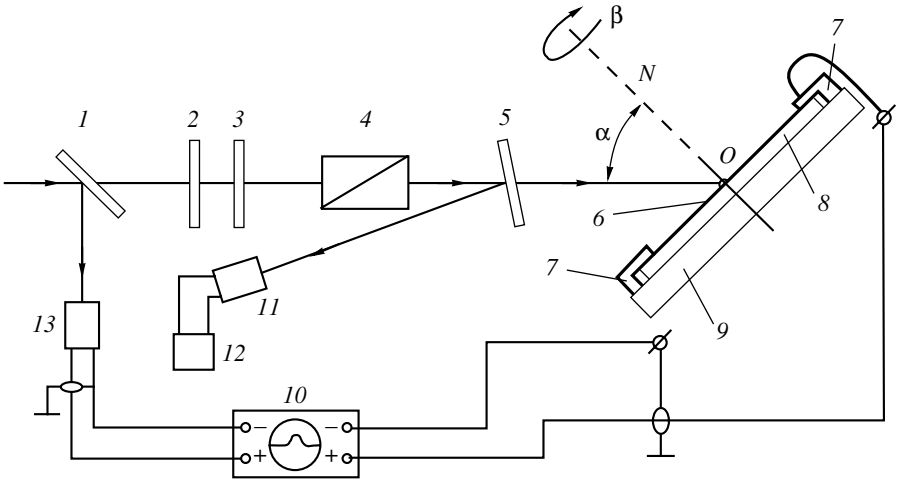


Fig. 1. Schematic diagram of the experimental arrangement (see the text for explanations).

Figure 1 shows a schematic diagram of the experimental arrangement. Nanostructured carbon film *6* on a silicon substrate *8* was pressed with two flat conducting electrodes *7* to a dielectric sample holder *9*, so that the substrate was in contact with the sample holder, while the film surface was almost entirely free (parallel electrodes covered only a small part of the film surface area). The dc resistance between electrodes was 100–200 Ω, depending on the distance between them.

The films were exposed to the radiation of a single-mode single-frequency YAG:Nd^{3+} laser. The laser cavity scheme allowed linearly polarized output radiation to be obtained [10]. The energy *W* of the linearly polarized radiation pulse at a wavelength of λ = 1064 nm reached 50 mJ. The pulses had a Gaussian shape with a full width at half maximum (FWHM) of about 22 ns. The laser beam diameter was 2 mm. The laser beam was attenuated by neutral filters and passed through λ/4 plates (*2*, *3*) and optical polarizer *4* (Fig. 1). By changing mutual orientation of the quarter-wave plates, it was possible to rotate the polarization plane in the beam entering the polarizer. This allowed the power of a linearly polarized radiation at the polarizer output to be smoothly controlled. The polarizer could be rotated within 90° around its axis so as to obtain *p*- and *s*-polarized output beams.

Then, the laser beam passed divider plate *5* and was directed to the carbon film studied (Fig. 1). Using the divider plate, together with photodiode *11* and an automated multichannel registration system *12* [11], it was possible to measure the energy of pulsed radiation incident onto the sample. Divider plate *1*, together with avalanche photodiode *13*, enabled the observation of laser radiation pulses with the aid of an oscillograph. The measurements according to this scheme were also performed using the second harmonic mode of the same laser. All the results presented below were obtained for laser pulse energies below a threshold for the visible laser damage of the nanostructures carbon films studied. The damage could be observed at a laser radiation intensity above 20 MW/cm^2 [12].

The amplitude of the voltage *U* appearing between the electrodes during laser irradiation of the film was measured using a storage oscillograph *10* with an input impedance of 1 MΩ, while the shape of electric pulses could be observed using an oscillograph with a pass band of 650 MHz.

3. EXPERIMENTAL RESULTS

The results of our experiments showed that irradiation of a nanostructured carbon film by laser pulses at λ = 1064 nm leads to the appearance of pulsed voltage between the electrodes. The shape of this signal was virtually identical to the laser pulse shape observed with the aid of the avalanche photodiode. Analogous results were obtained for the samples irradiated by second-harmonic radiation (λ = 532 nm). As is known, the conversion to second harmonic leads to a decrease in the laser pulse width by a factor of $\sqrt{2}$. In our experiments, the durations of pulses measured between electrodes during exposure of the sample to laser radiation with λ = 1064 and 532 nm also differed approximately 1.4 times.

Further investigations showed that the amplitude *U* and polarity of the pulsed response voltage strongly depend on the polarization of laser beam and on the spatial orientation of a carbon film with electrodes relative to the incident laser beam. When the laser beam was perpendicular to the film plane, the response signal amplitude was zero irrespective of the beam polarization and the arrangement of electrodes. Any other orientation of the film relative to the laser beam led to the appearance of a nonzero pulsed response voltage.

Figure 2 shows the dependence of the response voltage pulse amplitude *U* on the laser beam incidence

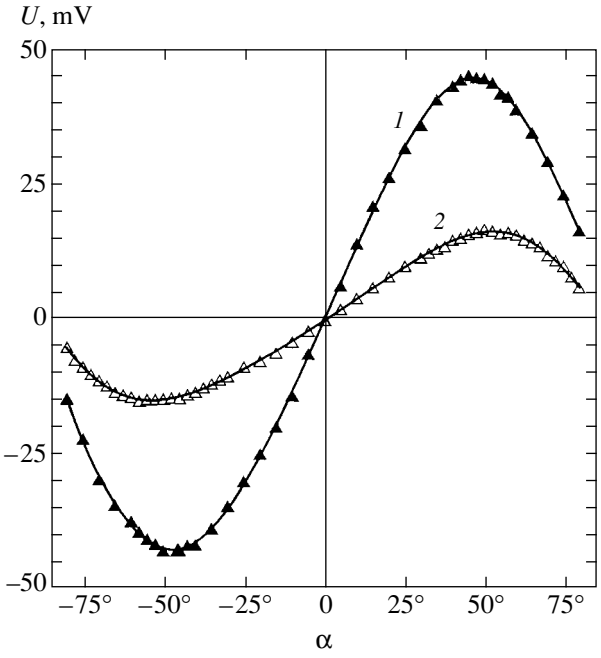


Fig. 2. Plots of the amplitude U of a pulsed response voltage between electrodes fixed on a nanostructured graphitelike carbon film versus the angle α of incidence of (*1*) *p*- and (*2*) *s*-polarized laser radiation (τ = 22 ns, λ = 1064 nm, W = 2 mJ).

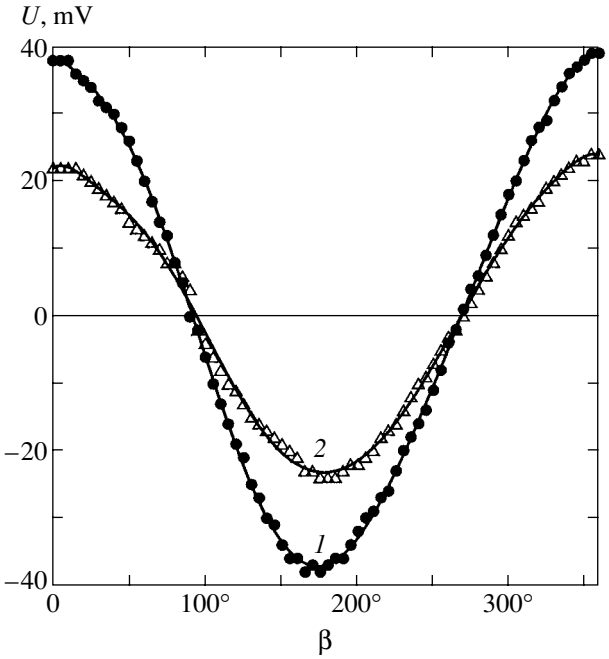


Fig. 3. Plots of the amplitude U of a pulsed response voltage between electrodes fixed on a nanostructured graphitelike carbon film versus the angle β of the sample rotation for (*1*) linear and (*2*) circular polarization of the incident laser radiation (τ = 22 ns, λ = 1064 nm, W = 1.7 mJ, α = 50°).

angle α relative to the film plane (Fig. 1) for the *p*- and *s*-polarized radiation (at a fixed laser pulse energy). Note that, according to Fig. 1, $\alpha = 0$ when the film surface is perpendicular to the laser beam (initial orientation) and α is positive (negative) when the film is rotated clockwise (anticlockwise) from the initial position. The axis of rotation passing through the point *O*, the plane of the carbon film, and the electrodes are perpendicular to the plane of the figure coinciding with the plane of incidence of the laser radiation (and the plane of observation). It should be also noted that data corresponding to incidence angles close to ±90° are not presented in Fig. 2 because laser radiation incident at grazing angles on the sample surface partly illuminated the electrodes, with the result that the response voltage pulse duration became significantly longer than the laser pulse width. Variation of the laser spot position in the region between electrodes at a fixed arrangement of the sample did not influence the amplitude and shape of the response signal.

As can be seen from Fig. 2, the experimental curves $U(\alpha)$ behave as uneven functions of the angle, the absolute value of U is maximum at $\alpha = \pm(45°–55°)$, and the $U(\alpha)$ curves for the *p*- and *s*-polarized laser radiation are significantly different in magnitude of the response voltage. At fixed laser beam energy, the response voltage pulse amplitude for the *p* polarization exceeds that for the *s* polarization in the entire range of α.

Figure 3 shows the plots of U versus β, the angle of rotation of film *6* with electrodes *7* relative to the *ON* axis (Fig. 1) coinciding with the normal to the surface, measured for the linear (*p*-polarized) and circularly polarized laser beam at a fixed incidence angle α = +50°. It should be noted that the values $\beta = 0°$ and 180° correspond to the positions of sample faces with electrodes *7* (perpendicular to the plane of Fig. 1). As can be seen from Fig. 3, rotation of the sample holder plate with electrodes around the *ON* axis leads to variation of the response voltage pulse amplitude according to the cosine law. When electrodes *7* are in the planes parallel to the plane of laser beam incidence (β = 90° and 270°) the response voltage pulse amplitude is zero. For β = 180°, when the electrodes change their mutual arrangement (upper electrode becomes lower and vice versa, see Fig. 1), the response signal becomes negative, while its amplitude approximately corresponds to the U value observed for $\beta = 0°$.

Figure 4 shows the plots of U versus the laser pulse energy W measured using the *p*-polarized radiation for two orientations of the film with electrodes corresponding to α = +50°, $\beta = 0°$ and α = –50°, $\beta = 0°$. As can be seen from these data, the dependences $U(W)$ can be approximated by straight lines with a good accuracy. Note that the plots corresponding to α = +50°, $\beta = 0°$ and α = –50°, $\beta = 0°$ are virtually symmetric with respect to the W axis. Taking into account that the laser pulse duration τ during the experiments at λ = 1064 nm was fixed, we may infer from Fig. 4 that the amplitude of the pulsed response voltage is a linear function of the laser power. The coefficient of conversion $\eta(\lambda$ =

1064 nm) of the pulsed laser power into response voltage amplitude was about 500 mV/MW, while the value of $\eta(\lambda = 532$ nm) obtained for second-harmonic radiation was about 650 mV/MW.

We have also studied the dependence of the response voltage pulse amplitude on the laser beam spot diameter on the carbon film surface at fixed values of W and τ. The laser beam diameter was increased or decreased with the aid of telescopes based on diverging and converging lenses, respectively. These experiments showed that an increase in the beam diameter by a factor of 6.25 led to a growth in the response voltage amplitude by a factor of about 1.3. Thus, at a fixed laser pulse power, a 39-fold decrease in the incident power density led to a slight (1.3-fold) increase, rather than to a decrease in amplitude of the response voltage between electrodes situated on the carbon film surface.

4. DISCUSSION OF RESULTS

The whole body of experimental results described above covers all features characteristic of the optical rectification effect (see, e.g., [3]). The obtained coefficients of conversion of the pulsed laser power into response voltage amplitude are many times greater than the values (observed for the optical rectification in nonlinear optical crystals of ADP and DKDK types (15 and 18 mV/MW, respectively) [3].

However, in contrast to the previously reported investigations, where the response voltage was measured between electrodes situated on the opposite faces of noncentrosymmetric transparent dielectric crystals, the optical rectification effect in our experiments was observed on absorbing conducting nanostructured carbon films with a graphitelike atomic structure. Since graphite (two- or three-dimensional) structure possesses the center of inversion (point symmetry group D_{6h} [13]), the affect observed in our samples is probably due to a quadrupole nonlinearity. This conclusion is confirmed by the observed dependence of the response voltage on the film orientation. Indeed, preliminary theoretical analysis showed that the signal due to the optical rectification on a quadrupole nonlinearity is proportional to $F(\alpha)\sin 2\alpha\cos\beta$, where $F(\alpha)$ is a slowly varying function of the angle α. This agrees well with our results presented in Figs. 2 and 3.

However, the symmetry of the graphitelike material in our nanostructured carbon films may significantly differ from that of the usual graphite because of a considerable amount of structural defects related to the high proportion of atoms situated in the surface of nanodimensional crystals and in the region of bending of the atomic planes [14]. Violation of the central symmetry caused by such structural defects may lead to an increase in the contribution of dipole nonlinearity to the static polarization. Such an analysis of the possible manifestations of the dipole, quadrupole, and other possible mechanisms is a subject for separate investigation.

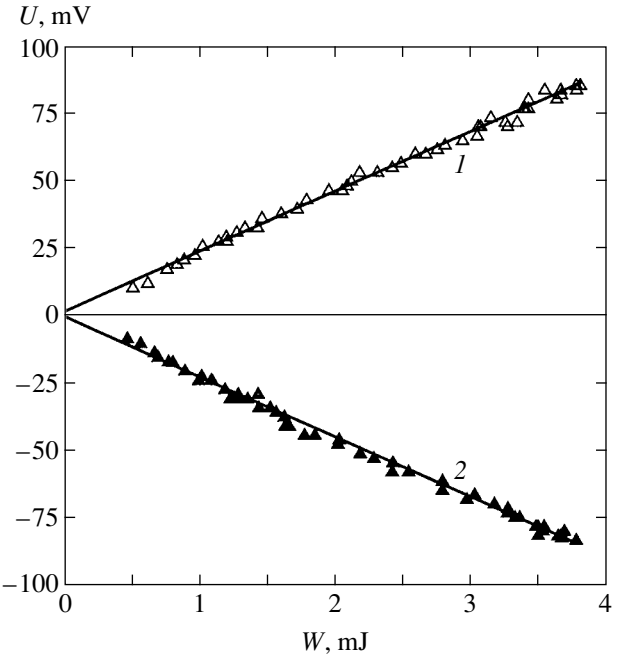


Fig. 4. Plots of the amplitude U of a pulsed response voltage between electrodes fixed on a nanostructured graphitelike carbon film versus the laser pulse energy W for (*1*) $\alpha = 50°$, $\beta = 0°$ and (*2*) $\alpha = -50°$, $\beta = 0°$ (*p*-polarized light).

5. CONCLUSIONS

We have experimentally observed the appearance of a pulsed electric signal in laser-irradiated conducting nanostructured carbon films composed of nanodimensional graphite crystals. The films were obtained by CVD on silicon substrates. The time variation of the pulsed response voltage measured between parallel electrodes on the surface of the film repeated the shape of the laser power pulse. The amplitude of the response voltage strongly depends on spatial orientation of a carbon film with electrodes relative to the laser beam and on the polarization of the incident laser radiation. The magnitude of the effect is directly proportional to the laser power, rather than to the laser power density as in the case of second harmonic generation.

ACKNOWLEDGMENTS

The authors are grateful to T.N. Mogileva for her help in performing experiments.

This study was supported by the INTAS Foundation (grant no. 01-0254).

REFERENCES

1. M. Bass, P. A. Franken, J. F. Ward, and G. Weinereich, Phys. Rev. Lett. **9**, 446 (1962).
2. J. F. Ward, Phys. Rev. **143**, 569 (1966).
3. B. N. Morozov and Yu. M. Aĭvazyan, Kvantovaya Élektron. (Moscow) **7**, 5 (1980).
4. M. C. Nuss and J. Orenstein, in *Millimeter and Submillimeter Wave Spectroscopy of Solids* (Springer, Berlin, 1998), p. 7.
5. J. L. W. Siders, S. A. Trugman, F. H. Garzon, *et al.*, Phys. Rev. B **61**, 13633 (2000).
6. Vl. Margulis, E. A. Gaiduk, and E. N. Zhidkin, Opt. Commun. **183**, 317 (2000).
7. A. A. Zolotukhin, A. N. Obraztsov, A. O. Ustinov, and A. P. Volkov, Zh. Éksp. Teor. Fiz. **124**, 1291 (2003) [JETP **97**, 1154 (2003)].
8. A. N. Obraztsov, A. A. Zolotukhin, A. O. Ustinov, *et al.*, Carbon **41**, 836 (2003).
9. A. N. Obraztsov, I. Yu. Pavlovsky, A. P. Volkov, *et al.*, Diamond Relat. Mater. **8**, 814 (1999).
10. G. M. Mikheev, D. I. Maleev, and T. N. Mogileva, Kvantovaya Élektron. (Moscow) **19**, 45 (1992).
11. G. M. Mikheev, T. N. Mogileva, A. Yu. Popov, and D. G. Kalyuzhnyĭ, Prib. Tekh. Éksp., No. 2, 101 (2003) [Instrum. Exp. Tech. **46**, 233 (2003)].
12. G. M. Mikheev, R. G. Zonov, A. N. Obraztsov, and A. P. Volkov, Zh. Éksp. Teor. Fiz. **125**, 548 (2004) [JETP **98**, 483 (2004)].
13. *Optical Properties of Semiconductors (Handbook)*, Ed. by V. I. Gavrilenko, A. M. Grekhov, D. V. Korbutyak, and V. G. Litovchenko (Naukova Dumka, Kiev, 1987), p. 69 [in Russian].
14. A. N. Obraztsov, A. P. Volkov, I. Yu. Pavlovskiĭ, *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **69**, 381 (1999) [JETP Lett. **69**, 411 (1999)].

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 99, No. 5, 2004, pp. 947–957.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 5, 2004, pp. 1089–1100.

ATOMS, SPECTRA, RADIATION

Nonclassical Light Generation in Quasi-Phase-Matched Parametric Self-Frequency Conversion

A. A. Novikov and A. S. Chirkin*

Department of Physics, Moscow State University, Moscow, 119992 Russia
**e-mail: chirkin@squeez.phys.msu.su, aschirkin@pisem.net*

Received April 30, 2004

Abstract—We present a quantum theory of the parametric self-conversion of the laser radiation frequency in active nonlinear crystals with a regular domain structure. Such crystals feature simultaneous lasing and quasi-phase-matched parametric conversion of the laser radiation frequency. These processes are described using the Heisenberg–Langevin equations in two regimes of the subharmonic generation: super- and subthreshold. The spectral properties of the quadrature components of the laser frequency and its subharmonic and the photon statistics have been studied as dependent on the pump power, crystal length, and reflectance of the laser cavity output mirror. Using the obtained analytical expressions, these characteristics are calculated for a active nonlinear Nd:Mg:$LiNbO_3$ crystal with a regular domain structure. In the subthreshold regime, the maximum decrease in the spectral density of fluctuations in the subharmonic quadrature component relative to the standard quantum limit may reach 90%; in the above-threshold regime, these fluctuations are virtually not suppressed. A decrease in the spectral density of fluctuations of the laser frequency quadrature does not exceed 10%. In the subthreshold excitation regime, the subharmonic photons obey a super-Poisson statistics; in the above-threshold regime, the photon statistics is Poisson-like. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

In recent years, much attention has been devoted to ferroelectric crystals possessing periodically inhomogeneous nonlinear properties. In such media, the direction of the spontaneous polarization vector changes to the opposite on passage from one domain to another. This may lead to reversal of the sign of the coefficient of quadratic nonlinear susceptibility [1, 2], whereby a "nonlinear" lattice is formed in the crystal possessing homogeneous linear properties. Nonlinear optical crystals with a regular domain structure, called RDS crystals or nonlinear photonic (periodically poled) crystals, have certain advantages over homogeneous nonlinear crystals. The main advantage is that, by selecting the period of modulation of the nonlinear susceptibility in an RDS crystal, it is possible to compensate mismatch of the interacting light waves (quasi-phase-matched interaction), thus providing conditions for almost arbitrary three-wave mixing. At present, the quasi-phase-matched interactions between light waves are used for generating coherent radiation in a broad spectral range from UV to IR. In addition, the quasi-phase-matched interactions are of interest from the standpoint of generating nonclassical light (squeezed light and entangled photon states) [3–6]. Nonclassical light can be used in various high-precision optical measurements and in optical data transmission and processing systems [7].

Active nonlinear RDS crystals open new prospects in nonlinear optics [8]. In such crystals, rare earth dopant ions provide for active (lasing) properties, while the crystalline matrix plays the role of a nonlinear medium. This system may feature the phenomenon of self-frequency conversion, whereby lasing proceeds simultaneously with nonlinear conversion of the laser frequency. Investigations into the properties of active nonlinear laser crystals are stimulated by the wide potential range of applications of compact and reliable lasers based on such crystals, generating in the visible and IR spectral range.

The possibilities of self-frequency conversion in homogeneous active nonlinear crystals are limited by their dispersion properties. As was noted above, these limitations can be by-passed in the presence of spatially modulated nonlinear susceptibility. Recently [8–11], it was demonstrated that RDS crystals combining the advantages of active nonlinear media with periodically inhomogeneous nonlinear properties are of interest from the standpoint of realization of various three-wave interactions, whereby one of the waves can be enhanced due to the active properties of the crystal. Such processes provide a basis for the creation of miniature self-frequency-conversion lasers.

This study was devoted to the quantum properties of light generated during the parametric self-frequency conversion of laser radiation. In the course of this process taking place in a resonator based on a active nonlinear RDS crystal, lasing proceeds simultaneously with the parametric conversion of the laser frequency.

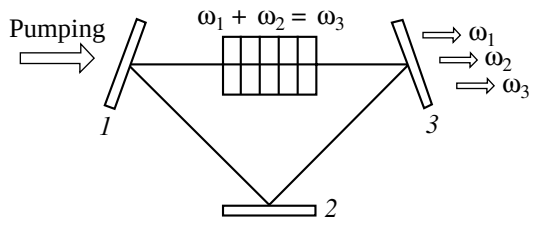


Fig. 1. Schematic diagram of ring cavity involving three mirrors (*1–3*) and an active nonlinear RDS crystal featuring three-wave interaction.

In the case under consideration, we are speaking of the second subharmonic generation. Up to now, quantum theory has been well developed separately for the lasing and the parametric frequency conversion processes (see, e.g., monographs [12–14]). However, in the case of self-frequency conversion, the two processes are correlated. This circumstance alters the physics of this phenomenon and complicates its theoretical analysis.

Previously [15, 16], the quantum theory of parametric self-frequency conversion was developed for gas lasers. In these lasers, the photon lifetime T in the laser cavity is much longer than the characteristic times of the inverse population relaxation ($T_\parallel$) and the active medium polarization ($T_\perp$): $T \gg T_\perp, T_\parallel$. This condition allows the inverse population and polarization to be adiabatically excluded from the equations describing generation of the laser radiation. However, the aforementioned active nonlinear crystals are characterized by a different relation between the characteristic times: $T \ll T_\perp \ll T_\parallel$. For this reason, the results of previous theoretical analysis [15, 16] are inapplicable to such solid-state laser systems.

In this study, a quantum analysis of the parametric self-frequency conversion is performed for an arbitrary relation between the characteristic times. We will consider the spectrum of fluctuations of the quadrature components and the photon statistics of the laser radiation and subharmonic fields. The general formulas will be used in numerical calculations of the parametric self-frequency conversion in a active nonlinear Nd:Mg:$LiNbO_3$ crystal with a regular domain structure.

This paper is organized as follows. In Section 2, the Heisenberg–Langevin equations are written separately for the laser generation and the quasi-phase-matched three-wave interactions. The parametric self-frequency conversion is analyzed in Section 3, where the corresponding Heisenberg–Langevin equations are written and solved by perturbation method. Section 4 presents the results of calculating the spectrum of fluctuations of the quadrature components of the laser radiation and its subharmonic in the sub- and above-threshold regimes of second subharmonic generation in a active nonlinear Nd:Mg:$LiNbO_3$ crystal. In Section 5, the main attention is devoted to analysis of the subharmonic photon statistics in the subthreshold generation regime.

2. LASER AND NONLINEAR OPTICAL EQUATIONS

The process of parametric self-frequency conversion will be analyzed according to the following scheme. First, we will separately describe laser generation in an active crystal and three-wave interactions in a nonlinear optical crystal. Then, the two descriptions will be combined to yield the joint system of equations for the field operators and atoms of the medium, which describes processes in a nonlinear active crystal. The system will be solved using the Heisenberg–Langevin method [12–14], which is known to be especially convenient for calculating the correlation functions of operators and, hence, of the spectral densities.

The process of laser and parametric generation is studied for a nonlinear active RDS crystal situated in a ring cavity (Fig. 1) provided with two mirrors (*1* and *2*) totally reflecting radiation at the excited frequencies and with an output mirror (*3*). The crystal is pumped through mirror *1*; mirrors *2* and *3* are fully transparent to pump radiation.

2.1. Laser Generation in an Active Crystal

The process of laser generation in an active crystal will be first considered without allowance for nonlinear optical properties. We use a quantum approach based on the Heisenberg operator equation, generally following the scheme [12]. Excluding operators related to the thermal reservoir and modeling the laser field interaction with other physical systems (except atoms) from the system of operator equations for the field and atoms of the medium, we obtain the system of Heisenberg–Langevin equations [12]:

$$\frac{da_l}{dt} = -k_l a_l - iT_C g\sigma + F_l(t), \quad (1)$$

$$\frac{d\sigma}{dt} = \frac{1}{\tau_\perp}(-\sigma + igT_\perp a_l N) + \Gamma(t), \quad (2)$$

$$\frac{d\sigma^\dagger}{dt} = \frac{1}{\tau_\perp}(-\sigma^\dagger) - igT_\perp a_l^\dagger N) + \Gamma^\dagger(t), \quad (3)$$

$$\frac{dN}{dt} = \frac{1}{\tau_\parallel}[P - N + i2T_\parallel g(\sigma a_l^\dagger - \sigma^\dagger a_l)] + \Gamma_N(t). \quad (4)$$

These equations are written in the interaction representation using the rotating wave approximation. The laser field frequency ω_l is assumed to coincide with the

frequency ω_0 of transitions between the laser levels ($\omega_0 = \omega_l$). In Eqs. (1)–(4),

$$\sigma = \sum_{\nu=1}^{M} \sigma_\nu, \quad N = \sum_{\nu=1}^{M} N_\nu$$

are operators of the polarization and the population difference for a laser- active medium containing M atoms, respectively; $a_l^\dagger$ (a_l) is the operator of creation (annihilation) of a photon with the frequency ω_l;

$$g = \sqrt{\frac{c\hbar\omega_l}{4I_S V T_\parallel T_\perp}}$$

is the constant of interaction between an atom and the electromagnetic field; $T_\parallel$ and $T_\perp$ are the times of relaxation of the inverse population and the polarization of medium, respectively; I_S is the saturation intensity of the active medium; V is the quantization volume determined by the transverse size of the pump beam and the crystal length; c is the speed of light in vacuum; T_C is the cavity round trip time by wave; t is the dimensionless time representing the current time normalized to T_C ($t \longrightarrow t/T_C$); $\tau_\parallel = T_\parallel/T_C$; $\tau_\perp = T_\perp/T_C$;

$$P = \frac{k_l(1+\eta)}{g^2 T_C T_\perp}$$

is the pump parameter; and η is the excess pump power (absorbed in the active medium) over threshold (the latter corresponds to $\eta = 0$, and the lasing condition is $\eta > 0$). Operators $\Gamma(t)$, $\Gamma_N(t)$, and $F_l(t)$ are the operators of noise related to the polarization, inverse population of the active medium, and losses, respectively. The appearance of these operators and the terms $\tau_\perp^{-1}\sigma$, $\tau_\perp^{-1}\sigma^\dagger$, $\tau_\parallel^{-1}(P-N)$, and $k_l a_l$ in Eqs. (1)–(4) reflects the interaction with thermal reservoir. The coefficient k_l is given by the formula

$$k_l = \frac{\alpha_l L + 1 - R_l}{2}, \tag{5}$$

where α_l are linear losses in the crystal, R_l is the coefficient of intensity reflection of the cavity output mirror at the frequency ω_l, and L is the laser crystal length. The random force operator F_l is defined as [13, 17]

$$F_l(t) = \sqrt{\alpha_l L}\, b_l(t) + \sqrt{1-R_l}\, c_l(t), \tag{6}$$

where b_l and c_l are operators related to the losses in the crystal and in the cavity output mirror, respectively.

We also introduce the operators b_j and c_j related to the modes with the frequencies ω_j, which are defined as follows (see, e.g., [12]):

$$[b_j(t), b_k^\dagger(t')] = [c_j(t), c_k^\dagger(t')] = \delta_{jk}\delta(t-t'), \tag{7}$$

$$\langle b_j^\dagger(t)\rangle = \langle b_j(t)\rangle = \langle c_j(t)\rangle = \langle c_j^\dagger(t)\rangle = 0, \tag{8}$$

$$\begin{aligned}\langle b_j(t)b_k(t')\rangle &= \langle b_j^\dagger(t)b_k^\dagger(t')\rangle \\ &= \langle c_j(t)c_k(t')\rangle = \langle c_j^\dagger(t)c_k^\dagger(t')\rangle = 0,\end{aligned} \tag{9}$$

$$\langle b_j^\dagger(t)\rangle b_k(t')\rangle = \langle c_j^\dagger(t)\rangle c_k(t')\rangle = \bar{n}_j(T)\delta_{jk}\delta(t-t'), \tag{10}$$

where $j, k = 1, 2, 3$ (operators with different subscripts refer to the waves of different frequencies), δ_{jk} is the Kronecker delta, and

$$\bar{n}_j(T) = \left(\exp\frac{\hbar\omega_j}{kT} - 1\right)^{-1}$$

is the average number of thermal phonons for the jth mode at the thermal reservoir temperature T (at room temperature, $\bar{n}_j(T) \ll 1$).

According to relations (6)–(10),

$$[F_l(t), F_l^\dagger(t')] = 2k_l\delta(t-t'), \tag{11}$$

$$\langle F_l(t)\rangle = \langle F_l^\dagger(t)\rangle = 0, \tag{12}$$

$$\langle F_l^2(t)\rangle = \langle (F_l^\dagger(t))^2\rangle = 0, \tag{13}$$

$$\langle F_l^\dagger(t)F_l(t')\rangle = 2k_l\bar{n}_l\delta(t-t'). \tag{14}$$

The noise operators $\Gamma(t)$ and $\Gamma_N(t)$ possess the following properties [12]:

$$\langle \Gamma(t)F_l(t')\rangle = 0, \quad \langle \Gamma(t)F_l^\dagger(t')\rangle = 0, \tag{15}$$

$$\langle \Gamma_N(t)F_l(t')\rangle = 0, \quad \langle \Gamma_N(t)F_l^\dagger(t')\rangle = 0, \tag{16}$$

$$\langle \Gamma(t)\Gamma(t')\rangle = 0, \quad \langle \Gamma^\dagger(t)\Gamma^\dagger(t')\rangle = 0, \tag{17}$$

$$\begin{aligned}&\langle \Gamma^\dagger(t)\Gamma(t')\rangle \\ &= \left(\frac{T_C}{\tau_\perp}(M+N_0) + \frac{T_C}{2\tau_\parallel}(P-N_0)\right)\delta(t-t'),\end{aligned} \tag{18}$$

$$\begin{aligned}&\langle \Gamma(t)\Gamma^\dagger(t')\rangle \\ &= \left(\frac{T_C}{\tau_\perp}(M-N_0) - \frac{T_C}{2\tau_\parallel}(P-N_0)\right)\delta(t-t'),\end{aligned} \tag{19}$$

$$\langle \Gamma_N(t)\Gamma_N(t')\rangle = \frac{2T_C}{\tau_\parallel}\left(M - \frac{PN_0}{M}\right)\delta(t-t'), \tag{20}$$

$$\langle \Gamma^\dagger(t)\Gamma_N(t')\rangle = -\frac{T_C}{\tau_\parallel}\left(1 + \frac{P}{M}\right)\sigma_0^*\delta(t-t'), \tag{21}$$

$$\langle \Gamma(t)\Gamma_N(t')\rangle = -\frac{T_C}{\tau_\parallel}\left(\frac{P}{M} - 1\right)\sigma_0\delta(t-t'), \tag{22}$$

where N_0 and σ_0 are the stationary values of the inverse population and the medium polarization operators and P is the pump power (see formulas (32) below).

Using the equations presented in this section, it is possible to analyze the process of laser generation in an active crystal with allowance for losses in the crystal and in the output mirror.

2.2. Quasi-Phase-Matched Wave Interactions in a Nonlinear Crystal

Now let us proceed to the three-wave process ($\omega_3 = \omega_1 + \omega_2$) in a nonlinear optical RDS crystal situated in a ring cavity. With neglect of losses, the Hamiltonian of this interaction is

$$H_{\mathrm{NL}} = \hbar\varepsilon(Qa_3^\dagger a_2 a_1 + Q^* a_1^\dagger a_2^\dagger a_3), \qquad (23)$$

where $a_j^\dagger$ (a_j) are the operators of creation (annihilation) of a photon with the frequency ω_j,

$$\varepsilon = \pi L d_{\mathrm{eff}} \sqrt{\frac{8\pi\hbar\omega_1\omega_2\omega_3}{n_1 n_2 n_3 c^2 V}}$$

is the nonlinear wave coupling coefficient (n_j being the refractive index for the wave with the frequency ω_j, and d_{eff} being an effective nonlinearity coefficient dependent on the polarizations of interacting waves), and Q is the factor taking into account the RDS parameters of the crystal. The latter quantity is given by the formula [18]:

$$Q = \exp\left(-\frac{i\Delta kL}{2}\right)\mathrm{sinc}\frac{\Delta kL}{2}\tan\frac{\Delta k\Lambda}{4},$$

where $L = S\Lambda$ is the crystal length, S is the number of RDS periods in the given crystal, Λ is the period of modulation of the nonlinear susceptibility (RDS period),

$$\Delta k = k(\omega_3) - k(\omega_2) - k(\omega_1)$$

is the phase mismatch, $k(\omega_j)$ is the wavenumber corresponding to the frequency ω_j. If the modulation period Λ is much shorter than the characteristic nonlinear interaction length and the quasi-phase-matching condition is valid ($\Delta k = 2\pi m/\Lambda$, where m is an odd integer indicating the phase-matching order), the wave interaction in the RDS crystal proceeds in the same way as in a homogeneous medium [2, 19]. In this case, $Q = 2/\pi m$.

According to Eq. (23), the system of Heisenberg–Langevin equations for the nonlinear process under consideration with allowance for losses is as follows:

$$\frac{da_{1,2}}{dt} = -k_{1,2}a_{1,2} + i\varepsilon a_{2,1}^\dagger a_3 + F_{1,2}(t), \qquad (24)$$

$$\frac{da_3}{dt} = -k_3 a_3 + i\varepsilon a_1 a_2 + F_3(t). \qquad (25)$$

In deriving these equations, we excluded operators related to the thermal reservoir (this is achieved using a procedure similar to that used in deriving laser equations (1)–(4)). The coefficients k_l and the fluctuation operators $F_l(t)$ are given by formulas (5) and (6), respectively. Taking into account these relations and following the conventional procedure (see, e.g., [20]), one may readily check that commutation relations for the operators $a_j^\dagger$ and a_j are retained with time, so that

$$[a_j(t), a_k^\dagger(t)] = \delta_{jk}. \qquad (26)$$

The quantum equations (24) and (25) describe the process of three-wave nonlinear optical interactions in a crystal situated in the laser cavity with losses. The joint system of equations (1)–(4) describing the laser generation process and Eqs. (24)–(25) describing wave interactions in the nonlinear optical medium is a basis of the quantum theory of the process of optical self-frequency conversion.

3. PARAMETRIC LASER FREQUENCY SELF-CONVERSION IN AN ACTIVE NONLINEAR CRYSTAL

Let us consider the parametric self-down-conversion of the laser frequency in an active nonlinear crystal. The process involves simultaneous lasing and quasi-phase-matched division of the laser frequency (i.e., generation of the subharmonic with the frequency $\omega/2$). In the three-wave interaction, we have $\omega_1 = \omega_2 = \omega/2$ and $\omega_3 = \omega$, so that the system of equations describing this process is as follows:

$$\frac{da_1}{dt} = -k_1 a_1 + i2\varepsilon a_1^\dagger a_3 + F_1(t), \qquad (27)$$

$$\frac{da_3}{dt} = -k_3 a_3 - iT_C g\sigma + i\varepsilon a_1^2 + F_3(t), \qquad (28)$$

$$\frac{d\sigma}{dt} = \frac{1}{\tau_\perp}(-\sigma + igT_\perp a_3 N) + \Gamma(t), \qquad (29)$$

$$\frac{dN}{dt} = \frac{1}{\tau_\parallel}(P - N + i2gT_\parallel(\sigma a_3^\dagger - \sigma^\dagger a_3)) + \Gamma_N(t). \qquad (30)$$

Equations (27)–(30) are obtained by combining the system of nonlinear optical equations (24), (25) with the system of equations (1)–(4) describing laser generation. In addition, we take into account that the wave with the frequency $\omega_3 = \omega$ is enhanced in the active medium and involved in the nonlinear interaction. Subscript 1 refers to the $\omega/2$ subharmonic wave.

Since the analysis of the nonlinear system of equations (27)–(30) in the general case is impossible, we will only analyze two typical regimes from the standpoint of the parametric process under consideration: subthreshold and above-threshold generation of the subharmonic mode. Let the pump power exceed the

laser generation threshold. Then, the atomic and field operators can be represented as follows:

$$a_j(t) = A_{j0} + \delta a_j(t), \quad \sigma(t) = \sigma_0 + \delta\sigma(t), \quad N(t) = N_0 + \delta N(t), \tag{31}$$

where A_{j0}, σ_0, and N_0 are the classical quantities corresponding to the stationary solution of Eqs. (27)–(30) in the absence of random forces, while $\delta a_j(t)$, $\delta\sigma(t)$, and $\delta N(t)$ are the operators taking into account fluctuations. As can be readily seen, the quantities A_{j0}, σ_0, and N_0 are in fact the stationary average values, because the operator forces $F_1(t)$ and $F_3(t)$ are mutually uncorrelated and are not correlated with the fluctuation operators $\Gamma(t)$ and $\Gamma_N(t)$.

A stationary solution of the system (27)–(30) has the following form:

$$A_{30} = -\frac{ik_1}{2\varepsilon}, \quad N_0 = \frac{P}{1 + k_1^2 g^2 T_\parallel T_\perp/\varepsilon^2},$$
$$\sigma_0 = \frac{k_1 g T_\perp P}{2(\varepsilon + k_1^2 g^2 T_\parallel T_\perp/\varepsilon)}, \tag{32}$$
$$A_{10} = \pm\left(\frac{k_1 k_3(\eta - k_1^2 g^2 T_\parallel T_\perp/\varepsilon^2)}{2(\varepsilon^2 + k_1^2 g^2 T_\parallel T_\perp)}\right)^{1/2}.$$

This solution exists provided that $\eta \geq k_1^2 g^2 T_\parallel T_\perp/\varepsilon^2$, which corresponds to the pump power exceeding the subharmonic generation threshold ($A_{10} \neq 0$). According to formulas (32), the laser wave amplitude A_{30} in this above-threshold regime is independent of the pump power P. In other words, the amplitude A_{30} remains constant when the pump power grows, and all the excess supplied power is spent on increasing the amplitude of the subharmonic wave. As expected, the subharmonic phase in this parametric process is shifted relative to the laser wave phase by $\pi/2$ or $-\pi/2$ [21].

If the parameters of the crystal, pump, and cavity are such that $\eta < k_1^2 g^2 T_\parallel T_\perp/\varepsilon^2$, then the pump power is below the subharmonic generation threshold. For a subthreshold regime of subharmonic generation, a stationary solution of the system (27)–(30) is as follows:

$$A_{30} = \frac{\eta^{1/2}}{(4g^2 T_\parallel T_\perp)^{1/2}}, \quad N_0 = \frac{k_3}{g^2 T_C T_\perp},$$
$$\sigma_0 = \frac{i\eta^{1/2} k^3}{(4g^4 T_C^2 T_\parallel T_\perp)^{1/2}}, \quad A_{10} = 0. \tag{33}$$

A comparison of Eqs. (32) and (33) shows that exceeding the subharmonic generation threshold leads to a change in the stationary laser generation amplitude. This variation is cased by additional losses of the laser wave energy for the generation of the subharmonic wave.

Now we will analyze the spectra of fluctuations of the laser radiation and its subharmonic in the two limiting excitation regimes.

3.1. Subharmonic Generation in Above-Threshold Regime

The above-threshold regime of subharmonic generation corresponds to stationary solutions of the type given by formulas (32). In this case, system (27)–(30) linearized in the vicinity of the stationary solution leads to the following equations for the fluctuation operators:

$$\frac{d(\delta a_1)}{dt} = -k_1\delta a_1 + i2\varepsilon(A_{30}\delta a_1^\dagger + A_{10}\delta a_3) + F_1(t), \tag{34}$$

$$\frac{d(\delta a_3)}{dt} = -k_3\delta a_3 - iT_C g\delta\sigma + i\varepsilon 2A_{10}\delta a_1 + F_3(t), \tag{35}$$

$$\frac{d(\delta\sigma)}{dt} = -\frac{1}{\tau_\perp}\delta\sigma + iT_C g(N_0\delta a_3 + A_{30}\delta N) + \Gamma(t), \tag{36}$$

$$\frac{d(\delta N)}{dt} = -\frac{1}{\tau_\parallel}\delta N + i2T_C g \times(\sigma_0\delta a_3^\dagger + A_{30}^\dagger\delta\sigma - \sigma_0\delta a_3 - A_{30}\delta\sigma^\dagger) + \Gamma_N(t). \tag{37}$$

Let us use the Fourier transform of Eqs. (34)–(37), for example,

$$\delta a_1(\Omega) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\delta a_1(t)e^{i\Omega t}dt.$$

Upon solving Eqs. (34)–(37) for the Fourier components $\delta a_{1,3}(\Omega)$, we obtain

$$\delta a_1(\Omega) = \frac{x_1^*(-\Omega)Z(\Omega) - y_1(\Omega)Z^\dagger(-\Omega)}{y_1(\Omega)y_1^*(-\Omega) - x_1(\Omega)x_1^*(-\Omega)}, \tag{38}$$

$$\delta a_3(\Omega) = i[(i\Omega - k_1)\delta a_1(\Omega) + i2\varepsilon A_{30}\delta a_1^\dagger(-\Omega) + F_1(\Omega)](2\varepsilon A_{10})^{-1}, \tag{39}$$

where

$$x_1(\Omega) = -4\varepsilon^2 A_{10} - i2\varepsilon y A_{30}^* - (i\Omega - k_1)(i\Omega - k_3 - gT_C L_\perp N_0 + x),$$
$$y_1(\Omega) = y(i\Omega - k_1) - i2\varepsilon A_{30}(i\Omega - k_3 - gT_C L_\perp N_0 + x),$$
$$Z(\Omega) = (k_3 - i\Omega + gT_C N_0 L_\perp - x)F_1(\Omega) + y(\Omega)$$

$$\times F_1^\dagger(-\Omega) + i2\varepsilon A_{10}[F_3(\Omega) + i(1 - 2\gamma A_{30}^*)L_\perp\Gamma(\Omega)$$

$$- i2A_{30}\gamma L_\perp\Gamma^\dagger(-\Omega) + \gamma\Gamma_N(\Omega)],$$

$$\gamma = \gamma(\Omega) = 2gT_C A_{30}L_\perp(i\Omega - \tau_{\bullet]\exists}^{-1} + 4gT_C L_\perp|A_{30}|^2)^{-1},$$

$$\gamma\Sigma\hat{}<\mathrm{dmM}(-\Omega)$$

$$x = x(\Omega) = 2gT_C\gamma(A_{30}^*N_0L_\perp - i\sigma_0), \quad x^* = x^*(-\Omega),$$

$$y = y(\Omega) = 2gT_C\gamma(A_{30}N_0L_\perp + i\sigma_0), \quad y^* = y^*(-\Omega),$$

$$L_\perp = L_\perp(\Omega) = \frac{gT_C}{i\Omega - \tau_\perp^{-1}}.$$

We are interested in statistical properties of the quadrature components of the laser frequency and its subharmonic at the laser cavity output, which are measured using the balance homodyne detection technique [22]. Let us introduce the quadrature Fourier components,

$$X_j(\Omega) = \delta a_{j,\,\mathrm{out}}(\Omega) + \delta a_{j,\,\mathrm{out}}^\dagger(\Omega), \tag{40}$$

$$Y_j(\Omega) = i(\delta a_{j,\,\mathrm{out}}^\dagger(\Omega) - \delta a_{j,\,\mathrm{out}}(\Omega)), \tag{41}$$

where $\delta a_{j,\,\mathrm{out}}$ is the Fourier component of the field operator with the frequency ω_j at the cavity output determined by the following boundary condition at the output mirror [17]:

$$\delta a_{j,\,\mathrm{out}}(\Omega) = \sqrt{1 - R_j}\,\delta a_j(\Omega) - c_j(\Omega). \tag{42}$$

According to relations (40) and (41), the quadrature components are δ-correlated:

$$\langle X_j^\dagger(\Omega)X_j(\Omega')\rangle = S_{X,\,j}(\Omega)\delta(\Omega - \Omega'), \tag{43}$$

$$\langle Y_j^\dagger(\Omega)X_j(\Omega')\rangle = S_{Y,\,j}(\Omega)\delta(\Omega - \Omega'), \tag{44}$$

where $S_{X,j}(\Omega)$ and $S_{Y,j}(\Omega)$ are the spectral densities of fluctuations of the quadrature components. Using Eqs. (38)–(44), it is possible to derive analytical expressions for the spectral densities $S_{X,j}(\Omega)$ and $S_{Y,j}(\Omega)$, but the resulting formulas are rather cumbersome and are not presented here. Below (see Section 4) we present the plots of the spectral densities $S_{X,j}(\Omega)$ and $S_{Y,j}(\Omega)$ versus various parameters of the problem under consideration.

3.2. Subharmonic Generation in Subthreshold Regime

In the subthreshold regime of subharmonic generation, the stationary solution of system (27)–(30) is given by formulas (33). In this case, equations for the fluctuation operators are as follows:

$$\frac{d(\delta a_1)}{dt} = -k_1\delta a_1 + i2\varepsilon A_{30}\delta a_1^\dagger + F_1(t), \tag{45}$$

$$\frac{d(\delta a_3)}{dt} = -k_3\delta a_3 - iT_C g\delta\sigma + F_3(t), \tag{46}$$

$$\frac{d(\delta\sigma)}{dt} = -\frac{1}{\tau_\perp}\delta\sigma + iT_C g(N_0\delta a_3 + A_{30}\delta N) + \Gamma(t), \tag{47}$$

$$\frac{d(\delta N)}{dt} = -\frac{1}{\tau_\parallel}\delta N + i2T_C g(\sigma_0\delta a_3^\dagger + A_{30}\delta\sigma - \sigma_0^*\delta a_3 - A_{30}\delta\sigma^\dagger) + \Gamma_N(t). \tag{48}$$

Solving these equations, we obtain the following expressions for the Fourier spectra of field operators:

$$\delta a_{1,\,\mathrm{out}}(\Omega) = \sqrt{1 - R_1} \times \frac{(i\Omega - k_1)F_1(\Omega) - i2\varepsilon A_{30}F_1^\dagger(-\Omega)}{4\varepsilon^2|A_{30}|^2 - (i\Omega - k_1)^2} - c_1(\Omega), \tag{49}$$

$$\delta a_{3,\,\mathrm{out}}(\Omega) = \frac{\sqrt{1 - R_3}}{yy^* - xx^*} \times [zF_3(\Omega) - yF_3^\dagger(-\Omega) + (zx_2 - iA_{30}L_\perp\gamma y)\Gamma(\Omega) + (-iA_{30}L_\perp\gamma z - yx_2^*)\Gamma^\dagger(-\Omega) + (z\gamma - y\gamma^*)\Gamma_N(-\Omega)] - c_3(\Omega), \tag{50}$$

where

$$x_2 = x_2(\Omega) = i(1 - 2\gamma A_{30}^*)L_\perp, \quad x_2^* = x_2^*(-\Omega),$$

$$z = z(\Omega) = i\Omega - k_3 - gT_C N_0 L_\perp + 2gT_C\gamma^*(A_{30}N_0L_\perp + i\sigma_0).$$

The quadrature components of the laser frequency are also δ-correlated and their spectral densities are described by relations (43) and (44). Using formulas (49) and (50), it is possible to obtain analytical expressions for the spectral densities $S_{X,j}(\Omega)$ and $S_{Y,j}(\Omega)$. The plots of these spectral densities are also presented below in Section 4.

Thus, the proposed theory allowed us to obtain analytical expressions for the spectrum of fluctuations of the quadrature components of the laser frequency and its subharmonic excited simultaneously in the same crystal in two limiting regimes of the parametric self-frequency conversion. The spectral densities of both quadrature components depend on various parameters of the system under consideration. For this reason, these dependences will be considered for a particular case of the active nonlinear Nd:Mg:$LiNbO_3$ crystal, in which one of the self-frequency conversion processes

(namely, the quasi-phase-matched parametric self-frequency-doubling) was observed [23].

4. QUADRATURE-SQUEEZED LIGHT GENERATION AT THE LASER FREQUENCY AND ITS SUBHARMONIC IN Nd:Mg:$LiNbO_3$ RDS CRYSTAL

Now we will calculate the spectra of quadrature components of the laser frequency and its subharmonic for the process of parametric self-frequency conversion in a active nonlinear Nd:Mg:$LiNbO_3$ RDS crystal. The crystal can be pumped by radiation with a wavelength of 0.81 μm, for example, from a diode laser [23]. Laser generation is observed at a wavelength of 1.084 μm, so that the subharmonic wavelength is 2.168 μm. In parametric interaction of the *ee–e* type, the laser and subharmonic waves possess extraordinary polarization. The condition of quasi-phase-matched interaction is obeyed for a nonlinear susceptibility modulation period of $\Lambda = 22$ μm [23]. The other parameters of the Nd:Mg:$LiNbO_3$ RDS crystal were as follows: effective nonlinearity coefficient, $d_{\text{eff}} = 34.4$ pm/V; linear losses at the frequencies $\omega/2$ and ω, $\alpha_1 = \alpha_3 = 0.08$ cm^{-1}; saturation intensity, $I_S = 10^4$ W/cm^2; number of active dopant atoms, $M = 10^{18}$; longitudinal and transverse relaxation times, $T_\parallel = 10^{-4}$ s and $T_\perp = 6.7 \times 10^{-10}$ s, respectively; typical cavity round trip time, $T_C = 3 \times 10^{-11}$ s.

4.1. Spectral Characteristics for Subharmonic Generation in Above-Threshold Regime

The characteristic spectral densities of quadrature components calculated using formulas (38)–(44) are presented in Figs. 2 and 3, where the unit spectral density corresponds to the standard quantum limit. Figure 2 shows the spectra of fluctuations in one of the laser radiation quadratures at various pump powers. As can be seen, the more the pump power exceeds the threshold level, the stronger the suppression of quadrature fluctuations relative to the quantum limit in a certain spectral region. In other words, a quadrature-squeezed light is generated in this spectral region. According to Fig. 2, the maximum efficiency of fluctuation suppression below the standard quantum limit in the quadrature component for the laser frequency can amount to 10%. These results refer to the *X*-quadrature field components; as for fluctuations of the *Y*-quadrature components, these must increase in accordance with the uncertainty relation, which was confirmed by the results of our calculations.

Figure 3 shows the spectra of fluctuations in one of the subharmonic field quadratures for various reflection coefficients of the cavity output mirror. As can be seen, fluctuations in this quadrature component are virtually not suppressed. Thus, the active nonlinear Nd:Mg:$LiNbO_3$ crystal with the above parameters exhibits rather insignificant suppression of laser radiation fluctuations in the regime of above-threshold subharmonic generation.

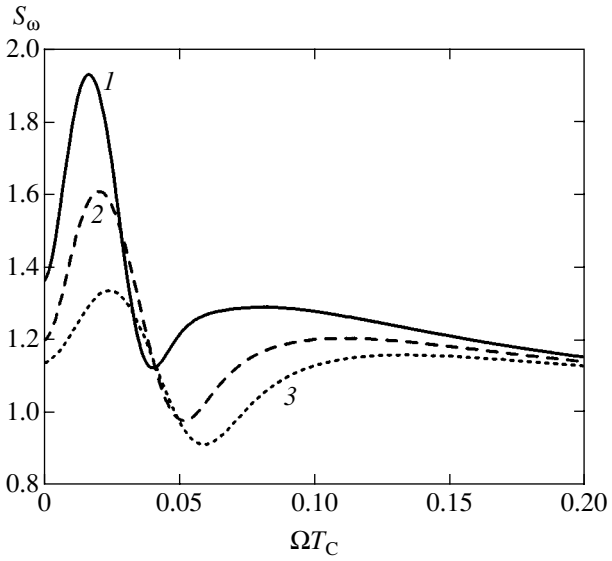


Fig. 2. The spectra of fluctuations in a laser field quadrature in a above-threshold regime of subharmonic generation, calculated for various ratios of the pump power to the threshold power $P_{\text{pump}}/P_{\text{th}} = 10$ (*1*), 15 (*2*), and 20 (*3*). Other parameters: $R_{\omega/2} = 1$; $R_\omega = 0.9$; $L = 0.5$ cm.

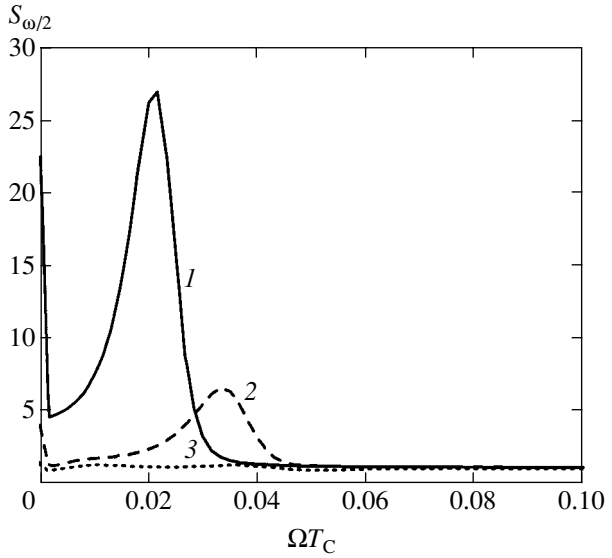


Fig. 3. The spectra of fluctuations in a subharmonic field quadrature in a above-threshold regime of generation, calculated for various coefficients of reflection of the cavity output mirror $R_{\omega/2} = 0.95$ (*1*), 0.97 (*2*), 0.99 (*3*). Other parameters: $R_\omega = 1$; $P_{\text{pump}}/P_{\text{th}} = 15$; $L = 0.5$ cm.

4.2. Spectral Characteristics for Subharmonic Generation in Subthreshold Regime

Figres 4–6 show the spectra of fluctuations in one of the quadrature components of the subharmonic field,

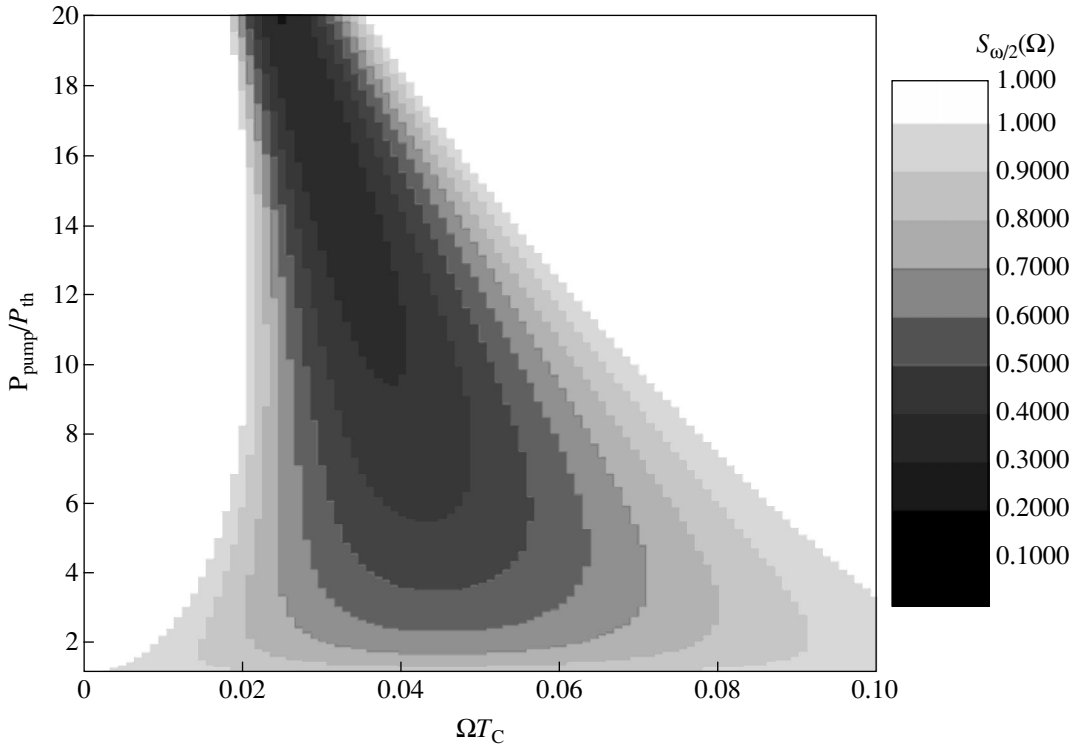


Fig. 4. Variation of the spectrum of fluctuations in a subharmonic field quadrature in a subthreshold regime of generation depending on the ratio of the pump power to the threshold power P_{pump}/P_{th} . Other parameters: $L = 0.5$ cm; $R_\omega = 1$; $R_{\omega/2} = 0.8$. The inset shows the scale of the spectral density levels, in which a darker color corresponds to a lower level of fluctuations.

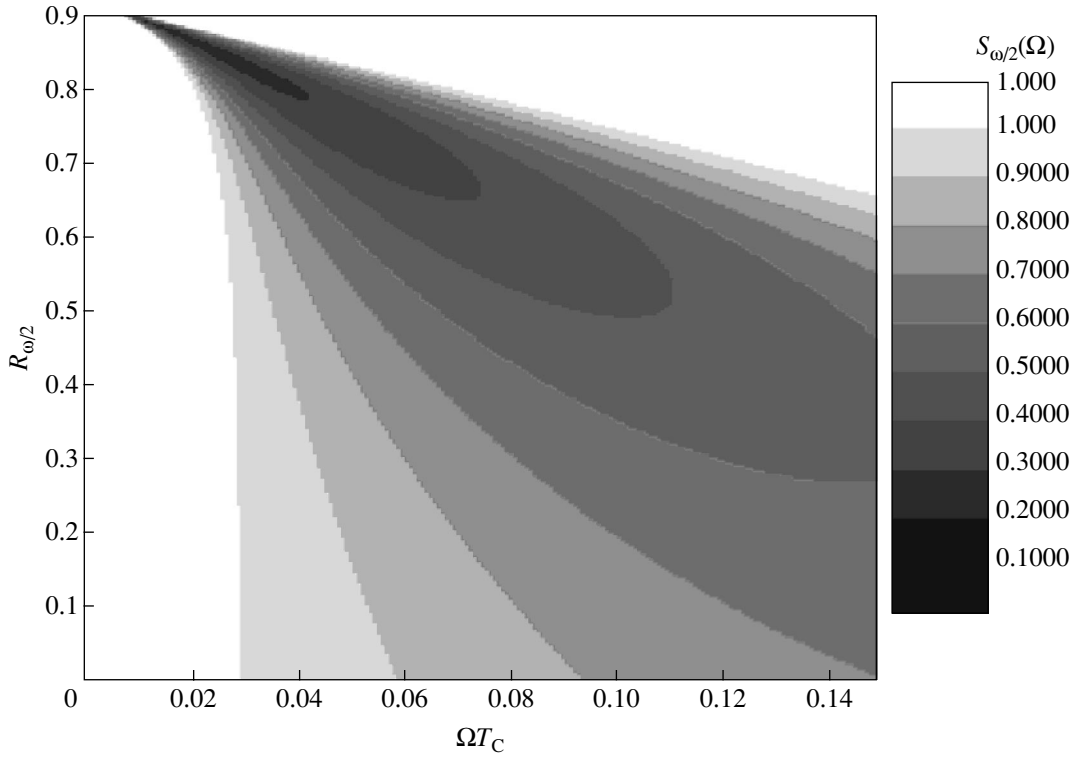


Fig. 5. Variation of the spectrum of fluctuations in a subharmonic field quadrature in a subthreshold regime of generation depending on the coefficient of reflection $R_{\omega/2}$ of the cavity output mirror at the subharmonic frequency $\omega/2$. Other parameters: $P_{pump}/P_{th} = 10$; $R_\omega = 1$; $L = 0.5$ cm. The inset shows the scale of the spectral density levels, in which a darker color corresponds to a lower level of fluctuations.

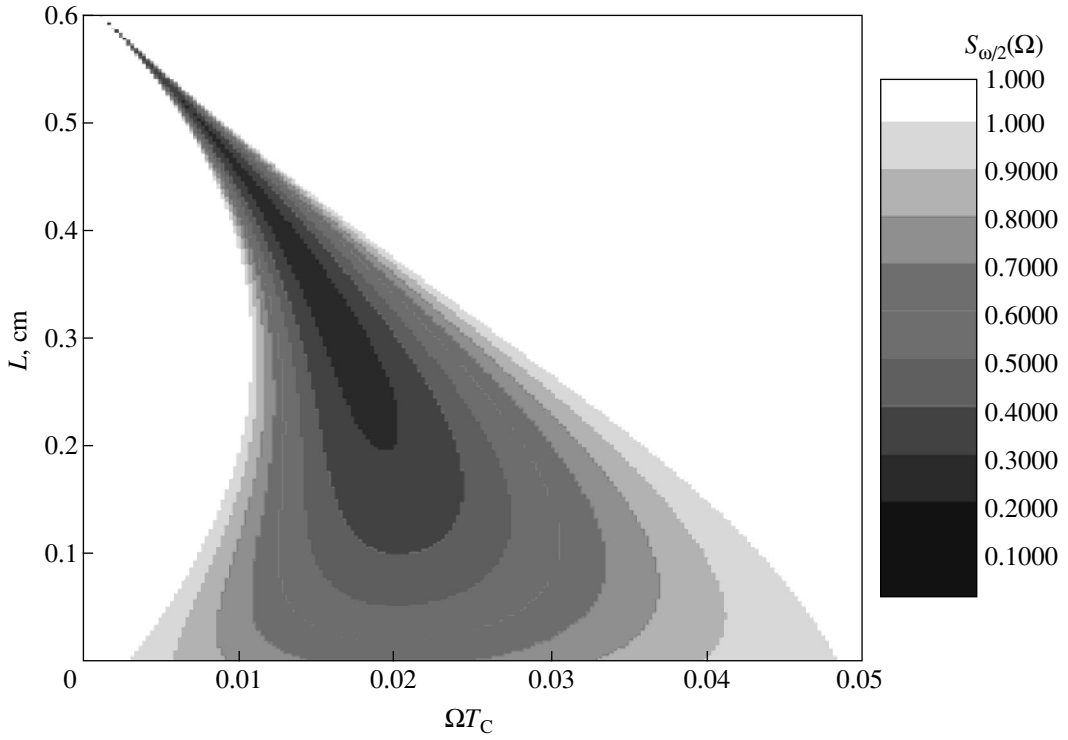


Fig. 6. Variation of the spectrum of fluctuations in a subharmonic field quadrature in a subthreshold regime of generation depending on the length L of the active nonlinear crystal. Other parameters: $P_{pump}/P_{th} = 10$; $R_\omega = 1$; $R_{\omega/2} = 0.9$. The inset shows the scale of the spectral density levels, in which darker color corresponds to a lower level of fluctuations.

which were calculated using formulas (40)–(44), (49), and (50). The calculations were performed for various pump powers, reflection coefficients of the cavity output mirror, and crystal lengths. The parameters of the Nd:Mg:$LiNbO_3$ crystal were the same as those used above for the above-threshold regime of subharmonic generation; other parameters are indicated in the figure captions. In Figs. 4–6, the level of the standard quantum limit also corresponds to the unit spectral density.

Analysis of the data presented in Figs. 4–6 shows that the maximum suppression (corresponding to the darkest area) of the quadrature field fluctuations takes place at a nonzero frequency, in contrast to the case of the above-threshold regime of subharmonic generation. Figure 4 shows that the higher the pump power, the stronger the suppression of fluctuations in the subharmonic quadrature. However, this is accompanied by narrowing of the spectral band in which the fluctuations are effectively suppressed. According to Fig. 5, an increase in the reflectance of the output mirror for the subharmonic wave leads to a significant growth in the efficiency of suppressing fluctuations in the corresponding quadrature. An analogous behavior is observed in response to a change in the length of the active nonlinear crystal as depicted in Fig. 6, which was calculated for

$$\eta + 1 = \frac{P_{pump}(1 - e^{-\alpha_p L})}{P_{th}},$$

where $P_{pump} = 0.5$ W, $P_{th} = 1.25k_3$ W, k_3 is the dimensionless quantity given by formula (5), and $\alpha_p = 2\ln 2$ cm^{-1}. The maximum efficiency of suppression of the subharmonic wave quadrature fluctuations for the indicated parameters is about 90% (for the X-quadrature field). As for the laser frequency generated in this regime, our calculations showed that virtually no suppression of fluctuations take place.

5. PHOTON STATISTICS IN PARAMETRIC SELF-FREQUENCY CONVERSION

In the general case, calculations of the photons distribution functions for the laser frequency and subharmonic frequency encounter considerable difficulties. For this reason, we will restrict the consideration to analysis of the statistical properties of photons within the framework of the second-order moments and calculate the average photon number and its dispersion. In order to simplify calculations, we consider the average photon number $\langle n_j \rangle = \langle a_j^\dagger a_j \rangle$ and the Fano factor

$$F_j = \frac{\langle n_j^2 \rangle - \langle n_j \rangle^2}{\langle n_j \rangle}$$

at the subharmonic frequency ($j = 1$) and laser frequency ($j = 3$) inside the cavity.

In the vicinity of the stationary solution of Eqs. (27)–(30), according to expressions (31),

$$\langle n_j \rangle = |A_{j0}|^2 + \langle \delta a_j^\dagger \delta a_j \rangle. \quad (51)$$

For the dispersion of the photon number, to within the same accuracy we have

$$\sigma_j^2 = \langle n_j^2 \rangle - \langle n_j \rangle^2 = |A_{j0}|^2 (1 + 2\langle \delta a_j^\dagger \delta a_j \rangle) + A_{j0}^2 (\langle \delta a_j \delta a_j \rangle + \langle \delta a_j^\dagger \delta a_j^\dagger \rangle). \quad (52)$$

In the above-threshold regime of subharmonic generation, we have $|A_{j0}|^2 \gg \langle \delta a_j^\dagger \delta a_j \rangle$ and, hence,

$$\langle n_j \rangle \approx |A_{j0}|^2, \quad \langle n_j^2 \rangle - \langle n_j \rangle^2 \approx |A_{j0}|^2. \quad (53)$$

This implies that the Fano factor is $F_j = \sigma_j^2 / \langle n_j \rangle \approx 1$ ($j = 1, 3$) and, therefore, the photon statistics in the laser cavity excited at these frequencies is Poisson-like.

In the subthreshold regime of subharmonic generation, the photon statistics at the laser frequency remains Poisson-like as well ($|A_{10}| = 0$). The photon statistics at the subharmonic frequency depends on the average values $\langle \delta a_1^\dagger \delta a_1 \rangle$ and $\langle (\delta a_1^\dagger \delta a_1)^2 \rangle$. The time variation of the operator $\delta a_1^\dagger (\delta a_1)$ under the action of the random force can be readily determined using Eq. (45) and the Hermitian-conjugated relation. As a result, we obtain

$$\delta a_1^\dagger(t) = \int_0^t e^{-k_1 \theta} (F_1^\dagger(t-\theta)\cosh(2\varepsilon A_{30}\theta) + iF_1(t-\theta)\sinh(2\varepsilon A_{30}\theta)) d\theta. \quad (54)$$

Using this solution, taking into account the statistical properties of fluctuations related to the operators $F_1^\dagger(t)$ and $F_1(t)$ according to Eqs. (12)–(14), and considering the stationary subthreshold regime of subharmonic generation, we obtain expressions for the average photon number,

$$\langle n_1 \rangle = \frac{(2\varepsilon A_{30})^2 + 2k_1^2 \bar{n}_1(T)}{2(k_1^2 - (2\varepsilon A_{30})^2)}, \quad (55)$$

and for the dispersion of this number,

$$\sigma_1^2 = \frac{k_1^2}{2(k_1^2 - (2\varepsilon A_{30})^2)} \times [3(2\varepsilon A_{30})^2 + 8(2\varepsilon A_{30})^2 \bar{n}_1(T) + 4(k_1^2 - (2\varepsilon A_{30})^2)\bar{n}_1^2(T)]. \quad (56)$$

Taking into account that, according to (33),

$$A_{30} = \frac{\eta^{1/2}}{(4g^2 T_\parallel T_\perp)^{1/2}},$$

we obtain the relation

$$k_1^2 - (2\varepsilon A_{30})^2 = k_1^2 - \frac{\varepsilon^2 \eta}{g^2 T_\parallel T_\perp}.$$

Thus, as expected, the average number of photons at the subharmonic frequency, the dispersion of this number, and the corresponding Fano factor sharply increase when the pump power approaches the subharmonic generation threshold ($\eta \approx k_1^2 g^2 T_\parallel T_\perp / \varepsilon^2$).

If the number of thermal phonons is $\bar{n}_1(T) \ll 1$, formulas (55) an (56) yield for the Fano factor

$$F_1 = 1.5 + 2\langle n_1 \rangle. \quad (57)$$

In the other limiting case ($\bar{n}_1(T) \gg 1$), the Fano factor is independent of proximity to the subharmonic generation threshold, being determined by the average number of thermal phonons: $F_1 = \bar{n}_1(T)$.

It should be noted that, far away from the subharmonic generation threshold at $\langle n_1 \rangle \ll 1$, the Fano factor for the excited biphotons, $F_1 = 1.5$, differs from the value ($F = 2$) for the biphotons generated as a result of the spontaneous parametric scattering [17, 22]. We believe that a decrease in the Fano factor is related to the inertial character of the response of the system under consideration to a random action (i.e., to the presence of a term describing losses in Eq. (45)).

6. CONCLUSIONS

We have developed a theory describing the generation of nonclassical light during laser self-frequency conversion in an active nonlinear crystal. Using this description, based on the Heisenberg–Langevin equations, we have thoroughly analyzed the process of self-down-conversion (halving) of the laser frequency. The calculations were performed for an arbitrary relation between the photon lifetime in the laser cavity and the characteristic times of the inverse population relaxation and the active medium polarization. We considered the process of subharmonic generation in the sub- and above-threshold regimes. It was established that there are optimum parameters of the crystal, pumping, and cavity favoring the maximum efficiency of suppressing fluctuations in the quadrature components of the laser frequency and subharmonic fields below the level of the standard quantum limit. We have also considered the photon statistics for the generated light fields.

The results of our theoretical analysis show that the active nonlinear RDS crystals can be used as effective sources of nonclassical light. Further expansion of the possibilities of such crystals with respect to self-frequency conversion due to the presence of RDS opens good prospects for the creation of small-size sources of nonclassical radiation in various wavelength ranges.

RDS crystals can feature, besides the traditional nonlinear optical interactions, consecutive three-wave interactions of the optical modes having the common pump wave. In this context, it is important to develop a quantum theory of such processes in active nonlinear RDS crystals. In such systems, the laser cavity features three coupled processes: lasing and two nonlinear wave interactions. The theory of self-frequency conversion during consecutive interactions in RDS crystals can be developed through generalization of the approach presented in this study.

ACKNOWLEDGMENTS

The authors are grateful to Yu.M. Golubev, T.Yu. Golubeva, I.V. Sokolov, and G.D. Laptev for fruitful discussions.

This study was supported in part by the INTAS Foundation, grant no. 01-2097.

REFERENCES

1. R. L. Byer, J. Nonlinear Opt. Phys. Mater. **6**, 549 (1997).
2. A. S. Chirkin, V. V. Volkov, G. D. Laptev, and E. Yu. Morozov, Kvantovaya Élektron. (Moscow) **30**, 847 (2000).
3. A. Levenson, P. Vidakovich, and C. Simonneau, J. Eur. Opt. Soc. A **7**, 81 (1998).
4. S. Tanzilli, H. De Riedmatten, W. Tittel, *et al.*, Electron. Lett. **37**, 1 (2001).
5. K. S. Zhang, M. Martinelli, T. Condreau, *et al.*, Phys. Rev. A **64**, 033815 (2001).
6. A. S. Chirkin, J. Opt. B: Quantum Semiclassic. Opt. **4**, S91 (2002).
7. *Physics of the Quantum Information*, Ed. by D. Bouwmeester, A. Ekert, and A. Zeilinger (Springer, Berlin, 2000; Postmarket, Moscow, 2002).
8. G. D. Laptev, A. A. Novikov, and A. S. Chirkin, J. Russ. Laser Res. **23**, 183 (2002).
9. K. S. Abedin, T. Tsuritani, M. Sato, and H. Ito, Appl. Phys. Lett. **70**, 10 (1997).
10. J. Capmany, D. Callejo, V. Bermudez, *et al.*, Appl. Phys. Lett. **79**, 293 (2001).
11. J. Capmany, Appl. Phys. Lett. **78**, 144 (2001).
12. H. Haken, *Light 2. Lasers Light Dynamics* (North-Holland, Amsterdam, 1985; Mir, Moscow, 1988).
13. D. F. Walls and G. J. Milburn, *Quantum Optics* (Springer, Berlin, 1995).
14. M. O. Skully and M. S. Zubairy, *Quantum Optics* (Cambridge University Press, Cambridge, 1997).
15. V. N. Gorbachev and E. S. Polzik, Zh. Éksp. Teor. Fiz. **96**, 1984 (1989) [Sov. Phys. JETP **69**, 1119 (1989)].
16. R. Schack, A. Sizmann, and A. Schenzle, Phys. Rev. A **43**, 6303 (1991).
17. C. W. Gardiner, *Quantum Noise* (Springer, Berlin, 1991).
18. G. D. Laptev and A. A. Novikov, Kvantovaya Élektron. (Moscow) **31**, 981 (2001).
19. E. V. Makeev and A. S. Chirkin, J. Russ. Laser Res. **24**, 544 (2003).
20. M. Lax, *Fluctuation and Coherence Phenomena in Classical and Quantum Physics* (Gordon and Breach, New York, 1968).
21. S. A. Akhmanov, Yu. E. D'yakov, and A. S. Chirkin, *Introduction to Statistical Radio Physics and Optics* (Nauka, Moscow, 1981) [in Russian].
22. L. Mandel and E. Wolf, *Optical Coherence and Quantum Optics* (Cambridge Univ. Press, Cambridge, 1995).
23. G. D. Laptev, A. A. Novikov, and A. S. Chirkin, Pis'ma Zh. Éksp. Teor. Fiz. **78**, 45 (2003) [JETP Lett. **78**, 38 (2003)].

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 99, No. 5, 2004, pp. 958–964.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 5, 2004, pp. 1101–1108.

ATOMS, SPECTRA, RADIATION

Polarization Bremsstrahlung of Heavy Charged Particles in a Polycrystal

V. A. Astapenko

Moscow Institute for Physics and Technology, Dolgoprudnyĭ, Moscow oblast, 141700 Russia
e-mail: astval@hotmail.com

Received May 6, 2004

Abstract—Polarization bremsstrahlung (PBS) of a fast ion scattered in a polycrystalline medium is calculated and analyzed with allowance for the contributions from the coherent and incoherent channels of the process. It is shown that scattering of a projectile from the crystal lattice of the target is responsible for typical features of the PBS spectrum. For example, PBS is suppressed (as compared to radiation emitted by a single atom) in the low-frequency part of the spectrum, where coherent PBS dominates. In the intermediate spectral region, a step structure is formed as a result of "elimination" of the contribution from the reciprocal lattice vector with a preset magnitude to the coherent component of the process. Finally, incoherent PBS dominates in the high-frequency part of the spectrum and the process occurs as in the case of a single atom. These spectral peculiarities of PBS are determined by the structure of the target and depend on the velocity of the projectile and the emission angle, and can be observed in experiments dealing with radiation emitted by fast charged particles in thin polycrystalline films. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The use of multiply charged ions in contemporary experiments on the interaction of charged particles with solid targets [1] necessitates analysis of possible mechanisms of radiation emission in the case of such scattering.

It is well known [2] that ordinary bremsstrahlung (which is also known as static bremsstrahlung) is strongly suppressed for small values of the charge-to-mass ratio of the particle being scattered. This study is devoted to calculation and analysis of bremsstrahlung of the emission channel, namely, polarization bremsstrahlung (PBS) [3, 4] induced by a heavy ion in a polycrystal. This type of radiation, which is supplementary to static bremsstrahlung, is independent of the projectile mass and should dominate over ordinary bremsstrahlung in the case considered here.

The existence of PBS of heavy charged particles was first indicated in [5], where the bremsstrahlung of a proton scattered from a hydrogen atom was calculated. Subsequently, the concept of PBS was used in [6] for explaining the emission spectra of protons with an energy on the order of 1 MeV during their scattering from thin aluminum films. The effect of the structure of a solid target on PBS was not considered in [6]. The band nature of the PBS energy spectrum was accounted for in [7], where the PBS spectrum near the adsorption edge was calculated. However, the collective effects in PBS associated with the crystal lattice of the target were not analyzed in [7]. Such effects were considered for the first time in [8] for PBS of a relativistic electron in a polycrystal. It was shown that the PBS spectrum of a relativistic charged particle is formed by an aggregate of peaks. The position of such peaks is determined by the Bragg condition for scattering of the intrinsic field of an electron from the crystal lattice of the target. These peculiarities of the PBS spectrum of a relativistic electron in a polycrystal were experimentally observed in [9], where the frequency-angular distribution of radiation emitted by 2.4-MeV electrons during their scattering from a thin aluminum foil was measured. The position and height of the peaks detected in these measurements were in good agreement with theoretical calculations [8].

As we pass to a fast, but nonrelativistic projectile, the PBS spectrum in a polycrystal must be transformed since the intrinsic field of the charged particle with a nonrelativistic velocity substantially differs from the field of a transverse electromagnetic wave. Accordingly, the analogy between PBS and Bragg scattering of virtual photons from a crystal lattice becomes inadequate. In addition, the noticeable contribution to the PBS intensity in the nonrelativistic case must lead to incoherent scattering of the intrinsic field of the projectile into a bremsstrahlung photon, which is insignificant in the relativistic case [8].

Thus, we will study PBS of a fast ion in a polycrystal in a wide energy range with allowance for both coherent and incoherent emission channels.

2. BASIC RELATIONS

We proceed from the expression for the PBS amplitude in the case when a charged particle is scattered from a single atom [3]. In the first Born approximation,

the corresponding formula for an electron and for an ion has the same appearance except for factor Z_p (Z_p is the ion charge) (we use the atomic units $\hbar = e = m_e = 1$ everywhere):

$$T_{if} = 2\pi i\delta(\varepsilon_f + E_f + \omega - \varepsilon_i - E_i)(q^0)^2$$

$$\times \sqrt{\frac{2\pi}{\omega}}\frac{1}{c}e^*_{\mathbf{k},\sigma,l}A_s^{(p)}(q)\langle f|\hat{c}^{ls}(\omega, \mathbf{k}; q^0, \mathbf{q})|i\rangle, \qquad (1)$$

$$q^0 = \varepsilon_f - \varepsilon_i, \quad \mathbf{q} = \mathbf{p}_f - \mathbf{p}_i,$$

where c is the velocity of light, ω and $\mathbf{k}$ are the frequency and the wave vector of the emitted photon, ε_j and $\mathbf{p}_j$ are the energy and momentum of the projectile, E_j is the energy of the target atom, $\mathbf{e}_{\mathbf{k},\sigma}$ is the unit vector of photon polarization, and $\mathbf{A}^{(p)}(q)$ is the vector potential of the electromagnetic field of the scattering charge,

$$\mathbf{A}^{(p)}(q) = \left(\frac{4\pi c Z_p}{q^0}\right)\frac{q^0\mathbf{v}/c^2 - \mathbf{q}}{(q^0/c)^2 - \mathbf{q}^2}, \qquad (2)$$

$\mathbf{v}$ being the velocity of the projectile. In expression (1), allowance is made for the possibility of target excitation in the course of PBS $|i\rangle \longrightarrow |f\rangle$; consequently, an atom is described by the matrix element of the scattering operator

$$\langle f|\hat{c}^{ls}(k, \mathbf{q})|i\rangle.$$

The use of the first order of perturbation theory in the interaction of a projectile with an atom in formula (1) in the case of PBS of a heavy charged particle is justified by the smallness of perturbation in the motion of the projectile in the course of the process up to velocities on the order of the atomic velocity. Indeed, it is well known [3] that the main contribution to PBS comes from transferred momenta which are smaller than the characteristic atomic momentum so that the minimal approach distance is longer than or on the order of the mean atomic radius. It can be easily seen, however, that the Coulomb interaction energy at such distances is much smaller than the kinetic energy of a heavy particle impinging at a velocity higher than the atomic velocity. Consequently, the condition for the applicability of perturbation theory in the interaction of a projectile with an atom is satisfied.

Expression (1) for the PBS amplitude makes it possible to interpret this process as the scattering of the intrinsic field of the projectile particle (virtual photon) into a real photon at bound electrons of the target. Such an interpretation discloses common features of PBS and scattering of a real photon. In particular, if an atomic electron is ionized in the course of the process, we are dealing with Compton scattering or radiation-induced ionization in the case of PBS [6, 10]. Henceforth, we will consider PBS without target excitation $|i\rangle = |f\rangle$; in this case, $q^0 = -\omega$, and the matrix element of the scattering operator can be expressed in terms of the dynamic polarizability of the atom,

$$\langle f|\hat{c}^{ls}(\omega, \mathbf{k}; q^0, \mathbf{q})|i\rangle \longrightarrow \alpha(\omega, \mathbf{q} + \mathbf{k})\delta_{ls}.$$

Using this substitution, we obtain the following expression for the PBS amplitude at the jth atom of the crystal lattice:

$$T_j = 2\pi i\delta(\omega + q\mathbf{v})\sqrt{\frac{2\pi}{\omega}}$$
$$\times \frac{1}{c}[\mathbf{n}\mathbf{A}^{(p)}(q)]\omega^2\alpha(\omega, \mathbf{q} + \mathbf{k})\exp(i(\mathbf{q} + \mathbf{k})\cdot\mathbf{r}_j), \qquad (3)$$

where $\mathbf{n}$ is the unit vector in the direction of emission of a photon. While deriving formula (3) from expression (1), we carried out averaging over polarization of the emitted photon. The factor $\exp(i(\mathbf{q} + \mathbf{k})\cdot\mathbf{r}_j)$ introduced in expression (3) describes the phase relations between the contributions from different atoms of the crystal lattice to the amplitude of the process ($\mathbf{r}_j$ is the radius vector of the nucleus of the jth atom).

It should be noted that PBS with excitation of an atom is an incoherent process in respect of the contribution of atomic electrons; the cross section of this process is proportional to the number of electrons in the target. On the contrary, PBS without excitation of the atom is a coherent process whose cross section is proportional to the squared number of electrons. Thus, the latter process dominates over the first one in the frequency range $\omega < v/r_a$ (r_a is the characteristic atomic radius); this is demonstrated, for example, in [11]. Here, we consider the case when the above inequality is satisfied and, hence, PBS with atomic excitation can be discarded.

To derive the expression for the PBS intensity per unit volume of the medium, we must sum expression (3) over the atoms in unit volume of the crystal lattice,

$$T_{\text{tot}} = \sum T_j.$$

Then, using the standard quantum-mechanical procedure, we have

$$\frac{dI}{d\omega d\Omega_{\mathbf{n}}} = \frac{\omega^3}{(2\pi c)^3}\int\frac{d\mathbf{q}_1}{(2\pi)^3}\lim_{t\to\infty}\frac{|T_{\text{tot}}|^2}{t}, \qquad (4)$$

where t is the time interval; and $\mathbf{q}_1 = \mathbf{q} + \mathbf{k}$ is the momentum imparted to the medium from the projectile

during emission. Substituting expression (3) into formula (4), we obtain

$$\frac{dI}{d\omega d\Omega_{\mathbf{n}}} = \frac{\omega^3}{(2\pi c)^3}\int\frac{d\mathbf{q}}{(2\pi)^3}\frac{(2\pi)^2}{\omega c^2}$$

$$\times\delta(\omega+\mathbf{q}\cdot\mathbf{v})[\mathbf{n}\cdot\mathbf{A}^{(p)}(q)]^2\left|\omega^2\alpha(\omega,\mathbf{q}_1)\right|^2 \qquad (5)$$

$$\times\sum_{j,j'}\exp(i\mathbf{q}_1\cdot(\mathbf{r}_j-\mathbf{r}_{j'})).$$

Here, $d\Omega_{\mathbf{n}}$ is the solid angle in the direction of emission of a photon. The double sum on the right-hand side of equality (5) is the form factor of the medium; for a crystal, this quantity has the form [12]

$$\sum_{j,j'}\exp(i(\mathbf{q}_1)(\mathbf{r}_j-\mathbf{r}_{j'})) = n_{\mathrm{i}}(1-\exp(-u^2q_1^2)) \qquad (6)$$

$$+n_{\mathrm{i}}^2(2\pi)^3\sum_{\mathbf{g}}\exp(-u^2g^2)S^2(\mathbf{g})\delta(\mathbf{q}_1-\mathbf{g}),$$

where n_{i} is the ion concentration in the lattice, $\mathbf{g}$ is the reciprocal lattice vector, u is the mean thermal deviation of lattice atoms from their equilibrium positions, and $S(\mathbf{g})$ is the form factor of a unit cell normalized to unity. The first term on the right-hand side of equality (6) describes incoherent scattering of electromagnetic field from lattice ions. It is proportional to the first power of the atomic concentration. The second term describes coherent scattering proportional to the squared concentration of ions. It can be seen from formula (6) that coherent scattering occurs only when the momentum transferred to the medium is equal to the reciprocal lattice vector ($\mathbf{q}_1=\mathbf{g}$) in contrast to an incoherent process. Formally, this circumstance is reflected in the presence of the delta function in the coherent term. It should be noted that these delta functions are "eliminated" after the integration of the coherent part of the PBS intensity in formula (4) with respect to transferred momentum $\mathbf{q}_1$.

In accordance with the above-mentioned division of the form factor of the medium into the coherent and incoherent parts, the differential PBS intensity per unit volume can be written in the form

$$\frac{dI}{d\omega d\Omega_{\mathbf{n}}dV} = \left(\frac{dI}{d\omega d\Omega_{\mathbf{n}}dV}\right)_{\mathrm{incoh}} + \left(\frac{dI}{d\omega d\Omega_{\mathbf{n}}dV}\right)_{\mathrm{coh}}. \qquad (7)$$

We will henceforth consider PBS in a polycrystalline medium. For this reason, intensity (7) should be averaged over the direction of the reciprocal lattice vector as in the case of X-ray scattering in a polycrystal [13]:

$$\left(\frac{dI}{d\omega d\Omega_{\mathbf{n}}dV}\right)_{\mathrm{polycr}} = \int\frac{dI}{d\omega d\Omega_{\mathbf{n}}dV}\frac{d\Omega_{\mathbf{g}}}{4\pi}. \qquad (8)$$

For convenience of comparison with experiment, it is expedient to pass from PBS intensity (8) to the differential yield of the number of photons from a unit length of the polycrystal into a unit solid angle in a unit frequency range:

$$\frac{dN}{d\omega d\Omega_{\mathbf{n}}dl} = \frac{1}{\omega v}\left(\frac{dI}{d\omega d\Omega_{\mathbf{n}}dV}\right). \qquad (9)$$

Using formulas (4)–(9), we arrive at the following expressions for the incoherent and coherent contributions to the differential yield of PBS photons emitted by a fast charged particle in the polycrystal:

$$\left(\frac{dN}{d\omega d\Omega_{\mathbf{n}}dl}\right)_{\mathrm{incoh}} = \frac{2n_{\mathrm{i}}Z_p^2\omega^3}{\pi v^2c^3}$$

$$\times\int_{q_{1\min}}^{2\mu v}|\alpha(\omega,q_1)|^2 I\phi(q_1,v,\omega,\theta) \qquad (10)$$

$$\times[1-\exp(-q_1^2u^2)]\frac{dq_1}{q_1},$$

$$\left(\frac{dN}{d\omega d\Omega_{\mathbf{n}}dl}\right)_{\mathrm{coh}} = \frac{4\pi n_{\mathrm{i}}^2Z_p^2\omega^3}{v^2c^3}$$

$$\times\sum_g N(g)\frac{|\alpha(\omega,g)|^2}{g^3}\exp(-g^2u^2)I\phi(q,v,\omega,\theta) \qquad (11)$$

$$\times\Theta\left(gv-\omega\left(1-\frac{v}{c}\cos\theta\right)\right).$$

Here, $\theta=\widehat{\mathbf{p}_i\mathbf{k}}$ is the photon emission angle, $N(g)$ is the number of reciprocal lattice vectors with a preset magnitude g, $\Theta(x)$ is the Heaviside step function, and μ is the reduced mass of a lattice ion and the projectile. The minimal momentum transferred to the medium is a function of the emission angle and is defined as

$$q_{1\,\min}(\omega,v,\theta) = \frac{\omega}{v}\left(1-\frac{v}{c}\cos\theta\right). \qquad (12)$$

This formula shows that the momentum transferred to the medium for zero emission angles in the relativistic limit decreases in inverse proportion to the squared energy. Function $I\phi(g,v,\omega,\theta)$ has emerged as a result of integration with respect to azimuth angle $\phi_{\mathbf{g}}$ of vector

$$\mathbf{g} = \mathbf{q}_1 = \mathbf{q}+\mathbf{k}.$$

The polar angle ϑ of this vector is determined by the energy conservation law, while the condition $|\cos\vartheta|\le 1$

leads to the emergence of the Heaviside function in formula (11). The expression for $I\phi(g, v, \omega, \theta)$ has the form

$$I\phi(g, v, \omega, \theta) = \frac{AD - BE - CD}{(D^2 - E^2)^{3/2}} + \frac{CD}{E^2\sqrt{D^2 - E^2}} - \frac{C}{E^2}, \quad (13)$$

where

$$A = g^2 + \frac{\omega^2 v^2}{c^4}\sin^2\theta + 2\frac{\omega^2 v}{c^3}\cos\theta - 2\frac{\omega^2}{c^2} - g\cos^2\theta\frac{\omega}{vg} \times\left(\frac{v}{c}\cos\theta - 1\right)\left[\frac{\omega}{v}\left(\frac{v}{c}\cos\theta - 1\right) + 2\frac{\omega v}{c^2}\right], \quad (14)$$

$$B = g\frac{\omega}{v}\sin 2\theta\sqrt{1 - \left[\frac{\omega}{vg}\left(\frac{v}{c}\cos\theta - 1\right)\right]^2} \times\left[\frac{v^2}{c^2} + \left(\frac{v}{c}\cos\theta - 1\right)\right], \quad (15)$$

$$C = g^2\sin^2\theta\left(1 - \left[\frac{\omega}{vg}\left(\frac{v}{c}\cos\theta - 1\right)\right]^2\right), \quad (16)$$

$$D = g - 2\cos\theta\frac{\omega^2}{cvg}\left(\frac{v}{c}\cos\theta - 1\right), \quad (17)$$

$$E = 2\sin\theta\frac{\omega}{c}\sqrt{1 - \left[\frac{\omega}{vg}\left(\frac{v}{c}\cos\theta - 1\right)\right]^2}. \quad (18)$$

Formulas (10)–(18) describe PBS of a fast charged particle in the polycrystal for all velocities satisfying the Born approximation, including the nonrelativistic case.

3. RESULTS AND DISCUSSION

Calculations based on formulas (10)–(18) were carried out for the following approximation of the nondipole dynamic polarizability:

$$\alpha(\omega, q) \approx \alpha(\omega)F(q), \quad (19)$$

where $\alpha(\omega)$ is the complex dipole polarizability and $F(q)$ is the atomic form factor.

Since the radiation considered here is characterized by a high frequency $\omega \gg 1$ at. unit, we will henceforth neglect the difference between the polarizability and the form factor of a single atom and the corresponding quantities for an ion in the crystal lattice.

We calculated dipole polarizability $\alpha(\omega)$ using the data on the photoabsorption cross section borrowed from the site of the Berkley National Laboratory [14]. With the help of these data and the optical theorem, we determined the imaginary part of the dipole polarizability. The real part was reconstructed from the imaginary part using the Kramers–Kronig relation. The value of the photoabsorption cross section [14] was normalized to satisfy the sum rule. The real and imaginary parts of the polarizability of the silver atom calculated in this way and multiplied by the square of the frequency are presented in Fig. 1. The peaks on the frequency dependence of the imaginary part of the polarizability correspond to the values of the electron binding energy in the *N*, *M*, and *L* subshells of the silver atom. Figure 1 also shows that, in the high-frequency limit, the real part of the polarizability multiplied by the squared frequency and taken with the opposite sign attains its asymptotic value equal to the number of bound electrons in the atomic system of units.

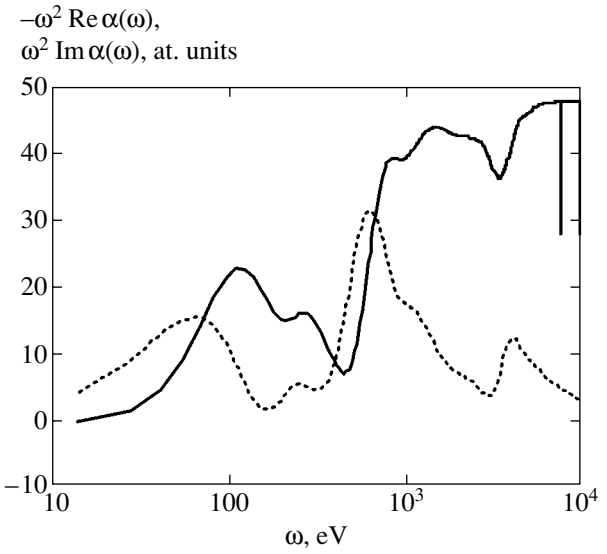


Fig. 1. Real (solid curve) and imaginary (dotted curve) parts of the dynamic polarizability of a silver atom, multiplied by squared frequency.

A comparison of the results of calculation of the complex polarizability of an atom, based on the method used here, with the results of calculation of the same quantity in the random-phase approximation with exchange [15] shows that the discrepancy typically does not exceed 10% in a wide frequency range.

The atomic form factor appearing on the right-hand side of equality (19) was calculated using the Slater wave functions using the formula proposed in [16], where it was shown that the difference between the form factor calculated in this way from the Hartree–Fock value does not exceed a few percent.

Figures 2–5 show the calculated differential yield of PBS photons generated by an ion with charge $Z_p = 30$ and velocity $v = c/3$ in polycrystalline silver and aluminum. We used the following values of mean standard deviation of lattice ions from their equilibrium positions (parameter u): $u_{Al} = 0.096$ Å and $u_{Ag} = 0.087$ Å. In the sum over reciprocal lattice vectors of the same magnitude, determining coherent PBS (11), we took into account 50 terms for which $S(\mathbf{g}) \neq 0$. It should be noted that the photon yield in the same spectral range changes by less than 1% when 40 terms are taken into account.

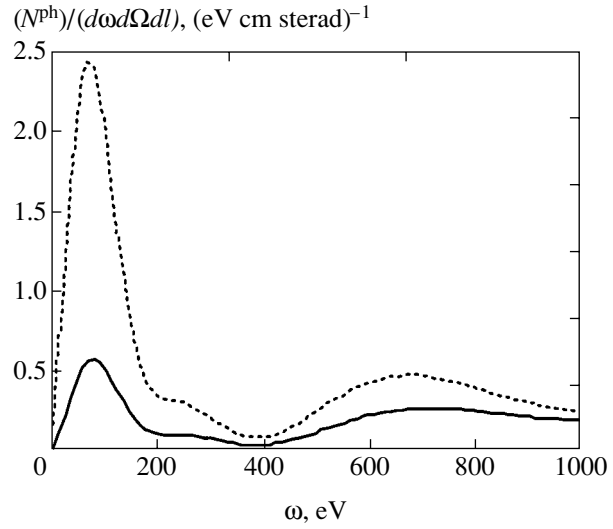


Fig. 2. PBS of an ion with charge $Z_p = 30$ in a polycrystal (solid curve) and at an atom (dotted curve) of silver in the low-frequency range.

Figure 2 shows the PBS photon yield at a silver polycrystal in the low-frequency range for an emission angle of 90°. The same figure shows the corresponding value for a single atom. It can be seen that PBS in the polycrystal is strongly suppressed as compared to the single atom in the given frequency range. This fact can be explained by the combined action of two factors. First, it can be seen from formula (10) that the incoherent term is small in the range of low frequencies $\omega < v/u$ and PBS is determined by the coherent component (11). Second, a contribution to coherent PBS in a polycrystal comes not from all momenta imparted to the target, but only from those with a magnitude equal to that of one of reciprocal lattice vectors. This leads to a decrease in the process intensity as compared to the intensity of radiation emitted by a single atom, when all momenta imparted to the target and allowed by the momentum conservation law contribute to the process. For example, for frequencies $\omega \ll vg$, modulo-small transferred momenta $\omega/v \le q < g$ make zero contribution to coherent PBS in a polycrystal, while precisely these momenta play an important role in the formation of PBS at a single atom.

The peaks on the frequency dependence in Fig. 2 are associated with the peaks of the imaginary part of the polarizability of the silver atom for photon energies close to the ionization potentials of the *N* and *M* shells.

Figure 3 shows PBS in a higher-frequency range, when the spectral dependence of the photon yield from a polycrystalline target acquires a characteristic structure in the from of "frequency steps." For preset values of the parameters of the problem (ion velocity and emission angle), three frequency steps can be seen in the frequency interval depicted in Fig. 3; the position of these steps is virtually the same for both types of the target. The reason for the emergence of such a spectral structure, which is not observed for a single atom, is associated with the coherent PBS component in a polycrystal, which dominates in the spectral range considered here. It can easily be seen from formula (11) that the frequency ω_j of the step associated with the reciprocal lattice vector **g** is defined as

$$\omega_j(g, v, \theta) = \frac{gv}{1-(v/c)\cos\theta}. \qquad (20)$$

For frequencies higher than ω_j, the contribution from a preset magnitude g of the reciprocal lattice vector to the probability of the process is zero since the energy-momentum conservation law is violated for it. This

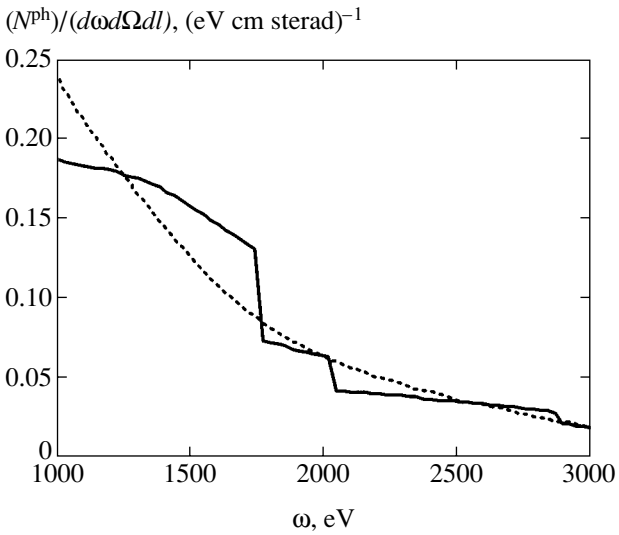


Fig. 3. The same as in Fig. 2 for a higher-frequency range.

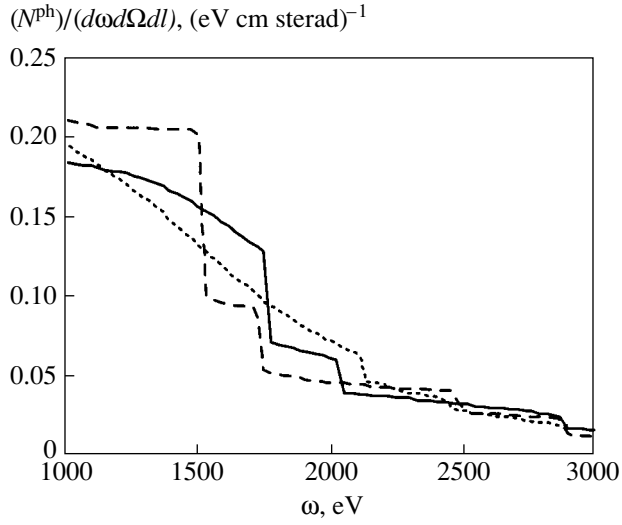


Fig. 4. Differential yield of PBS photons from an ion scattered in polycrystalline silver for various emission angles: 90° (solid curve), 60° (dotted curve) and 120° (dashed curve).

leads to the emergence of a frequency step on the spectral dependence of the PBS photon yield. Since the value of frequency ω_j is determined by the magnitude of vector $\mathbf{g}$ for which $S(\mathbf{g}) \neq 0$, the form of the PBS spectrum in a polycrystal depends on the crystal structure of the target. For example, for a diamond-type crystal lattice, the number of frequency steps is much smaller than for the face-centered lattice considered here. Indeed, in the case of the diamond lattice, an additional limitation is imposed on reciprocal lattice vectors for which the form factor of a unit cell differs from zero.

The "sharpness" of a frequency step depends on the relation between the coherent and incoherent contributions to PBS. If incoherent PBS prevails, the frequency step will be "blurred." To avoid this, the following condition must be satisfied:

$$g < \frac{1}{u}\left(1 - \frac{v}{c}\cos\theta\right). \tag{21}$$

It follows from the above inequality that the step structure in the PBS spectrum for a given magnitude of the reciprocal lattice vector has greater contrast for large emission angles θ. Indeed, with increasing θ, the minimal momentum (12) transferred to the target increases, while the contribution from the incoherent PBS component decreases.

Figure 4 shows the dependence of PBS in a silver polycrystal on the emission angle. It can be seen from the figure that the relative value of the frequency jump increases with the emission angle and its position is shifted to the region of lower frequencies in accordance with formulas (20) and (21).

Figure 5 shows the dependence of the PBS spectrum in an aluminum polycrystal on the velocity of the projectile. The solid curve is the emission spectrum for a high value of the ion velocity ($v = c/3$). With decreasing velocity, first, the contribution from the incoherent process increases; second, the position of frequency steps is shifted to the low-frequency region. Thus, the position of a frequency step in the PBS spectrum may serve as a measure of the energy of a scattering ion; i.e., the energy loss of the projectile can be judged from the displacement of ω_j. The dashed curve in Fig. 5 corresponds to an ion velocity equal to the velocity of 1-MeV protons used in experiments [6]. It can be seen that the PBS spectrum in this case does not exhibit the characteristic solid-state structure, but coincides with the emission spectrum for a single atom. This was precisely the case in experiments [6], in which no step structure of the spectrum was observed. This can be explained by the fact that, for low ion velocities ($v = c/20$ for the dashed curve in Fig. 5), the incoherent PBS component prevails over the coherent component even beginning with a photon energy of 500 eV or higher so that the step structure is completely hidden behind the incoherent background.

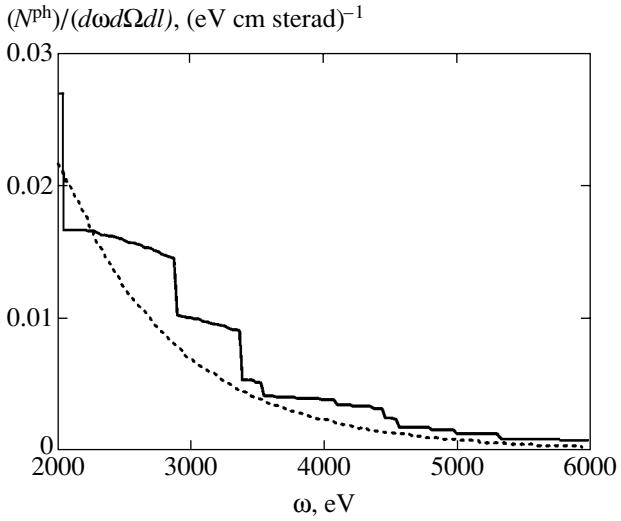


Fig. 5. Differential yield of PBS photons from an ion scattered in polycrystalline aluminum for various ion velocities: $v = c/3$ (the ordinate is magnified by a factor of 5) (solid curve) and $v = c/20$ (dotted curve).

In the high-frequency region of the PBS spectrum ($\omega > 10$ keV), which is characterized by high values of the momentum imparted to the target (or small values of the impact parameter), incoherent PBS plays a dominant role. For this reason, the solid-state structure of the spectrum associated with the coherent interaction of the projectile with the target becomes weakly distinguishable. As a result, the PBS spectrum in a polycrystal approaches the spectrum at a single atom, as should be expected in accordance with the physical pattern of the process. Thus, to observe frequency steps in the PBS spectrum at a polycrystal, ions with a sufficiently high energy should be used and observation should be carried out in the range of intermediate photon energies from 1.5–2 to approximately 6 keV.

4. CONCLUSIONS

We calculated and analyzed PBS of a fast ion in a polycrystalline medium taking into account coherent and incoherent emission channels. It is shown that in the range of low photon energies $\omega \ll 2\pi v/a$ (a is the lattice constant), in which the coherent component of the process dominates, PBS in the polycrystal is suppressed as compared to radiation emitted by a single atom. This suppression is due to the fact that small momenta transferred to the target and determining PBS at a single atom make zero contribution to coherent scattering of a fast ion from the crystal lattice in the low-frequency range.

In the intermediate photon energy range $\omega \geq 2\pi v/a$ (2–6 keV), a clearly distinguishable step structure (frequency steps) is observed in the PBS spectrum. This structure is associated with coherent scattering of a projectile from a crystal lattice, during which a momentum

modulo equal to one of reciprocal lattice vectors is transferred to the target. A frequency step is formed as a result of elimination of the contribution from one of the reciprocal vector moduli (g) to the process, when the minimal transferred momentum determined by conservation laws is modulo greater than g. The position and height of frequency steps reflect the features of the target structures and are functions of the velocity of the projectile particle and the photon emission angle.

In the high-frequency spectral range $\omega \gg 2\pi v/a$ ($\omega > 8$–10 keV), the incoherent component of the PBS spectrum dominates and the emission spectrum approaches the spectrum for a single atom.

These spectral features of PBS in polycrystalline targets may also be observed in experiments on emission from fast ions scattered in thin films, when the absorption of radiation in the target can be neglected as well as radiation emitted by secondary electrons.

Since the typical features of the PBS spectrum in polycrystals are determined by the structure of the target and depend on the velocity of the projectile, this radiation can be used for structural analysis of the target material and for determining bremsstrahlung of fast ions in a polycrystalline medium.

REFERENCES

1. I. E. Bakhmetjev, A. D. Fertman, A. A. Golubev, *et al.*, Laser Part. Beams **21**, 1 (2003).
2. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 5: *Statistical Physics*, 5th ed. (Fizmatlit, Moscow, 2003; Pergamon, New York, 1987).
3. *Polarized Bremsstrahlung from Particles and Atoms*, Ed. by V. N. Tsytovich and I. M. Oĭringel' (Nauka, Moscow, 1987) [in Russian].
4. V. A. Astapenko, L. A. Bureeva, and V. S. Lisitsa, Usp. Fiz. Nauk **172**, 155 (2002) [Phys. Usp. **45**, 149 (2002)].
5. V. M. Buĭmistrov, Yu. A. Krotov, and L. I. Trakhtenberg, Zh. Éksp. Teor. Fiz. **79**, 808 (1980) [Sov. Phys. JETP **52**, 411 (1980)].
6. K. Ishii and S. Morita, Phys. Rev. A **30**, 2278 (1984).
7. V. A. Astapenko, Zh. Éksp. Teor. Fiz. **99**, 165 (1991) [Sov. Phys. JETP **72**, 92 (1991)].
8. N. N. Nasonov, Nucl. Instrum. Methods Phys. Res. B **145**, 19 (1998).
9. S. Blazhevich, A. Chepurnov, V. Grishin, *et al.*, Phys. Lett. A **254**, 230 (1999).
10. V. A. Astapenko, L. A. Bureyeva, and V. S. Lisitsa, Laser Phys. **10**, 960 (2000).
11. V. A. Astapenko, V. M. Buĭmistrov, and Yu. A. Krotov, Zh. Éksp. Teor. Fiz. **93**, 825 (1987) [Sov. Phys. JETP **66**, 464 (1987)].
12. M. L. Ter-Mikaelyan, *High Energy Electromagnetic Processes in Condensed Media* (Akad. Nauk Arm. SSR, Yerevan, 1969; Wiley, New York, 1972).
13. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 8: *Electrodynamics of Continuous Media*, 4th ed. (Fizmatlit, Moscow, 2003; Pergamon, New York, 1984).
14. http://cindy.lbl.gov/optical_constants.
15. A. V. Korol', A. G. Lyalin, O. I. Obolenskiĭ, and A. V. Solov'ev, Zh. Éksp. Teor. Fiz. **114**, 458 (1998) [JETP **87**, 251 (1998)].
16. V. P. Shevelko, I. Yu. Tolstikhina, and Th. Stolker, Nucl. Instrum. Methods Phys. Res. B **184**, 295 (2001).

Translated by N. Wadhwa

Journal of Experimental and Theoretical Physics, Vol. 99, No. 5, 2004, pp. 965–977.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 5, 2004, pp. 1109–1122.

ATOMS, SPECTRA, RADIATION

Extraordinary Ray Refraction in a Large Pitch Helical Medium

E. V. Aksenova[a,*], A. Yu. Val'kov[b,], A. A. Karetnikov[a], A. P. Kovshik[a], V. P. Romanov[a], and E. I. Ryumtsev[a]**

[a] *St. Petersburg State University, Petrodvorets, St. Petersburg, 198504 Russia*
[b] *St. Petersburg Institute for Foreign Economic Relations, Economics, and Law, St. Petersburg, 191104 Russia*
*e-mail: *aksev@mail.ru; **alexvalk@mail.ru*

Received May 17, 2004

Abstract—The phenomenon of extraordinary ray refraction in a helical liquid crystal with large (compared to the light wavelength) pitch has been studied by theoretical and experimental methods. At a sufficiently large angle of incidence relative to the pitch axis, the extraordinary ray exhibits reflection (reversal) from a certain layer of the medium. The ordinary ray, for which the system is optically isotropic, exhibits no such reflection. The experimental dependences of the transmitted and reflected (reversed) rays are described using the geometrical optics approximation taking into account the optical losses for scattering inside the liquid crystal.

1. INTRODUCTION

Among the main distinctive features of liquid crystals (LCs) are their unusual optical properties, in particular—anisotropy (both uniaxial and biaxial), very high optical activity, selective reflection, unusually strong scattering of light, etc. [1]. From this standpoint, of considerable interest for optical applications are the LCs possessing one-dimensional periodicity, primarily cholesteric LCs. These LCs are locally uniaxial, with the optical axis uniformly rotating around the pitch axis so as to form a helical structure. In the case of normal incidence, whereby light propagates along the pitch axis, the wave equation admits exact solutions [1, 2]. For oblique incidence, systems with a cholesteric pitch P on the order of the light wavelength λ can be described using approximate methods [3] based on the Floquet theorem and the theory of diffraction. The validity of this approach was repeatedly confirmed in experiment.

Cholesteric LCs with a pitch greater then the light wavelength, as well as low-twist nematic LCs, have been studied to a much lower extent, although the first results for the normal incidence of light (the so-called Mouguin adiabatic regime) were obtained about a century ago [4]. In recent years, the lack of knowledge about such media has become especially challenging in view of the wide use of twist LC cells in data displays. The properties of helical LC media smoothly vary on a scale of λ, which makes it possible to use methods of the Wentzel–Kramers–Brillouin type. The problem of electromagnetic wave propagation in a locally isotropic medium with smooth inhomogeneities was considered in [5].

In recent years, we have studied the propagation of electromagnetic waves in the case of oblique incidence in cholesteric LCs with large pitch [6]. In this case, light also propagates in the adiabatic regime, whereby there are two normal waves—local ordinary and local extraordinary—with the polarization vectors determined by directions of the optical axis and the wave vector at a given point. The wave vector component perpendicular to the pitch axis is conserved, and the length of this vector (wavenumber) is determined from the local dispersion equation [6, 7].

In helical LCs, where the director orientation changes from point to point, the trajectories of rays exhibit a complicated character. Description of such systems encounters problems of two types. Problems of the first type are related to determination of the director distribution in an LC cell using the results of optical measurements. Such investigations were performed, for example, in [8]. In particular, the director field structure in hybrid cells of nematic LCs was reconstructed using the results of measurements of the reflected light intensity in the vicinity of a point of the total internal reflection. Problems of the second type are related to investigations of the light propagation and scattering in LC cells with known structures [3]. This paper deals with a problem of the second type.

We have studied propagation of the ordinary and extraordinary rays in twist LC cells with the helical structure period much greater than the light wavelength. It was established that the extraordinary ray propagating at a sufficiently large angle to the pitch axis exhibits reversal of the propagation direction inside the medium and leaves this medium similarly to the case of total internal reflection. However, in contrast to the lat-

ter effect, the extraordinary ray exhibits reflection in depth of the medium, rather than on the sample surface. We have studied this phenomenon theoretically and confirmed it experimentally. The experimental data have been analyzed in detail within the framework of the geometrical optics approximation.

This paper is organized as follows. Section 2 provides a theoretical description of the propagation of light in cholesteric LCs with large pitch. In particular, Section 2.1 presents general relations of the geometrical optics of such cholesteric LCs and Section 2.2 considers the trajectory shape and reversal of the extraordinary ray. In Section 3, we describe the experiment and present the experimental results. Section 4 is devoted to a comparative analysis of the theory and experiment with allowance for extinction in the cholesteric LC. In the Conclusions section, we will consider some other possible factors influencing the intensity of transmitted and reflected rays in the vicinity of the turning point. The Appendix gives formulas for the calculation of extinction in cholesteric LCs with large pitch.

2. THEORY OF THE LIGHT PROPAGATION IN TWIST LIQUID CRYSTAL CELLS WITH LARGE PITCH

2.1. Geometrical Optics Approximation

Consider a plane-parallel twist LC (cholesteric or nematic) cell with a pitch $P \gg \lambda$ and a thickness d, occupying a layer $0 \le z \le d$ in a Cartesian coordinate system xyz with the z axis perpendicular to the boundary planes. Let the pitch axis be oriented parallel to the z axis and the director vector $\mathbf{n}(\mathbf{r})$ be perpendicular to this direction and homogeneous in the xy plane:

$$\mathbf{n}(\mathbf{r}) \equiv \mathbf{n}(z) = (\cos(q_0 z + \phi_0), \sin(q_0 z + \phi_0), 0). \quad (2.1)$$

Here, $q_0 = \pi/P$ and ϕ_0 is the initial phase. According to formula (2.1), the period of the director $\mathbf{n}(z)$ is $2P$. However, since the directions $\mathbf{n}$ and $-\mathbf{n}$ in cholesteric and nematic LCs are equivalent, the actual pitch is P.

The wave equation for the electromagnetic field in such a medium can be written as

$$(\mathrm{curl}\,\mathrm{curl} - k_0^2 \hat{\varepsilon}(z))\mathbf{E}(\mathbf{r}) = 0, \quad (2.2)$$

where $\mathbf{E}(\mathbf{r})$ is the electric field vector, $k_0 = \omega/c$, ω is the circular frequency, c is the speed of light in vacuum, and $\hat{\varepsilon}(z)$ is the permittivity tensor. In a cholesteric (as well as in a planar twist nematic) LC, the latter tensor has the following form [1]:

$$\varepsilon_{\alpha\beta}(z) = \varepsilon_\perp \delta_{\alpha\beta} + \varepsilon_a n_\alpha(z) n_\beta(z), \quad (2.3)$$

where $\varepsilon_a = \varepsilon_\parallel - \varepsilon_\perp$; $\varepsilon_\parallel$ and $\varepsilon_\perp$ are the permittivity components parallel and perpendicular to $\mathbf{n}(z)$.

In view of the homogeneity of the system in the xy plane, it is convenient to pass to the Fourier harmonics $\mathbf{E}(\mathbf{k}_\perp, z)$ with respect to the transverse variables $\mathbf{r}_\perp = (x, y)$.

Let us fix the transverse wave vector $\mathbf{k}_\perp$ and seek a solution of the wave equation (2.2) inside the medium in the geometrical optics approximation:

$$\mathbf{E}(\mathbf{r}) = E_0(z)\mathbf{e}(z) \times \exp\left(i\mathbf{k}_\perp \cdot \mathbf{r}_\perp + i\int_0^z k_z(z')dz'\right), \quad (2.4)$$

where $E_0(z)$ and $k_z(z)$ are scalar functions and $\mathbf{e}(z)$ is the unit vector ($\mathbf{e} \cdot \mathbf{e}^* = 1$), which can be considered as local values of the amplitude, the z-components of the complete wave vector $\mathbf{k}(z) = (\mathbf{k}_\perp, k_z(z))$, and the vector of polarization of a quasi-plane wave.

The condition of applicability of the geometrical optics approximation (2.4) is represented by the set of inequalities,

$$\frac{dE_0(z)}{dz} \ll E_0(z)k_z(z), \quad \left|\frac{d\mathbf{e}(z)}{dz}\right| \ll k_z(z), \quad \frac{dk_z(z)}{dz} \ll k_z^2(z), \quad (2.5)$$

which must be obeyed for all z in the interval $0 \le z \le \min(P, d)$. The first two inequalities (2.5) imply smooth variation of the field amplitude and the polarization vector on a wavelength scale ($\lambda \sim k^{-1} \sim k_z^{-1}$), while the third inequality reflects the smooth variation of the $k_z(z)$ component of the wave vector $\mathbf{k}(z)$ on this scale. Note that, in view of the relation

$$\frac{dk_z(z)}{dz} \sim \frac{k_z(z)}{P},$$

the third inequality corresponds to the condition

$$k_z(z)P \gg 1. \quad (2.6)$$

According to formulas (2.5) and (2.6), the $k_z(z)$ component in the geometrical optics approximation is not very small for all z.

Substituting expression (2.4) into wave equation (2.2) and considering the main order with respect to the large parameter $k_z(z)P \sim P/\lambda$, we obtain the eikonal equation

$$\mathbf{k}(z) \times \mathbf{k}(z) \times \mathbf{e}(z) + k_0^2 \hat{\varepsilon}(z)\mathbf{e}(z) = 0. \quad (2.7)$$

At a fixed direction of the wave vector $\mathbf{t} = \mathbf{k}/k$ and a fixed value of z, the homogeneous system of linear equations (2.7) for a uniaxial permittivity tensor (2.3)

has two well-known solutions corresponding to the ordinary (o) and extraordinary (e) waves [9]:

$$k^{(o)} = k_0 n^{(o)}, \quad k^{(e)} = k_0 n^{(e)}. \tag{2.8}$$

Here, $n^{(o)}$ and $n^{(e)} = n^{(e)}(\mathbf{t})$ are the refractive indices for the ordinary and extraordinary rays, respectively:

$$n^{(o)} = \sqrt{\varepsilon_\perp}, \quad n^{(e)} = \sqrt{\frac{\varepsilon_\perp \varepsilon_\parallel}{\varepsilon_\parallel \cos^2\theta + \varepsilon_\perp \sin^2\theta}}, \tag{2.9}$$

where θ is the angle between vectors $\mathbf{n}$ and $\mathbf{k}$. The polarization vectors $\mathbf{e}^{(o)} = \mathbf{e}^{(o)}(\mathbf{t})$ and $\mathbf{e}^{(e)} = \mathbf{e}^{(e)}(\mathbf{t})$ are defined by the conditions

$$\mathbf{e}^{(o)} \perp \mathbf{n}, \quad \mathbf{e}^{(o)} \perp \mathbf{k}, \quad \hat{\varepsilon}\mathbf{e}^{(e)} \perp \mathbf{k}, \tag{2.10}$$

and by the fact that vector $\mathbf{e}^{(e)}$ occurs in the plane of vectors $\mathbf{k}$ and $\mathbf{n}$. Using Eqs. (2.10) and (2.3), we determine the directions of the unit polarization vectors $\mathbf{e}^{(o)}$ and $\mathbf{e}^{(e)}$:

$$\mathbf{e}^{(o)} \parallel \mathbf{k} \times \mathbf{n}, \quad \mathbf{e}^{(e)} \parallel \mathbf{n}(\mathbf{k}\hat{\varepsilon}\mathbf{k}) - \mathbf{k}(\mathbf{k}\hat{\varepsilon}\mathbf{n}). \tag{2.11}$$

In contrast to the usual formulation of the problem of wave propagation in a homogeneous anisotropic medium, where the direction of the wave vector $\mathbf{t}$ and the optical axis $\mathbf{n}$ do not change in space, our system has a fixed transverse component $\mathbf{k}_\perp$ of the wave vector, while the optical axis $\mathbf{n}(z)$ is rotating in space. In this case, the wave vectors $\mathbf{k}$ and the polarization vectors $\mathbf{e}$ are the functions of the coordinate z and the two-dimensional vector $\mathbf{k}_\perp$. For the ordinary wave, the first equation (2.8) yields

$$k_z^{(o)}(\mathbf{k}_\perp, z) \equiv k_z^{(o)}(k_\perp) = \sqrt{\varepsilon_\perp k_0^2 - k_\perp^2}. \tag{2.12}$$

For the extraordinary wave, we have

$$\cos\theta = \frac{\mathbf{n}(z) \cdot \mathbf{k}^{(e)}(\mathbf{k}_\perp, z)}{k^{(e)}(\mathbf{k}_\perp, z)} = \frac{\mathbf{n}(z) \cdot \mathbf{k}_\perp}{k^{(e)}(\mathbf{k}_\perp, z)}, \tag{2.13}$$

therefore, the second relation (2.8) is an equation for the $k_z^{(e)}(\mathbf{k}_\perp, z)$ component. Solving this equation, we obtain

$$k_z^{(e)}(\mathbf{k}_\perp, z) = \sqrt{\varepsilon_\parallel k_0^2 - k_\perp^2 - \frac{\varepsilon_a}{\varepsilon_\perp}(\mathbf{k}_\perp \cdot \mathbf{n}(z))^2}. \tag{2.14}$$

The polarization vectors $\mathbf{e}^{(o)}(\mathbf{k}_\perp, z)$, $\mathbf{e}^{(e)}(\mathbf{k}_\perp, z)$ can be determined using formulas (2.11) with $\mathbf{n} = \mathbf{n}(z)$ and $\mathbf{k} = \mathbf{k}^{(o)}(\mathbf{k}_\perp)$ or $\mathbf{k} = \mathbf{k}^{(e)}(\mathbf{k}_\perp, z)$, where $\mathbf{k}^{(o)} = (\mathbf{k}_\perp, k_z^{(o)}(k_\perp))$ and $\mathbf{k}^{(e)} = (\mathbf{k}_\perp, k_z^{(e)}(\mathbf{k}_\perp, z))$.

It should be noted that, within the framework of the geometrical optics approximation (2.4), the independence of the transverse wave vector component $\mathbf{k}_\perp$ of z is equivalent to Snell's law. Indeed, selecting the y axis orientation in the xyz system along the fixed vector $\mathbf{k}_\perp$, we can write $\mathbf{t}(z) = (0, \sin\chi(z), \cos\chi(z))$, where $\chi(z)$ is the z-dependent angle between the wave vector and the z axis. Therefore, $n^{(e)}(\mathbf{t}) = n^{(e)}(\mathbf{t}(z)) = n^{(e)}(z)$ and Eq. (2.8) yields $k_\perp = k_0 n^{(e)} \sin\chi(z)$ for the extraordinary wave. Thus, the condition that $\mathbf{k}_\perp$ is constant corresponds to the usual Snell's formula

$$n^{(e)}(z)\sin\chi(z) = \text{const}. \tag{2.15}$$

For the ordinary wave, this condition also corresponds to the usual Snell law, $n^{(o)}(z)\sin\chi(z) = \text{const}$. However, both quantities $n^{(o)}(z)$ and $\chi(z)$ are in fact independent of z (see relations (2.9) and (2.12)) and this law is trivial.

In order to find the field amplitude $E_0(z)$ in the geometrical optics approximation, it is necessary to retain terms of the next order with respect to the small parameter λ/P. The resulting "transfer equation" [10] for the field amplitude $E_0(z)$ is equivalent to the law of energy conservation, $\operatorname{div}\mathbf{S} = 0$, for the Poynting vector [9]:

$$\mathbf{S}(\mathbf{r}) = \frac{c}{8\pi k_0}[\mathbf{k}|E|^2 - \mathbf{E}(\mathbf{E}^* \cdot \mathbf{k})]. \tag{2.16}$$

For the waves satisfying Eq. (2.4), $\mathbf{S}(\mathbf{r}) \equiv \mathbf{S}(z)$ and

$$\mathbf{S}(z) = \frac{c|E_0(z)|^2}{8\pi k_0}[\mathbf{k}(z) - \mathbf{e}(z)(\mathbf{k}(z) \cdot \mathbf{e}(z))]. \tag{2.17}$$

Thus, in our case, the energy conservation law takes the form $\operatorname{div}\mathbf{S} = \partial_z S_z(z) = 0$, which implies that $S_z(z)$ is independent of z. As a result, relation (2.17) gives the law of variation of the field amplitude:

$$|E_0(z)|^2 = C_0 \frac{k_0}{k_z(z) - e_z(z)\mathbf{k}(z) \cdot \mathbf{e}(z)}, \tag{2.18}$$

where C_0 is an arbitrary constant. Substituting expressions (2.11)–(2.14) into (2.18), we eventually obtain

$$|E_0(z)|^2 = C_0 \begin{cases} \dfrac{k_0}{k_z^{(o)}(k_\perp)} & \text{for (o)-ray,} \\ \dfrac{k_0^2\varepsilon_\perp^2 + \varepsilon_a(\mathbf{k}_\perp \cdot \mathbf{n}(z))^2}{k_z^{(e)}(\mathbf{k}_\perp, z) k_0 \varepsilon_\perp^2} & \text{for (e)-ray.} \end{cases} \tag{2.19}$$

In concluding this Section, it can be noted that terms of the next order of smallness in λ/P in the wave equation yield a correction to the phase of the wave (2.4). The condition of neglect of this correction implies a restriction on the wave optical path in an inhomogeneous medium [10, 11]:

$$|z|\lambda \ll P^2, \tag{2.20}$$

which is an additional condition, supplementary to inequalities (2.5), determining the applicability of the geometrical optics approximation.

2.2. The Phenomenon of Extraordinary Ray Reversal

Let us consider the conditions under which the light wave can propagate in a medium with a preset value of $\mathbf{k}_\perp$. According to relation (2.12), the condition of the ordinary wave propagation appears as $\sqrt{\varepsilon_\perp}k_0 \geq k_\perp$. Since the wave vector $\mathbf{k}^{(o)}$ is independent of the coordinate z, the wave in this medium will exhibit rectilinear propagation in the entire region $0 \leq z \leq d$. According to relation (2.11), the polarization vector $\mathbf{e}^{(o)}(z)$ of this wave depends on z, since its direction is determined by the local director vector $\mathbf{n}(z)$. According to formula (2.19), the wave amplitude is constant.

The conditions of propagation of the extraordinary wave are significantly different because the longitudinal wave vector component $k_z^{(e)}$ depends on the coordinate z. According to Eq. (2.14), the extraordinary wave will propagate in the medium for all z, provided that vectors $\mathbf{k}_\perp$ and $\mathbf{n}(z)$ in the entire interval obey the inequality

$$\varepsilon_\parallel k_0^2 - k_\perp^2 > \frac{\varepsilon_a}{\varepsilon_\perp}k_\perp^2\cos^2\phi(z). \qquad (2.21)$$

This inequality corresponds to the condition $k_\perp^2 < \min(\varepsilon_\perp, \varepsilon_\parallel)k_0^2$.

On the contrary, if the reverse inequality is valid for all values of $\phi(z)$, which corresponds to the condition $k_\perp^2 > \max(\varepsilon_\perp, \varepsilon_\parallel)k_0^2$ the extraordinary wave will not propagate in the medium.

In the intermediate case of

$$\min(\varepsilon_\perp, \varepsilon_\parallel)k_0^2 \leq k_\perp^2 \leq \max(\varepsilon_\perp, \varepsilon_\parallel)k_0^2,$$

we observe a situation where inequality (2. 21) is valid only within a certain interval of angles $\phi(z)$. In this case, there exists an angle $\phi(z_t)$ for which the inequality in (2.21) changes to equality, so that the $k_z^{(e)}$ component vanishes. As the angle $\phi(z)$ decreases further, the $k_z^{(e)}$ component becomes imaginary and the extraordinary wave exhibits exponential decay. At the point z_t, the $k_z^{(e)}$ component changes sign and, accordingly, the wave changes the direction of propagation along the z axis. Therefore, z_t is the turning point for the extraordinary wave. In a certain sense, this effect is analogous to the phenomenon of total internal reflection from a certain plane in depth of the medium.

The polarization vector $\mathbf{e}^{(e)}(z)$ of the extraordinary wave is adiabatically adjusted to the local values of the wave vector and the director in accordance with formula (2.11).

In the vicinity of the point $z = z_t$, the inequalities in (2.5) are not satisfied and the geometrical optics approximation fails to be valid. Here, we occur in a caustic region where the field amplitude, according to Eq. (2.19), tends to infinity because $k_z^{(e)}(z) \longrightarrow 0$ as $z \longrightarrow z_t$. In order to describe the wave behavior in this region, it is necessary to analyze wave equation (2.2) for field (2.4) with $\hat{\varepsilon}(z)$ expanded into series in the vicinity of the turning point. For scalar waves, the phenomenon of wave reversal and the behavior of the wave field in the vicinity of this point have been described in much detail, for example, in [11]. This analysis showed that, in particular, after passage via the turning point, the field not only changes the sign of the $k_z^{(e)}$ component, but acquires an additional phase shift $\exp(-i\pi/2)$.

Now let us consider the shapes of trajectories of the ordinary and extraordinary rays in a cholesteric LC with large pitch. The ray trajectory is defined as the set of points at which the tangent coincides with the direction of the group velocity of the wave. In our case, it is convenient to replace the group velocity vector by the Poynting vector $\mathbf{S}$ having the same direction [9]. In an isotropic medium, $\mathbf{S} \parallel \mathbf{k}$ and the trajectory is usually described in terms of the wave vector. In an anisotropic medium, where (in the general case) $\mathbf{S} \nparallel \mathbf{k}$, the trajectory should be described in terms of the Poynting vector. Introducing parametrization of the ray trajectory as $(\mathbf{r}_\perp(z), z)$ and writing the condition that the tangent at a given point is parallel to the Poynting vector $(\mathbf{S}_\perp(z), S_z(z))$ of the wave, we obtain an equation for the ray trajectory $\mathbf{r}_\perp(z)$:

$$\frac{d\mathbf{r}_\perp(z)}{dz} = \frac{\mathbf{S}_\perp(z)}{S_z(z)}. \qquad (2.22)$$

For the ordinary ray, relations (2.10) show that $\mathbf{k}^{(o)}\mathbf{e}^{(o)}(z) = 0$, and Eq. (2.17) shows that $\mathbf{S}^{(o)} \parallel \mathbf{k}^{(o)}$ and $\mathbf{S}^{(o)}$ is independent of the coordinate z. Therefore, the right-hand side of Eq. (2.22) is constant and we conclude that the ordinary ray trajectory is rectilinear and coincides in direction with the wave vector $\mathbf{k}^{(o)}$.

For the extraordinary ray, the vector $\mathbf{S}^{(e)}(z)$ exhibits a more complicated trajectory because it is not parallel to the wave vector ($\mathbf{S}^{(e)}(z) \nparallel \mathbf{k}^{(e)}(z)$) and changes direction when z varies. In order to determine this trajectory, let us note that, in an arbitrary anisotropic medium, the vectors $\mathbf{k}$, $\mathbf{E}$, and $\mathbf{S}$ occur in the same plane so that $\mathbf{S} \perp \mathbf{E}$ and $\mathbf{k} \perp \mathbf{D} = \hat{\varepsilon}\mathbf{E}$; therefore, $\mathbf{k}\hat{\varepsilon}\mathbf{E} = 0$ [9]. Taking into account the symmetry of the permittivity tensor $\hat{\varepsilon}$, the

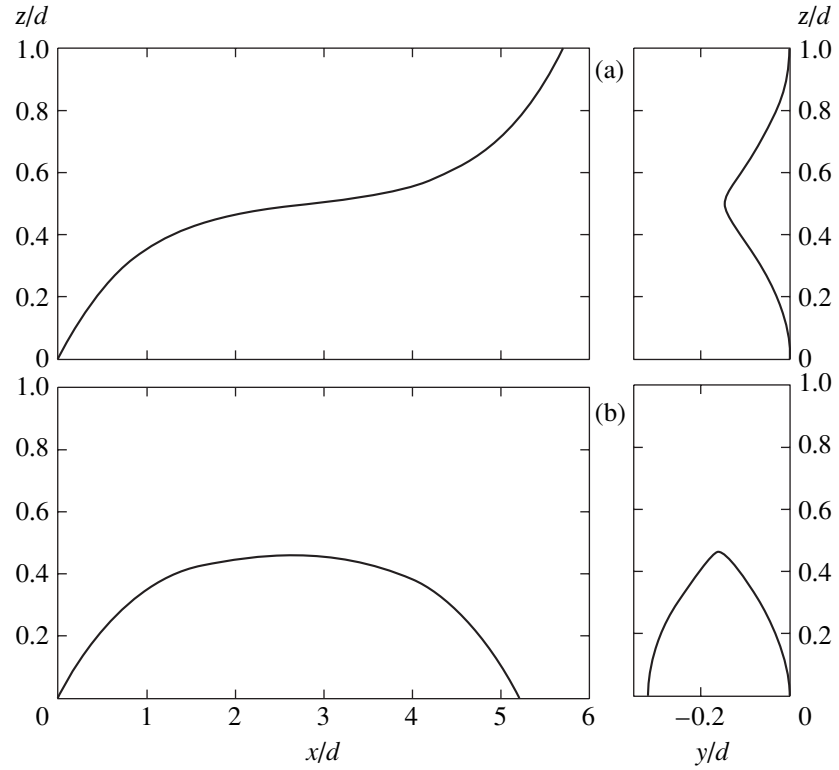


Fig. 1. Projections of the ray trajectories in an LC cell onto *xz* and *yz* planes for two angles of entrance χ into the LC: (a) extraordinary ray ($\chi = 63.2°$); (b) reverse extraordinary ray ($\chi = 63.5°$). Calculations were performed for $\sqrt{\varepsilon_\perp} = 1.51$, $\sqrt{\varepsilon_\parallel} = 1.69$, $\phi_0 = -\pi/2$ in formula (2.1); *x* axis is parallel to $\mathbf{k}_\perp$.

latter relation can be written as $\mathbf{e}\hat{\varepsilon}\mathbf{k} = 0$, so that $\mathbf{e} \perp \hat{\varepsilon}\mathbf{k}$. Formula (2.3) yields the relation

$$\hat{\varepsilon}\mathbf{k} = \varepsilon_\perp \mathbf{k} + \varepsilon_a(\mathbf{n}\cdot\mathbf{k})\mathbf{n}, \tag{2.23}$$

which implies that, in a uniaxial anisotropic medium, vector $\hat{\varepsilon}\mathbf{k}$ is a linear combination of the **n** and **k** vectors. Note also that, according to relation (2.11) for the extraordinary wave, the polarization vector $\mathbf{e}^{(e)}$ is also a linear combination of the **n** and **k** vectors. As a result, we conclude that the vectors $\mathbf{e}^{(e)}$, $\mathbf{S}^{(e)}$, and $\hat{\varepsilon}\mathbf{k}^{(e)}$ lie in the same plane formed by the vectors **n** and $\mathbf{k}^{(e)}$, so that $\mathbf{e} \perp \hat{\varepsilon}\mathbf{k}$ and $\mathbf{e} \perp \mathbf{S}$. Therefore, the uniaxial anisotropic media (to which the systems under consideration belong on a local level) are characterized by $\mathbf{S}^{(e)} \parallel \hat{\varepsilon}\mathbf{k}^{(e)}$.

The above considerations imply that

$$\frac{\mathbf{S}_\perp(z)}{S_z(z)} = \frac{(\hat{\varepsilon}(z)\mathbf{k}^{(e)}(z))_\perp}{(\hat{\varepsilon}(z)\mathbf{k}^{(e)}(z))_z}, \tag{2.24}$$

and trajectory equation (2.22) with allowance for (2.23) takes the following form:

$$\frac{d\mathbf{r}_\perp(z)}{dz} = \frac{\mathbf{n}(z)k_\perp\cos\phi(z)\varepsilon_a + \mathbf{k}_\perp\varepsilon_\perp}{k_z^{(e)}(z)\varepsilon_\perp}. \tag{2.25}$$

Integrating this relation, we obtain an explicit expression for the ray trajectory:

$$\mathbf{r}_\perp(z) = \frac{\varepsilon_a k_\perp}{\varepsilon_\perp}\int_0^z \frac{\mathbf{n}(z')\cos\phi(z')}{k_z^{(e)}(z')}dz' + \mathbf{k}_\perp\int_0^z \frac{dz'}{k_z^{(e)}(z')}. \tag{2.26}$$

The right-hand side of Eq. (2.26) reduces to elementary functions and an incomplete elliptic integral of the first kind. The results of numerical calculations of the ray trajectories are presented in Fig. 1. As can be seen, the extraordinary ray for certain angles of incidence exhibits reversal inside the crystal and turns back to the medium from which the light wave was incident onto the crystal. The trajectory of the extraordinary ray in such a medium is flat neither inside nor outside of the wave channel. However, the period-average Poynting vector occurs in the $(z, \mathbf{k}_\perp)$ plane.

3. EXPERIMENTAL

We have verified the effect of extraordinary ray reversal by measuring the parameters of rays outside a crystal and interpreting the results within the framework of the geometrical optics approximation.

The system studied comprised the mixture of a nematic LC (ZhKM1466, Institute of Organic Semiprod-

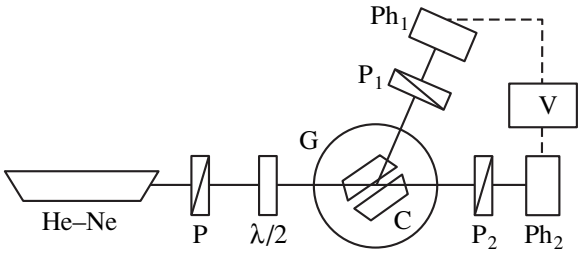


Fig. 2. Schematic diagram of the experimental setup: (He–Ne) helium–neon laser; (P, P_1, P_2) polarizers; ($\lambda/2$) half-wave plate; (Ph_1, Ph_2) photodetectors; (V) digital voltmeter V7-35; (G) goniometer; (C) optical cell with liquid crystal.

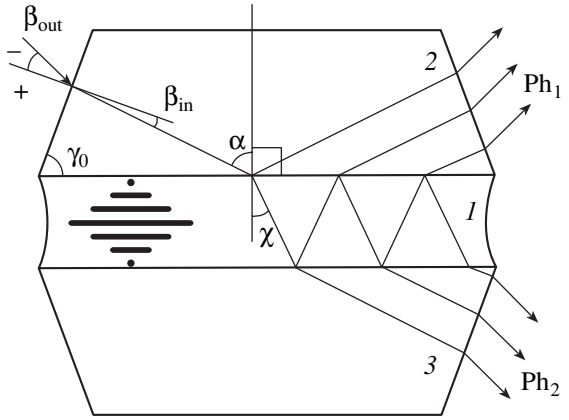
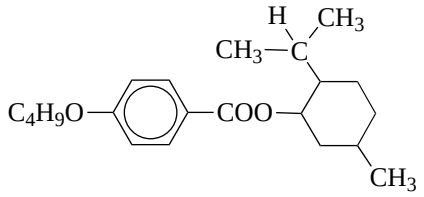


Fig. 3. Schematic diagram of the LC cell showing the ray paths: (*1*) planar twist LC; (*2*, *3*) glass prisms with a height of 12 mm, a larger base of 37 mm, and a base angle $\gamma_0 = 70°$; β_{out} and β_{in}, the angles of incidence and refraction on the prism entrance surface, respectively; α the angle of incidence onto the LC surface.

ucts and Dyes, Moscow) and a chiral additive. The latter was an optically active compound with the structural formula



synthesized at the Vilnius State University (Lithuania). The surfaces of prisms contacting with the LC medium were processed so as to ensure the planar orientation.

Figure 2 shows a schematic diagram of the experimental setup. The light source was a He–Ne laser operating at $\lambda = 632.8$ nm. After polarizer P, the linearly polarized laser beam passed through a half-wave plate (which allowed the polarization plane direction to be controlled) and struck the entrance face of a prism in cell C mounted on goniometer G. We measured the reflected and transmitted light intensities (I_1 and I_2) as functions of the angle of the laser beam incidence onto the LC for two polarizations of the incident light beam, which corresponded to the ordinary or extraordinary ray entering the LC. Using polarizers P_1 and P_2 placed in front of the photodetectors (Ph_1 and Ph_2, respectively), it was possible to separate the components having the same polarization as that in the incident light.

The measurements were performed with a special cell schematically depicted in Fig. 3, comprising a thin lane-parallel LC layer (*1*) sandwiched between two trapezoidal glass prisms (*2*, *3*) with the base angles $\gamma_0 = 70°$. This shape of the prism allowed the incident beam to be introduced at rather large angles α. The experiments were performed for two LC samples with $d = 8$ and 100 μm. The easy director orientation axes on both surfaces were perpendicular to the plane of the figure. The concentration of the optically active additive was selected so as to provide that the pitch P would be close to the LC layer thickness d.[1] As a result, the cell featured planar twist structure with a director orthogonal to the plane of the figure on the alignment surfaces and parallel to this plane at the center of the LC layer (thus, making one turn over the layer thickness).

The values of $\sqrt{\varepsilon_\perp}$ and $\sqrt{\varepsilon_\parallel}$ in the LC mixtures for $\lambda = 632.8$ nm were $\sqrt{\varepsilon_\perp} = 1.51$ and $\sqrt{\varepsilon_\parallel} = 1.69$, thus virtually coinciding with the corresponding values for the nematic matrix. This is explained by a very small concentration of the chiral additive (about 2 and 0.07% for the first and second sample, respectively). The refractive index of the glass prisms was $n_p = 1.644$. This value falls between $\sqrt{\varepsilon_\perp}$ and $\sqrt{\varepsilon_\parallel}$, being three times closer to $\sqrt{\varepsilon_\parallel}$ than to $\sqrt{\varepsilon_\perp}$.

The incident, transmitted, and reflected rays occur in the plane of Fig. 3. The angles of the light incidence onto the LC surface (α) and onto the prism entrance surface (β_{out}) are related as

$$\alpha = \gamma_0 \pm \beta_{in}, \quad n_p \sin\beta_{in} = \sin\beta_{out}, \qquad (3.1)$$

where the sign is selected as indicated in Fig. 3, depending on the position of the incident ray relative to the outer normal to the prism entrance surface. Relations (3.1) show that α can vary from $\gamma_0 - \arcsin(1/n_p) \approx 32.54°$ to 90°.

First, let us consider how the incident light intensity would be distributed between photodetectors Ph_1 and Ph_2 if the extraordinary ray reversal were not taking

[1] The pitch P was determined using two independent methods: first, by observation of the Grangin texture in a Canot wedge; second, by monochromatic light diffraction on a confocal structure.

place. According to Snall's law for the ordinary ray refraction at the glass–LC interface,

$$n_p \sin\alpha = n^{(o)} \sin\chi = \sqrt{\varepsilon_\perp}\sin\chi \le \sqrt{\varepsilon_\perp}. \qquad (3.2)$$

This relation poses limitation on the angle of incidence, $\alpha < \alpha_* = \arcsin(\sqrt{\varepsilon_\perp}/n_p)$ so that the ordinary ray does not enter the LC when $\alpha > \alpha_*$ Substituting the values of n_p and $\sqrt{\varepsilon_\perp}$, we obtain $\alpha_* \approx 66.7°$ for the total internal reflection angle of the ordinary ray at the glass–LC interface.

For the extraordinary ray, Snall's law with allowance for relation (2.9) yields

$$n_p \sin\alpha = n^{(e)}(\theta)\sin\chi = \sqrt{\varepsilon_\parallel}\sin\chi, \qquad (3.3)$$

because $\theta = 90°$ at the at the glass–LC interface (Fig. 3). When χ varies from 0 to 90°, the right-hand side in relation (3.3) changes from 0 to $\sqrt{\varepsilon_\parallel}$. Similarly, the left-hand side varies (depending on α) from 0 to n_p. By virtue of the condition $n_p < \sqrt{\varepsilon_\parallel}$, for any α there is an angle χ,

$$\chi = \arcsin\frac{n_p \sin\alpha}{\sqrt{\varepsilon_\parallel}}, \qquad (3.4)$$

such that relation (3.3) is satisfied. This implies that the extraordinary ray does not exhibit total internal reflection at the glass–LC interface in the entire range of angles α.

Therefore, it might be expected that the extraordinary ray would exhibit reflection and refraction at all angles of incidence and the output light intensity measured by both photodetectors would smoothly vary depending on α. The intensity of the transmitted light (I_2) must significantly exceed that of the reflected light (I_1). For the ordinary ray, it might be expected that both reflected and refracted (transmitted) signals would be observed for $\alpha < \alpha_*$, and only the reflected signal, for $\alpha > \alpha_*$.

The results of our measurements for the ordinary and extraordinary rays are presented in Fig. 4. As can be seen from these data, there exists an angle $\alpha_* \approx 66.7°$ (the same for both rays) such that the light reaches both photodetectors Ph_1 and Ph_2 for $\alpha < \alpha_*$, and only photodetector Ph_1 for $\alpha > \alpha_*$. The results presented in Fig. 4 refer to the sample with $d = 100$ μm. The pattern observed for the sample with $d = 8$ μm is qualitatively the same, but the signal intensity variations in the vicinity of $\alpha \lesssim \alpha_*$ are less pronounced. Thus, the experimental results for the ordinary ray are qualitatively consistent with the above predictions, whereas the data for the extraordinary ray disagree with these predictions.

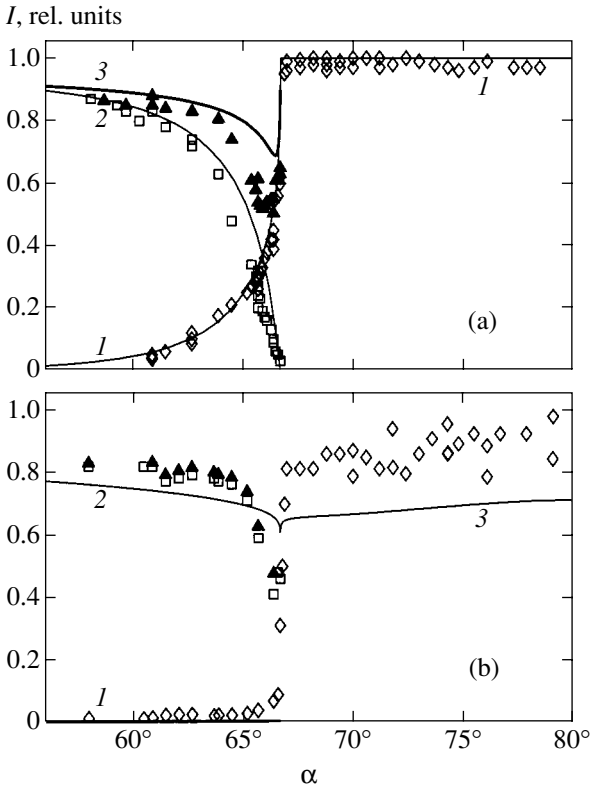


Fig. 4. Angular dependence of the reflected and transmitted light intensity for the (a) ordinary and (b) extraordinary rays. Experimental data: (◇) reflected intensity; (□) transmitted intensity; (▲) total intensity. Solid curves show theoretical curves: (*1*) reflected intensity I_1; (*2*) transmitted intensity I_2 calculated using formulas (4.4) (a) and (4.8) (b); (*3*) total intensity $I_1 + I_2$ (a); reflected intensity I_1 calculated using formula (4.9) (b). All values are normalized to the incident light intensity.

The discrepancy can be eliminated if we take into account the phenomenon of the extraordinary ray reversal described in Section 2.2. According to this, the extraordinary wave with a sufficiently large value of $k_\perp$ (i.e., at a sufficiently large angle of incidence α) cannot penetrate into the LC deeper than to a certain layer where $k_z = 0$. Upon reaching this depth, k_z changes sign and, in fact, this wave will escape from the LC layer with the same value of the transverse component $\mathbf{k}_\perp$ but with the opposite longitudinal component k_z. This will appear as if the wave would be reflected. As was demonstrated in Section 2.2, the extraordinary ray reversal for $\varepsilon_\perp < \varepsilon_\parallel$ takes place for waves with $k_\perp$ such that $\varepsilon_\perp k_0^2 < k_\perp^2 \le \varepsilon_\parallel k_0^2$. Taking into account the relation $k_\perp = k^{(e)}\sin\chi$ and Snall's law (3.3), we obtain

$$\frac{\sqrt{\varepsilon_\perp}}{n_p} < \sin\alpha \le \frac{\sqrt{\varepsilon_\parallel}}{n_p}.$$

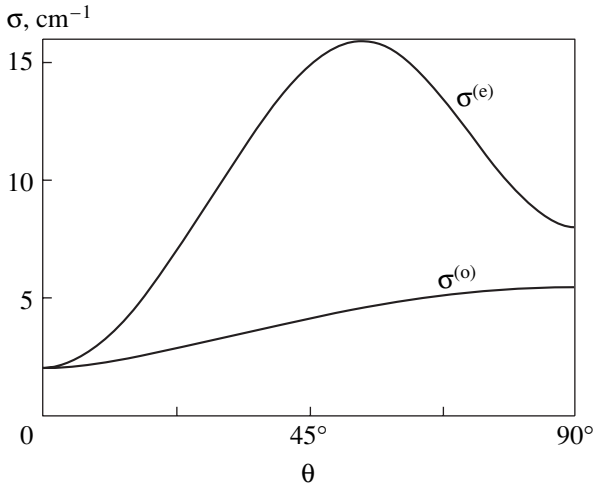


Fig. 5. Angular dependences of the extinction coefficients for the ordinary ($\sigma^{(o)}$) and extraordinary ($\sigma^{(e)}$) rays calculated using formula (A.2).

Since $\sqrt{\varepsilon_\parallel} > n_p$ and $\sqrt{\varepsilon_\perp}/n_p \equiv \sin\alpha_*$, we infer that the extraordinary ray reversal must be observed in the interval $\alpha_* < \alpha \le 90°$, in agreement with our experimental results.

The fact of coincidence of the limiting angles α_* for the ordinary and extraordinary rays can be explained as follows. If the refractive index (2.9) of the extraordinary ray decreases in the course of its propagation in the LC layer, while the $k_\perp$ component of the wave vector remains constant, the $k_z^{(e)}$ component along the ray tends to decrease. The minimum angle α corresponding to the onset of the extraordinary ray reversal is determined by the condition that $k_z^{(e)}$ is zero on a plane in depth of the LC, where the refractive index $n^{(e)}$ is minimum. In our case, $\varepsilon_\perp < \varepsilon_\parallel$ and, hence, the minimum value of $n^{(e)}$ is $\sqrt{\varepsilon_\perp}$, thus coinciding with the refractive index $n^{(o)}$ for the ordinary ray. Therefore, the minimum angle α corresponding to the onset of the extraordinary ray reversal exactly coincides with the angle of total internal reflectance α_* of the ordinary ray.

4. THE INFLUENCE OF EXTINCTION

Now we will analyze the angular dependence of intensities $I_{1,2}$ for the rays of both types in more detail. First, let us consider the case of incidence of the ordinary ray. The first factor determining a smooth decrease in intensity of the transmitted ray (I_2) and an increase in intensity of the reflected ray (I_1) at $\alpha \longrightarrow \alpha_*$ in the region of $\alpha < \alpha_*$ is related to a redistribution of the energy between the rays refracted and reflected at the LC–glass interface. The intensities of the transmitted (I_{tr}) and reflected (I_{rf}) rays are described by the Fresnel formulas [9],

$$I_{rf} = I_0 R_\parallel, \quad I_{tr} = I_0(1 - R_\parallel), \tag{4.1}$$

where I_0 is the incident light intensity and

$$R_\parallel = \frac{\tan^2(\chi - \alpha)}{\tan^2(\chi + \alpha)} \tag{4.2}$$

is the reflection coefficient for the ray polarized in the plane of incidence.

As α approaches the critical value α_*, the reflection coefficient increases and tends to unity in the limit of $\alpha \longrightarrow \alpha_*$. Therefore, in the vicinity of α_*, it is necessary to take into account multiple reflections at both LC–glass interfaces (Fig. 3).

It should be also noted that, for $\alpha \longrightarrow \alpha_*$, relation (3.2) yields $\chi \longrightarrow 90°$ and, hence, the refracted ray propagates in the LC at $\alpha \approx \alpha_*$ at almost glancing angles. As a result, the optical path

$$l^{(o)} = d\frac{\sin\alpha_*}{\sqrt{\sin^2\alpha_* - \sin^2\alpha}} \tag{4.3}$$

traveled by the ray between two sequential reflections in the LC can be rather large so that the losses of light scattered in the LC (extinction) may become significant. The general formulas for the extinction coefficients in a twist LC cell with large pitch are presented in the Appendix. We used formulas (A.2) and (A.4) to calculate the angular dependence of the extinction coefficient for the ordinary ($\sigma^{(o)}(\theta)$) and extraordinary ($\sigma^{(e)}(\theta)$) rays in the LC under consideration. The results of these calculations are presented in Fig. 5. For a nematic matrix of this LC, the ratio of the Frank moduli is $K_{33}/K_{11} \approx 0.95$ and $K_{11} \sim 10^{-6}$ dyn. For this reason we have calculated estimates for a single-constant approximation with $K_{11} = K_{22} = K_{33} = 10^{-6}$ dyn. The interval of integration in (A.2) was truncated at a scattering angle of 0.7° corresponding to the detector aperture. It should be noted that the maximum values of $\sigma^{(e)}$ (for $\theta \approx 45°$) are several times larger than the characteristic values of $\sigma^{(o)}$ (cf. [12, 13]).

Thus, taking into account the multiple reflections and extinction, we obtain for the ordinary ray

$$I_1 = I_0 R_\parallel\left[1 + \frac{(a_{(o)} - \gamma_{(o)})^2}{1 - \gamma_{(o)}^2}(1 - \gamma_{(o)}^{2M_{(o)}+2})\right],$$
$$I_2 = \frac{I_0}{a_{(o)}}\frac{(a_{(o)} - \gamma_{(o)})^2}{1 - \gamma_{(o)}^2}(1 - \gamma_{(o)}^{2M_{(o)}+2}) \tag{4.4}$$

for $\alpha < \alpha_*$, and

$$I_1 = I_0, \quad I_2 = 0 \tag{4.5}$$

for $\alpha > \alpha_*$. Here, $\gamma_{(o)} = a_{(o)}R_\parallel$,

$$a_{(o)} = \exp\left(-\frac{l^{(o)}}{d}\int_0^d \sigma^{(o)}(\theta(z))dz\right), \tag{4.6}$$

$\sigma^{(o)}(\theta)$ is the extinction coefficient for the ordinary ray, $\theta = \theta(z)$ is the angle between the director $\mathbf{n}(z)$ and the wave vector $\mathbf{k}^{(o)}$,

$$\cos\theta(z) = \frac{\mathbf{n}(z)\cdot\mathbf{k}_\perp}{k_0\sqrt{\varepsilon_\perp}} = \cos(q_0 z + \phi_0)\frac{\sin\alpha}{\sin\alpha_*},$$

and

$$M_{(o)} \approx 1 + \frac{D\sqrt{\sin^2\alpha_* - \sin^2\alpha}}{d\sin 2\alpha}$$

is the number of reflections of the ordinary ray inside the sample within the aperture $D \gg d$ (in our experiments, $D = 4$ mm). In the geometry under consideration, $\phi_0 = -\pi/2$ and $\mathbf{k}_\perp = (k_\perp, 0, 0)$; we also took into account that the ordinary ray has a rectilinear trajectory, and an increment of the trajectory length in (A.1) is $dl = (l^{(o)}/d)dz$.

In Fig. 4a, solid curves show the results of calculations using formulas (4.4) and (4.5) with the extinction coefficient $\sigma^{(o)}$ presented in Fig. 5. It should be noted that the curves $I_{1,2}(\alpha)$ are very sensitive to this extinction coefficient. Therefore, the agreement of theory and experiment reached in Fig. 4a without any fitting parameters can be considered as quite satisfactory. Note that the sum of intensities of the reflected and transmitted rays for $\alpha < \alpha_*$ is smaller than the intensity of the ray reflected for $\alpha > \alpha_*$. The difference is especially pronounced for $\alpha_* - \alpha \sim 1$–$2°$. This behavior is related to considerable losses for scattering in the LC for α values close to the total internal refection angle (when the ordinary ray travels over a large distance inside the crystal).

Now let us consider the case of incidence of the extraordinary ray. Note certain differences between Figs. 4b and 4a. First, in the region where α is smaller than α_* and sufficiently far from the critical value, the transmitted intensity I_2 is smaller for the extraordinary ray than for the ordinary one: for example, $I_2^{(e)}(56°) \approx 0.8I_0$ and $I_2^{(o)}(56°) \approx 0.9I_0$. Second, the total intensity $I_1 + I_2$ for the extraordinary ray at $\alpha < \alpha_*$ in a rather wide region ($\alpha_* - \alpha \gtrsim 1°$) is close to the value of I_1 for $\alpha > \alpha_*$, whereas for the ordinary ray, the intensity I_1 at $\alpha > \alpha_*$ is significantly greater than $I_1 + I_2$ at $\alpha < \alpha_*$. Finally, the region of $\alpha < \alpha_*$ in Fig. 4b where I_1 and I_2 significantly vary when $\alpha \longrightarrow \alpha_*$ is much narrower than the analogous region for the ordinary ray in Fig. 4a. This region in Fig. 4b features a weak growth in the reflected intensity I_1 and a significant decrease in the transmitted intensity I_2 for the extraordinary ray, in contrast to the case of the ordinary ray (for which the variations of I_1 and I_2 are of the same order of magnitude).

In order to explain these distinctions, it should be recalled that the refractive index for the extraordinary ray ($\sqrt{\varepsilon_\parallel}$) in our system is almost three times closer than the value for the ordinary ray ($\sqrt{\varepsilon_\perp}$) to the value for the glass prisms (n_p). From this we infer that the coefficient of reflection for the extraordinary ray polarized perpendicularly to the plane of incidence [9],[2]

$$R_\perp = \frac{\sin^2(\chi - \alpha)}{\sin^2(\chi + \alpha)}, \tag{4.7}$$

for χ given by formula (3.4) is about ten times smaller than the coefficient of reflection for the ordinary ray ($R_\parallel$). Therefore, the main factor significantly influencing the values of I_1 and I_2 for the extraordinary ray in this case is extinction, whereas multiple reflections at the LC–glass and glass–LC interfaces play a much less significant role. This circumstance explains the absence of significant angular dependence of the reflected intensity I_1 for the ordinary beam in the region of $\alpha < \alpha_*$.

Thus, the intensities of transmitted and reflected extraordinary rays are given by the expressions

$$I_1 = I_0R_\perp\left[1 + \frac{(a_{1(e)} - \gamma_{1(e)})^2}{1 - \gamma_{1(e)}^2}(1 - \gamma_{1(e)}^{2M_{1(e)}+2})\right],$$
$$I_2 = \frac{I_0}{a_{1(e)}}\frac{(a_{1(e)} - \gamma_{1(e)})^2}{1 - \gamma_{1(e)}^2}(1 - \gamma_{1(e)}^{2M_{1(e)}+2}) \tag{4.8}$$

for $\alpha < \alpha_*$ and

$$I_1 = I_0R_\perp + I_0\frac{a_{2(e)} - \gamma_{2(e)}}{1 - \gamma_{2(e)}}(1 - \gamma_{2(e)}^{M_{2(e)}+1}),$$
$$I_2 = 0 \tag{4.9}$$

[2] These estimates are obtained using the formula for an isotropic medium.

for $\alpha > \alpha_*$. Here, $\gamma_{j(e)} = a_{j(e)}R_\perp$ ($j = 1, 2$),

$$a_{1(e)} = \exp\left(-\int_{z=0}^{z=d} \sigma^{(e)}(z)dl^{(e)}(z)\right),$$
$$a_{2(e)} = \exp\left(-2\int_{z=0}^{z=z_t} \sigma^{(e)}(z)dl^{(e)}(z)\right), \quad (4.10)$$

$\sigma^{(e)}(\theta)$ is the extinction coefficient for the extraordinary ray, $\theta(z)$ is the angle given by the formula (2.13),

$$dl^{(e)}(z) = \sqrt{1 + \left|\frac{d\mathbf{r}_\perp(z)}{dz}\right|^2}dz \quad (4.11)$$

is the element of optical path traveled by the extraordinary ray in the LC according to formula (2.25), $M_{j(e)}$ is the number of reflections of the extraordinary ray inside the sample for $\alpha < \alpha_*$ ($j = 1$) and $\alpha > \alpha_*$ ($j = 2$). In the case under consideration, $R_\perp \ll 1$ and, hence $\gamma_{j(e)} \ll 1$, so that we may put $M_{j(e)} = \infty$ in formulas (4.8) and (4.9).

In Fig. 4b, the solid curve shows the results of calculations using formulas (4.8) and (4.9) with the extinction coefficient $\sigma^{(e)}$ presented in Fig. 5. Here, it should be also emphasized that the curves $I_2(\alpha)$ at $\alpha < \alpha_*$ and $I_1(\alpha)$ at $\alpha > \alpha_*$ are very sensitive to the extinction coefficient $\sigma^{(e)}$. Note that the coefficient $\sigma^{(e)}$ was calculated with cutoff of the integral (A.2) in the region of scattering angles below 0.7° (which corresponded to the condition $q \gtrsim 4q_0$ in (A.4)). At the same time, formula (A.4) is valid in the region of $q \gg q_0$. In the region of $q \lesssim q_0$, fluctuations of the director (A.4) exhibit smecticlike [14], rather than nematiclike behavior. In thin LC layers, the interaction of LC molecules with the alignment surfaces becomes more significant, which leads to the suppression of fluctuations near the interfaces and, hence, decreases the extinction coefficient. For this reason, the values of $\sigma^{(e)}$ in Fig. 5, which are more sensitive than $\sigma^{(o)}$ with respect to the scattering at small angles, should be treated as estimates. This probably accounts for a greater difference between theory and experiment for the extraordinary ray (Fig. 4b) than for the ordinary ray (Fig. 4a).

Now we can explain the behavior of I_1 and I_2 in the regions where the angle α is far from α_*. The lower values of I_2 for the (e)-ray than for the (o)-ray are related to the fact that the integrals in (4.8) and (4.9) contain significant contributions from the regions of z where $\sigma^{(e)}$ is large, that is, where $\theta(z)$ in (2.13) is close to 45°. As a result, even relatively short optical paths of the ray at α far away from α_*, the exponential factors in (4.10) significantly differ from unity. The regions of $\theta(z) \approx 45°$ correspond to two relatively narrow regions in the intervals (0, $P/2$) and ($P/2$, P). The transmitted ray crosses one of these regions producing a significant contribution to the integral at $0 < z < P/2$, and another, at $P/2 < z < P$. The extraordinary ray exhibiting a reversal passes twice through one of these regions ($0 < z < P/2$) when propagating in the forward and reverse directions. For this reason, the values of I_2 at $\alpha < \alpha_*$ and I_1 at $\alpha > \alpha_*$ are approximately equal.

In the close vicinity of α_*, the length of the trajectory of the extraordinary ray begins to increase and the integrals in (4.8) and (4.9) acquire significant contributions from the regions with small $\sigma^{(e)}$ and large length. The width $\alpha - \alpha_*$ of the region where this contribution becomes predominant is significantly smaller for the (e)-ray than for the (o)-ray, which is caused by two factors. First, the length $l^{(e)}$ of the extraordinary ray trajectory is shorter than that ($l^{(o)}$) of the ordinary ray. Indeed, as $\alpha - \alpha_* \longrightarrow 0$, $l^{(e)}$ increases logarithmically, while $l^{(o)} \sim (\alpha - \alpha_*)^{-1/2}$ according to (2.4). Second, the longest part of the trajectory in the vicinity of the turning point, $z_t(\alpha_*) = P/2$ corresponds to a nearly zero angle θ between the director $\mathbf{n}(z) \approx \mathbf{n}(P/2)$ and the wave vector $\mathbf{k}^{(e)} \approx \mathbf{k}_\perp$. However, as can be seen from Fig. 5, the value $\sigma^{(e)}(0) \equiv \sigma^{(o)}(0)$ corresponds to the minimum of $\sigma^{(o)}(\theta)$, so that the values of the integrand in (4.8) and (4.9) on the longest part of the extraordinary ray trajectory are smaller than the analogous values for the ordinary ray in (4.6).

5. CONCLUSIONS

Now let us consider some additional factors influencing the intensities of transmitted and reflected extraordinary rays in the close vicinity of α_*.

(i) As is known, the extraordinary ray in nematic LCs exhibits anomalous scattering by small angles into extraordinary rays [1]. For this reason, the extraordinary ray exhibits multiple scattering over a length on the order of $(\sigma^{(e)})^{-1} \ll (\sigma^{(o)})^{-1}$ and transforms from coherent into diffuse with a small angular expansion and retained polarization [15]. Since the system under consideration is locally close to nematic LCs, we may expect that an analogous effect can take place in our system as well. According to experimental data [15], the characteristic angular size of a diffuse beam was 1°–1.5° at a ray optical path in the LC on the order of 0.1 cm. Thus, significant angular expansion begins approximately in the same region of α where I_2 considerably changes due to losses due the scattering by large angles. As a result, the beam at $\alpha_* - \alpha \sim 1°–1.5°$ contains a considerable proportion of extraordinary rays exhibiting reversal, which appear as reflected when leaving the LC. For this reason, the region of $\alpha_* - \alpha \sim 1°–1.5°$ features an increase ion I_1, which becomes more pronounced as $\alpha_* - \alpha$ approaches zero.

(ii) For angles $\alpha > \alpha_*$ but still sufficiently close to α_*, the angular dependence of $I_{1,2}$ may be influenced by the so-called permeation effect [11, 16], whereby a wave partly penetrates through the forbidden region, where it exponentially decays in the z direction, to reach the adjacent allowed region. This leads to a decrease in the reflected intensity I_1 (a part of the energy leaks through the forbidden region) and an increase in the transmitted intensity I_2 (more energy penetrates through the forbidden region). The smaller the width of the forbidden region, the more pronounced is the permeation phenomenon. In our case, the forbidden region corresponds to $z_t < z < P - z_t$, and its width $P - 2z_t > \alpha_*$ tends to zero when $\alpha \longrightarrow \alpha_*$. The characteristic interval of angles for which the permeation is significant, $\Delta\alpha_{per} = \alpha - \alpha_*$, is determined by the condition

$$\int_{z_t}^{P/2} \mathrm{Im} k_z^{(e)}(z)dz \lesssim 1. \qquad (5.1)$$

Using the condition $k_z^{(e)}(\mathbf{k}_\perp, z_t) = 0$, it is possible to determine z_t as the function of α:

$$z_t(\alpha) = q_0^{-1} \arcsin \sqrt{\frac{\varepsilon_\perp}{\varepsilon_a}\left(\frac{\varepsilon_\parallel}{n_p^2 \sin^2\alpha} - 1\right)}.$$

In particular, $z_t(\alpha_*) = P/2$. This yields, for $\Delta z_t(\alpha) = z_t(\alpha_*) - z_t(\alpha)$ in the limit as $\Delta\alpha = \alpha - \alpha_* \longrightarrow 0$,

$$\Delta z_t(\alpha) \approx \mu q_0^{-1} (\Delta\alpha)^{1/2},$$

where

$$\mu = 2^{1/2} \varepsilon_\parallel^{1/2} \varepsilon_\perp^{-1/4} \varepsilon_a^{-1/2} (n_p^2 - \varepsilon_\perp)^{1/4}.$$

In the vicinity of the point $z_t = z_t(\alpha)$ in the forbidden region, the longitudinal component of the wave vector (2.14) has the following form:

$$k_z^{(e)}(\mathbf{k}_\perp, z) \approx ik_0\sqrt{\varepsilon_a} \times \sqrt{2\mu(\Delta\alpha)^{1/2} q_0 (z - z_t) - q_0^2 (z - z_t)^2}. \qquad (5.2)$$

Evaluating the integral in (5.1), we obtain

$$\frac{\pi k_0}{4q_0} \mu^2 \sqrt{\varepsilon_a} \Delta\alpha \lesssim 1,$$

which yields

$$\Delta\alpha_{per} \sim \frac{\lambda}{\pi\varepsilon_\parallel P} \sqrt{\frac{\varepsilon_\perp \varepsilon_a}{n_p^2 - \varepsilon_\perp}}.$$

Using the experimentally determined values of n_p, $\varepsilon_\perp$, and $\varepsilon_\parallel$, we obtain $\Delta\alpha_{per} \sim 0.2\lambda/P$. For a 100-μm-thick LC sample, this yields $\Delta\alpha_{per} \sim 0.07°$, while for $d = 8$ μm, $\Delta\alpha_{per} \sim 0.9°$.

(iii) For $\alpha = \alpha_*$, the turning point is at $z_t(\alpha_*) = P/2$. The components of the extraordinary wave vector $\mathbf{k}^{(e)}$ at this point are $k_\perp^{(e)} = k_0\sqrt{\varepsilon_\perp}$, $k_z^{(e)} = 0$. For the ordinary wave, $\mathbf{k}_\perp^{(o)} = \mathbf{k}_\perp = \mathbf{k}_\perp^{(e)}$. According to (2.12), in this case we have $k_z^{(o)}(k_\perp) = 0$ and $\mathbf{k}^{(e)} = \mathbf{k}^{(o)}$. At the turning point, $\mathbf{n}(P/2) \parallel \mathbf{k}^{(o,e)}$ and the modes are degenerate, whereby the difference between ordinary and extraordinary waves disappears. This situation may lead to the phenomenon of mode transformation [10] between the ordinary and extraordinary rays. As a result, in the vicinity of the turning point at $\alpha \approx \alpha_*$, the energy is partly transferred from extraordinary to ordinary ray and the latter propagates in a narrow vicinity of the plane $z = d/2$, not escaping from the outer surfaces of the prisms. The characteristic interval of angles $\Delta\alpha_{int}$ in which the interconversion of two rays may take place is determined by the condition

$$\left|\mathbf{k}^{(o)} - \mathbf{k}^{(e)}\right| \lesssim q_0 = \pi/P.$$

Taking into account that, in the vicinity of the turning point, $k_z^{(e)} \ll k_\perp$, we can write this condition as

$$\left|\mathbf{k}_\perp^{(o)}(\alpha_*) - \mathbf{k}_\perp^{(e)}(\alpha)\right| \lesssim q_0,$$

or as

$$k_0 n_p \cos\alpha_* \Delta\alpha = k_0\sqrt{n_p^2 - \varepsilon_\perp}\Delta\alpha \lesssim q_0.$$

This yields an estimate of $\Delta\alpha_{int} \lesssim \lambda/P$ or $\Delta\alpha_{int} \sim 0.4°$ for $d = 100$ μm and $\Delta\alpha_{int} \sim 4.5°$ for $d = 8$ μm. It should be emphasized that this estimate only indicates the interval of angles where the interconversion of extraordinary into ordinary ray may take place, rather than giving the absolute values of such transformation. Calculations of the latter values requires solving the wave equation in the vicinity of the point $\alpha = \alpha_*$ with allowance for he interaction of modes, which is a nontrivial problem.

In conclusion, it should be noted that the observed phenomenon of refraction of the extraordinary ray has analogs in the propagation of waves in the media with smoothly variable optical and acoustical characteristics. These phenomena include tropospheric refraction [16], the formation of a submerged waveguide channel in the ocean [11], and some effects accompanying the propagation of seismic waves in the Earth crust.

The investigations of refraction in LC cells are of independent interest for the optics of LCs. At the same time the cells of chiral LCs offer a convenient model system for studying fine effects such as the wave propagation inside a waveguide channel; permeation of

waves from one to another channel, interaction of modes in the vicinity of caustics, etc.

ACKNOWLEDGMENTS

The authors are grateful to V. A. Belyakov for fruitful discussions and useful remarks.

This study was supported in part by the Russian Foundation for Basic Research (project. nos. 02-02-16577, 03-02-16173, and 03-03-32823) and the Ministry of Education of the Russian Federation (project no. PD02-1.2-297).

Appendix

Calculation of the extinction coefficient for a twist cell with large pitch. When a light ray propagates in a fluctuating medium, the energy is partly lost as a result of scattering. These losses are described by the extinction coefficient σ, which coincides with the total scattering cross section representing the intensity of scattered radiation per unit volume and per unit incident intensity, integrated over the entire range of scattering angles.

With allowance for the extinction, the intensity $I(l)$ of a ray traveling over a distance l in a medium is determined by the formula

$$I(l) = I(0)\exp\left(-\int_0^l \sigma(l)dl\right), \tag{A.1}$$

where dl is the element of the trajectory length.

In a homogeneous anisotropic medium, there are two extinction coefficients, $\sigma^{(o)}$ and $\sigma^{(e)}$. In the Born approximation, these coefficients are as follows [12, 13, 17]:

$$\sigma^{(i)} = \frac{k_0^4}{16\pi^2}\frac{e_\alpha^{(i)}e_\beta^{(i)}}{n^{(i)}\cos\delta^{(i)}} \times \sum_{s=1,2}\int d\Omega_{\mathbf{k}^{(s)}}\frac{n^{(s)}e_\mu^{(s)}e_\nu^{(s)}}{\cos^2\delta^{(s)}}G_{\alpha\mu\beta\nu}, \tag{A.2}$$

where the superscripts (i) and (s) refer to the incident and scattered waves, respectively, and take two values corresponding to the (o) and (e) modes in a uniaxial medium; $\delta^{(j)}$ are the angles between vectors $\mathbf{e}^{(j)}$ and $\hat{\varepsilon}\,\mathbf{e}^{(j)}$,

$$G_{\alpha\mu\beta\nu} = G_{\alpha\mu\beta\nu}(\mathbf{k}^{(s)} - \mathbf{k}^{(i)})$$

is the Fourier transform of the correlation function of the director fluctuations,

$$\hat{G}(\mathbf{r} - \mathbf{r}') = \langle\delta\hat{\varepsilon}(\mathbf{r})\hat{\delta}\varepsilon^*(\mathbf{r}')\rangle,$$
$$\delta\varepsilon_{\alpha\beta} = \varepsilon_a(n_\alpha\delta n_\beta + n_\beta\delta n_\alpha),$$

and $\int d\Omega_{\mathbf{k}^{(s)}}$ denotes integration over all directions of the unit vector $\mathbf{k}^{(s)}/k^{(s)}$.

In a helical medium with large pitch, formula (A.2) retains is meaning, but the quantities $\mathbf{e}^{(j)}$, $n^{(j)}$, and $\delta^{(j)}$, being the functions of the angle $\theta^{(j)}(z)$ between the vectors $\mathbf{k}^{(j)}$ and $\mathbf{n}(z)$, also depend on z. The correlation function $\hat{G}(\mathbf{r}, \mathbf{r}')$ in the coordinate representation in the helical medium can be written as $\hat{G}(\mathbf{r} - \mathbf{r}'; (z + z')/2)$. For the light scattering, a significant part of the function $\hat{G}$ with respect to the first argument $|\mathbf{r} - \mathbf{r}'| \sim \lambda$, which is much smaller than P (the characteristic scale with respect to the argument $(z + z')/2$. This difference between the two scales allows us to perform substitution $(z + z')/2 \approx z \approx z'$ in all smoothly varying quantities and perform the Fourier transform with the wave vector $\mathbf{q} = \mathbf{k}^{(s)}(z) - \mathbf{k}^{(i)}(z)$ at a fixed $z \approx z'$, assuming the condition $|\mathbf{k}^{(s)}(z) - \mathbf{k}^{(i)}(z)| \gg q_0$ to be valid [7, 18].

The Fourier transform of the correlation function of the director fluctuations

$$g_{\alpha\beta}\left(\mathbf{r} - \mathbf{r}', \frac{z + z'}{2}\right) = \langle\delta n_\alpha(\mathbf{r})\delta n_\beta(\mathbf{r}')\rangle \tag{A.3}$$

with respect ti the variable $\mathbf{r} - \mathbf{r}'$ for a cholesteric LC with large pitch has, in the adiabatic approximation under the condition $q \gg q_0$, a form analogous to that for the nematic LC [1] with allowance for a smooth precession of the director along the z axis:

$$g_{\alpha\beta}(\mathbf{q}, z) = k_B T \times \sum_{j=1,2}\frac{e_{j\alpha}(\mathbf{q}, z)e_{j\beta}(\mathbf{q}, z)}{(K_{33} - K_{jj})(\mathbf{q}\cdot\mathbf{n}(z))^2 + K_{jj}q^2}, \tag{A.4}$$

where K_{jj} $(j = 1, 2, 3)$ are the Frank moduli, and

$$\mathbf{e}_1(\mathbf{q}, z) = \frac{\mathbf{q} - \mathbf{n}(z)(\mathbf{q}\cdot\mathbf{n}(z))}{\sqrt{q^2 - (\mathbf{q}\cdot\mathbf{n}(z))^2}},$$
$$\mathbf{e}_2(\mathbf{q}, z) = \mathbf{e}_1(\mathbf{q}, z)\times\mathbf{n}(z)$$

(see [18]).

This allows a formula of the (A.2) type to be used with the correlation function (A.4) for calculating the contribution to the extinction coefficient $\sigma^{(j)}(\theta^{(j)}(z))$ in a helical medium due to scattering by not very small angles.

REFERENCES

1. P. G. de Gennes and J. Prost, *The Physics of Liquid Crystals*, 2nd ed. (Clarendon Press, Oxford, 1993).
2. S. Chandrasekhar, *Liquid Crystals* (Cambridge Univ. Press, Cambridge, 1977).
3. V. A. Belyakov, *Diffraction Optics of Complex-Structured Periodic Media* (Nauka, Moscow, 1988; Springer, New York, 1992).

4. M. C. Mouguin, Bull. Soc. Fr. Mineral. Cristallogr. **34**, 71 (1911).
5. V. S. Liberman and B. Ya. Zel'dovich, Phys. Rev. E **49**, 2389 (1994); A. Yu. Savchenko and B. Ya. Zel'dovich, Phys. Rev. E **50**, 2287 (1994).
6. A. Yu. Val'kov, R. V. Grinin, and V. P. Romanov, Opt. Spektrosk. **83**, 239 (1997) [Opt. Spectrosc. **83**, 221 (1997)]; E. V. Aksenova, A. Yu. Val'kov, and V. P. Romanov, Mol. Cryst. Liq. Cryst. **359**, 351 (2001); Opt. Spektrosk. **91**, 1030 (2001) [Opt. Spectrosc. **91**, 969 (2001)].
7. E. V. Aksenova, A. Yu. Val'kov, and V. P. Romanov, Zh. Éksp. Teor. Fiz. **125**, 72 (2004) [JETP **98**, 62 (2004)].
8. M. Warenghem, M. Ismaili, and D. Hector, J. Phys. III **2**, 765 (1992); F. Simoni, F. Bloisi, L. Vicari, *et al.*, Europhys. Lett. **21**, 189 (1993); M. Warenghem, M. Ismaili, F. Simoni, *et al.*, Mol. Cryst. Liq. Cryst. **251**, 43 (1994); **251**, 61 (1994); M. Warenghem, D. Louvergneaux, and F. Simoni, Mol. Cryst. Liq. Cryst. **282**, 235 (1996).
9. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 8: *Electrodynamics of Continuous Media*, 2nd ed. (Fizmatgiz, Moscow, 2001; Pergamon Press, Oxford, 1984).
10. Yu. A. Kravtsov and Yu. I. Orlov, *Geometrical Optics of Inhomogeneous Media* (Nauka, Moscow, 1980) [in Russian].
11. L. M. Brekhovskikh, *Waves in Layered Media* (Akad. Nauk SSSR, Moscow, 1957; Academic, New York, 1960); L. M. Brekhovskikh and O. A. Godin, *Acoustics of Layered Media* (Nauka, Moscow, 1989; Springer, New York, 1990).
12. E. Miraldi, I. Trossi, and P. Taverna Valabreda, Nuovo Cimento B **60**, 165 (1980).
13. A. Yu. Val'kov and V. P. Romanov, Zh. Éksp. Teor. Fiz. **90**, 1264 (1986) [Sov. Phys. JETP **63**, 737 (1986)].
14. T. C. Lubensky, Phys. Rev. A **6**, 452 (1972).
15. A. Yu. Val'kov, L. A. Zubkov, A. P. Kovshik, and V. P. Romanov, Pis'ma Zh. Éksp. Teor. Fiz. **40**, 281 (1984) [JETP Lett. **40**, 1064 (1984)].
16. V. L. Ginzburg, *The Propagation of Electromagnetic Waves in Plasmas*, 2nd ed. (Nauka, Moscow, 1967; Pergamon Press, Oxford, 1970).
17. D. Langevin and M.-A. Bouchiat, J. Phys. (Paris) **36**, 197 (1975).
18. E. V. Aksenova, V. P. Romanov, and A. Yu. Val'kov, J. Math. Phys. **45**, 2420 (2004).

Translated by P. Pozdeev

Journal of Experimental and Theoretical Physics, Vol. 99, No. 5, 2004, pp. 978–986.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 5, 2004, pp. 1123–1132.

PLASMA, GASES

Compression of Whistler Waves in a Plasma with a Nonstationary Magnetic Field

M. E. Gushchin, S. V. Korobkov, A. V. Kostrov*, and A. V. Strikovsky

Institute of Applied Physics, Russian Academy of Sciences, Nizhni Novgorod, 603950 Russia

**e-mail: kstr@appl.sci-nnov.ru*

Received April 19, 2004

Abstract—We present the results of our experiments in which the propagation of whistler waves in a plasma with a nonstationary magnetic-field perturbation ($B = B_0 + \delta B(t)$, $\delta B/B_0 \le 5\%$) was investigated. The parametric and dispersive phenomena in a variable magnetic field were studied on the unique Krot plasma bench (the plasma column was 4 m in length and 1.5 m in diameter). A periodic field perturbation is shown to lead to an amplitude–frequency modulation of the whistler wave and to fragmentation of the signal into separate frequency-modulated wavepackets followed by their compression. The formation and compression of pulses is attributable to strong whistler group-velocity dispersion near the electron cyclotron frequency ($\omega \le \omega_H$). The results can be used to interpret the spectral shapes of the signals received from the Earth's magnetosphere and ionosphere in the electron and ion whistler frequency ranges. *© 2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

Investigating the excitation and propagation of whistlers in the Earth's magnetosphere and ionosphere is of current interest in diagnosing plasma parameters, studying nonlinear phenomena, and solving problems related to VLF radio communication [1]. It is well known that artificial signals can be significantly distorted as they propagate in the near-Earth plasma. Such phenomena as the amplitude–frequency modulation, the generation of satellites, and the broadening of the whistler frequency spectrum are often recorded in ground-based and satellite experiments [2, 3]. As regards the natural emissions, modulation can arise both directly during their generation in the magnetosphere and on the path of the signal in the plasma [4]. Nonlinear processes in the Earth's radiation belts [1] and the parametric interaction of waves with plasma-density and magnetic-field variations [3, 5] may be responsible for the variations in the amplitude–frequency characteristics of whistlers.

When parametric phenomena are analyzed, resonant processes are traditionally considered, implying that the space and time scales of the variations in parameters of the medium are related in a certain way to the frequency and wavenumber of the waves propagating in a plasma [5]. However, in our opinion, the most typical phenomena are nonresonant in nature [6]. In this case, no special constraints are imposed on the pattern of variations in parameters; in particular, it can be aperiodic.

Below, the linear interaction of waves with nonstationary variations in parameters of the medium, variations that exist independently of the traveling probe wave, is called parametric. The nonlinear effects attributable to the self-action of intense waves also belong to the broad class of parametric phenomena, but they are not considered here.

The geometrical-optics method generalized to nonstationary media [6, 7] can be used to describe waves in a medium with parameters that are relatively slow functions of the time and space coordinates. This method allows the frequency transfer equation for a quasi-monochromatic wave field to be directly written [6]:

$$\frac{\partial \omega}{\partial t} + (\mathbf{v}_g \nabla)\omega = -\omega \frac{(\partial n/\partial t)_{\omega, \mathbf{k}}}{\partial(n\omega)/\partial\omega}, \qquad (1)$$

where ω and $\mathbf{k}$ are the frequency and the wave vector, respectively; $\mathbf{v}_g = \mathbf{v}_g(\omega)$ is the group velocity of the wave; and n is the refractive index of the medium. Equation (1) is of fundamental importance in understanding how the frequency spectrum of the emission in nonstationary media is enriched. We see from Eq. (1) that the modulation of the refractive index is accompanied by the frequency modulation (FM) of the signal. The group-velocity dispersion causes different elements of the signal to propagate with different velocities; hence, the compression of individual parts of the wave can be observed.

The refractive index for longitudinal whistlers with frequencies

$$\omega_{LH} < \omega < \omega_H \ll \omega_p$$

(ω_H is the electron cyclotron frequency, ω_p is the electron plasma frequency, $\omega_{LH} = \sqrt{\omega_H \Omega_H}$ is the lower-

hybrid resonance frequency, and Ω_H is the ion cyclotron frequency) is

$$n = \frac{\omega_p}{\sqrt{\omega(\omega_H - \omega)}}. \quad (2)$$

In previous papers devoted to nonresonant phenomena in a magnetoactive plasma [8, 9], only the electron density was considered as a nonstationary parameter. However, it follows from expression (2) that the propagation characteristics of whistler waves are determined to a large extent by the magnetic field strength. Nonstationary disturbances of the Earth's magnetic field are recorded often, particularly during enhanced solar activity. Both periodic (associated with the excitation of magnetohydrodynamic modes) disturbances and aperiodic variations of the geomagnetic field are observed when solar-wind plasma streams interact with the Earth's magnetosphere. The magnetic field can also be disturbed by intense low-frequency whistlers; in this case, the field modulation is not accompanied by electron-density variations [10].

Of particular interest are ion whistlers, waves with left-hand polarization and with frequencies close to the ion cyclotron frequency: $\omega \le \Omega_H$ [11, 12]. The similarity of the electron and ion whistler dispersions near the cyclotron frequencies ($n_{\text{ion}} \approx (\Omega_H - \omega)^{-1/2}$) allows the results obtained in the electron whistler frequency range to be used to interpret phenomena in the ion whistler frequency range.

In this paper, we present the results of our experiments on the propagation of whistlers in a plasma with a weakly periodic magnetic-field modulation. The experiments were carried out on a Krot bench that was specially designed to simulate phenomena in space plasma.

2. DESCRIPTION OF THE EXPERIMENT

The experimental facility is a vacuum chamber 10 m in length and 3 m in diameter (Fig. 1a). A mirror-configuration (with a mirror ratio of $R = 2.4$, Fig. 1b) magnetic field is generated by a solenoid placed inside the vacuum chamber. Under experimental conditions, the magnetic field strength in the central cross section of the solenoid was $B_0 = 65$ G. A cylindrical plasma column (4 m in length, 1.5 m in diameter) that is not in contact with the metal walls of the chamber is produced by a pulsed inductive discharge ($f_{\text{dis}} = 5$ MHz, $P_{\text{gen}} = 250$ kW, $\tau_{\text{pulse}} = 1$ ms) in an argon atmosphere at a pressure of $p = 7 \times 10^{-4}$ Torr. The maximum plasma density at the discharge time reaches $\sim 10^{13}$ cm^{-3}, the electron temperature is $T_e \approx 10$ eV, and the ion temperature is $T_i \le 0.5$ eV. The plasma decay is governed by the ambipolar diffusion of electrons along the magnetic field.

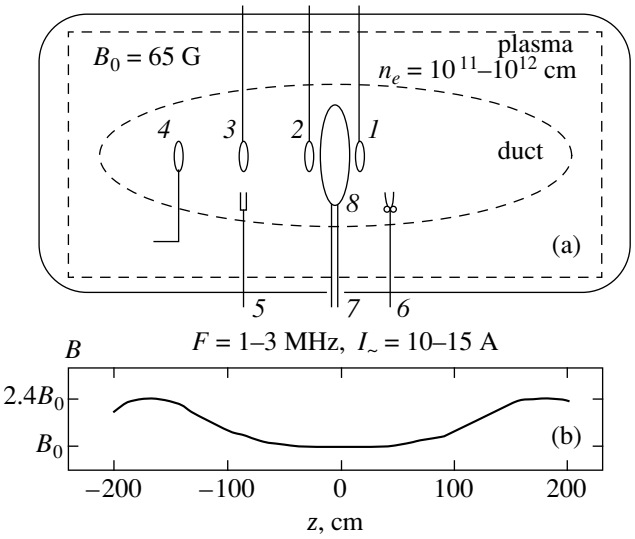


Fig. 1. (a) A scheme of the Krot experimental facility: *1*—emitting antenna; *2*, *3*, and *4*—receiving antennas; *5*—double probe; *6*—microwave probe; *7*—pulse generator (F = 1–3 MHz, I = 10–15 A); *8*—inductor. (b) The magnetic-field distribution along the axis of the vacuum chamber.

The experiments were carried out in a decaying plasma after the plasma-producing generators were turned off when the electron density reached $n_e = 10^{12}$ cm^{-3}. To decrease the divergence of whistler waves in a plasma, a duct with a reduced electron density extended along the axis of the system was produced; the whistlers excited by the antennas in this duct were kept in ducting mode [13]. An additional pulse generator ($\tau_{\text{pulse}} = 1$ ms) operated at frequencies of $F = 1$–3 MHz was used to produce the duct. The generator was loaded onto a loop antenna (an inductor $D = 20$ cm in diameter) located in the center of the chamber. The power of the generator was about 10 kW; the variable current induced in the antenna by the generator reached $I_{\max} = 30$ A. The duct was formed through the local heating of electrons in the near field of the inductor and the subsequent thermodiffusion-induced plasma redistribution [14]. The duct formation is governed by unipolar diffusion: nonmagnetized ions and electrons diffuse, respectively, across and along the magnetic field, while the quasi-neutrality is ensured by the closure of the current through the background plasma. A quasi-stationary plasma density distribution was established in a time of ~500 μs; the length of the duct L reached ~3 m, and its diameter was determined by the diameter of the antenna and the electron heat-conduction length across the magnetic field:

$$\Lambda_\perp \approx \rho_e/\delta^{1/2} \approx 20 \text{ cm}$$

(ρ_e is the Larmor electron radius, and $\delta = 2m/M$, where m and M are the electron and ion masses, respectively).

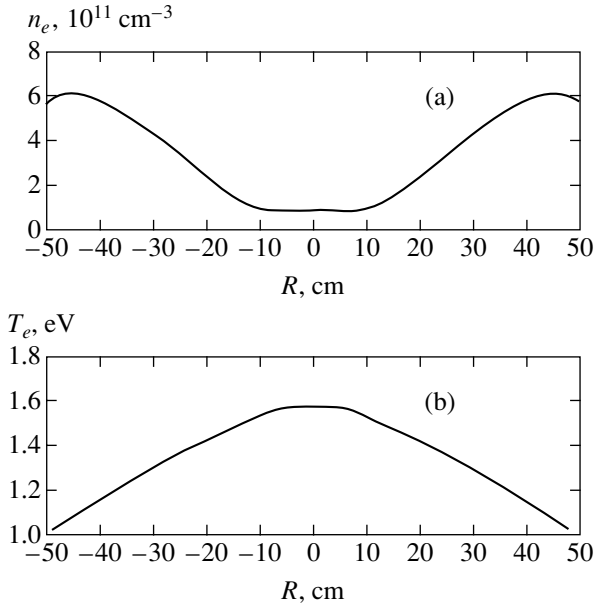


Fig. 2. (a) The transverse plasma density distribution 900 μs after the onset of a LF pulse (F = 1.2 MHz). (b) The transverse plasma electron temperature distribution at the same time.

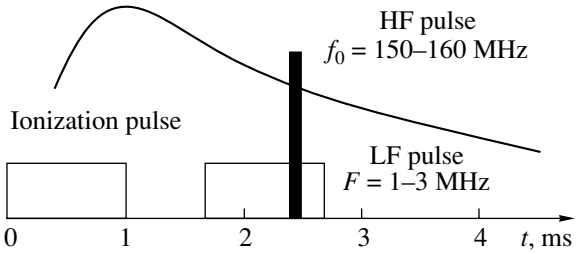


Fig. 3. A time diagram for the operation of the experimental facility.

The radial plasma density and temperature distributions in a stationary duct 900 μs after the onset of a heating pulse (F = 1.2 MHz) are shown in Fig. 2. The plasma density was measured by a tiny probe with a microwave cavity produced on a segment of a double-wire line [15]; a double probe was used to measure the electron temperature. The plasma density on the axis of the duct was $n_e = 10^{11}$ cm^{-3}, and the electron temperature was T_e = 1.5 eV.

As the source of a nonstationary magnetic-field perturbation, we used the same inductor as that used to produce the duct. After a quasi-stationary density distribution was established, the external magnetic-field perturbation was determined by the variable near field of the antenna (F = 1–3 MHz). The structure of the low-frequency (LF) field near the inductor (the axial magnetic-field component) in the formed density duct matched the field distribution of a current loop in a vacuum. The amplitude of the variable magnetic field in the plane of the inductor reached $\delta B_{\max} = 3$ G, which corresponds to a relative perturbation $\delta B_{\max}/B_0 \approx 5\%$. At distances $z > D$ from the inductor, LF whistler waves with

$$k_\perp \approx \frac{\omega_p}{c}, \quad k_\parallel \approx 2\frac{\omega_p}{c}\frac{\omega}{\omega_H}$$

were excited in the plasma. It is easy to show that the relative magnetic-field perturbation in waves of this type is much larger than the plasma density perturbation [10]:

$$\frac{\delta n}{n_0} \approx \left(\frac{\omega_H}{\omega_p}\right)^2 \frac{\delta B}{B_0} \ll \frac{\delta B}{B_0}.$$

Under experimental conditions, $\omega_H < 0.1\omega_p$, δn and δB are the plasma density perturbation and the magnetic-field perturbation of a LF whistler wave.

A time diagram for the operation of the experimental facility is shown in Fig. 3. The propagation of high-frequency (HF) whistlers in a plasma with a nonstationary magnetic field was investigated at the frequencies of the probe wave f_0 = 150–160 MHz. In this case, the condition $f_H/2 < f_0 \le f_H$ ($f_H = \omega_H/2\pi$) was satisfied, and the quasi-longitudinal whistlers with wavelengths $\lambda = 2\pi/k \approx$ 4–5 cm were confined in a duct with a reduced plasma density [13]. The probe wave was emitted into the plasma in the form of a pulse τ = 0.1–10 μs in duration.

The chosen whistler frequencies are close to the cyclotron frequency, $f_0 \approx 0.9 f_H$, and the whistler group-velocity dispersion is fairly strong. In this frequency range, the whistlers are also called cyclotron waves. Their propagation characteristics are determined to a large extent by the collisional damping and the kinetic collisionless absorption by resonant particles — plasma electrons.

HF waves were emitted and received by shielded single-turn magnetic loop antennas 1–2 cm in diameter placed in different cross sections of the facility. The antennas were coated with a dielectric layer to reduce the plasma influence on their impedance characteristics. In general, the antennas in the experiments were placed on the axis of the plasma duct; the plane of the antennas was oriented along the external magnetic field lines. The emitting antenna was located near the inductor, at a distance of Δz = 5 cm from its plane. The loops installed on the opposite side of the inductor at different distances from the emitting antenna were used as the receiving antennas (In Fig. 1a, *1* is the emitting antenna, and *2*, *3*, and *4* are the receiving antennas).

The HF signals passed through the plasma were recorded by a digital oscillograph; their spectral analysis was performed by using a numerical Fourier transform.

3. EXPERIMENTAL RESULTS

The main experimental results are presented in Fig. 4. When passing through the region with a perturbed field, a continuous HF signal (f_0 = 160 MHz) breaks up into separate wavepackets that follow one another with the field modulation period (F = 1.2 MHz). The amplitude modulation depth of the signal increases as the receiving antenna recedes from the emitting antenna. The compression of HF pulses is observed—their duration decreases as they propagate in the plasma. A spectral analysis shows that the received signal is frequency-modulated; the relative frequency deviation $\Delta f/f$ does not exceed 1–3%. The filling frequency decreases from the beginning to the end of each pulse; the slope of the time–frequency relation increases with increasing distance between the receiving and emitting antennas (Figs. 4d–4f). The minimum duration to which the frequency-modulated (FM) signal is compressed is $\tau_{min} \approx \Delta f^{-1} \approx 100$ ns (Fig. 4c), where $\Delta f \approx 10$ MHz is the width of the frequency spectrum for the FM whistler wave (Figs. 4d–4f).

The cyclotron absorption of whistler waves is clearly seen (Fig. 5a). At large distances from the emitting antenna (Δz = 260 cm), the signal spectrum is shifted as a whole downward in frequency. The observed pattern is the result of nonuniform damping of various elements of the FM-signal spectrum, which must have been symmetric about the carrier frequency of the probe wave, $\Delta f \approx 159$ MHz, in the absence of absorption.

The signal frequency modulation is relatively small when a HF wave is excited and received outside the region of an intense magnetic-field perturbation. Figure 6 presents the results obtained at a distance of L = 70 cm from the inductor, the distance between the emitting and receiving antennas placed on the duct axis was $\Delta z \approx 15$ cm. The amplitude of the magnetic-field pertur-

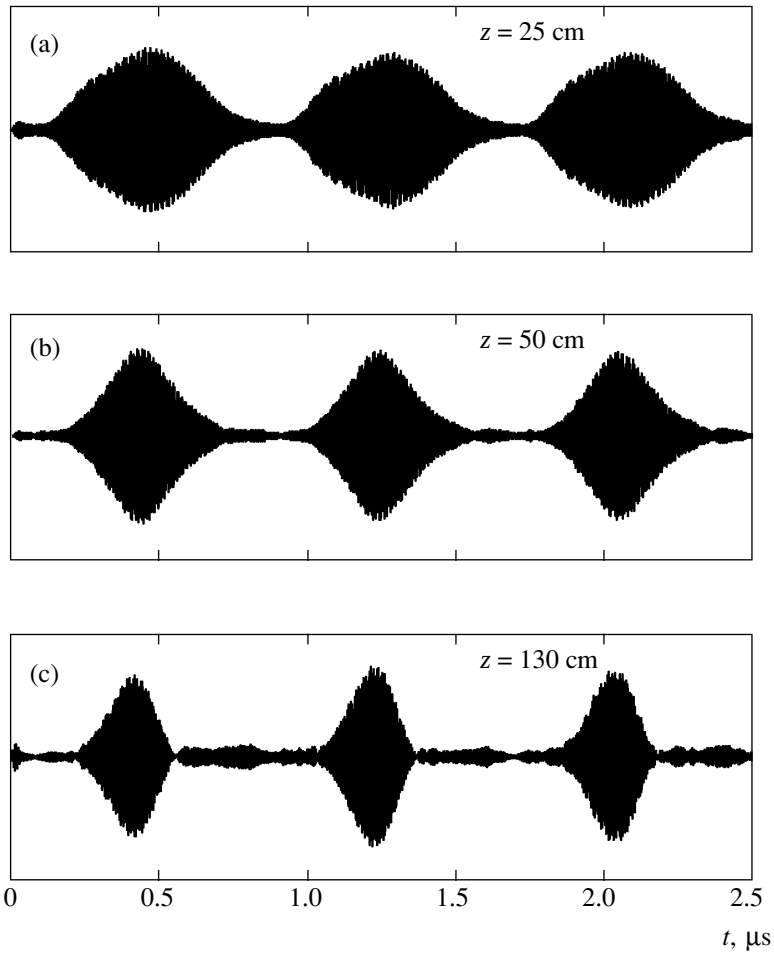


Fig. 4. Oscillograms (a–c) and spectrograms (d–f) of the HF signals (f_0 = 160 MHz) received from the plasma at various distances from the emitting antenna in the presence of a magnetic-field perturbation at F = 1.2 MHz ($\delta B/B_0 \approx 3\%$).

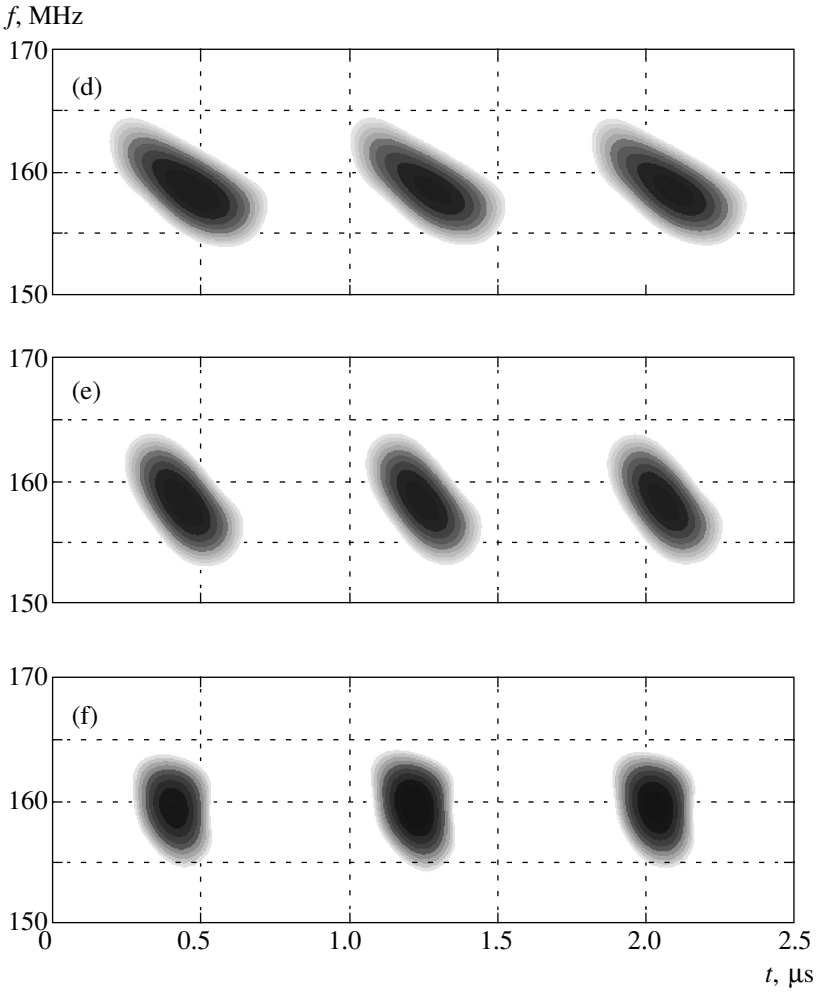


Fig. 4. (Contd.)

bation by LF whistler waves (F = 1.2 MHz) is smaller by almost a factor of 10 than the field perturbation in the plane of the inductor; the full width of the modulated frequency spectrum is $\Delta f < 1$ MHz (Fig. 6b). Despite the low frequency modulation and the small wave path, the group-velocity dispersion still causes the signal envelope to be distorted appreciably (Fig. 6a).

The wave frequency transformation was investigated in experiments with short pulses ($\tau \approx 150$ ns, f_0 = 160 MHz) that were fed to the emitting antenna with different time delays. The delays were chosen in such a way that a HF pulse was emitted into the plasma at different phases of the periodic magnetic-field perturbation (F = 3 MHz, $\delta B_{max}/B_0 \sim 5\%$). Depending of the phase of the variable field, signals with "redshifted" and "blueshifted" frequencies were recorded (Fig. 7). The frequency shift was determined by the time at which the wave passed through the plasma: at the maximum of the magnetic field, when the amplitude of the periodic perturbation was at a maximum and the variable magnetic field was aligned with the external field,

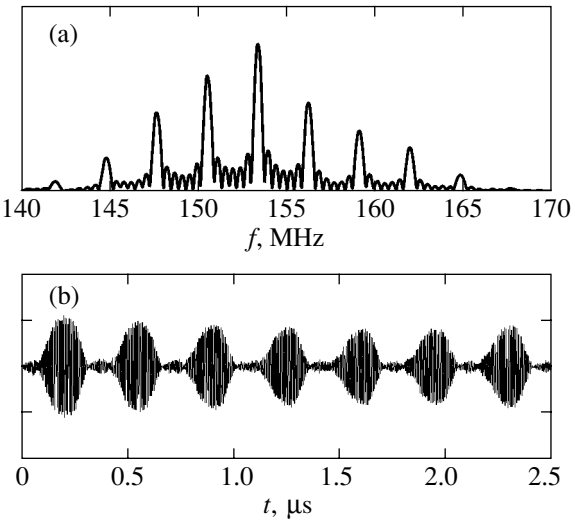


Fig. 5. A spectrum (a) and oscillogram (b) of the HF signal (f_0 = 159 MHz) received from the plasma in the presence of a LF magnetic-field perturbation at F = 3 MHz, $\delta B/B_0 \approx 5\%$. The distance between the receiving and emitting antennas is $\Delta z \approx 260$ cm; both antennas were placed on the axis of the plasma duct.

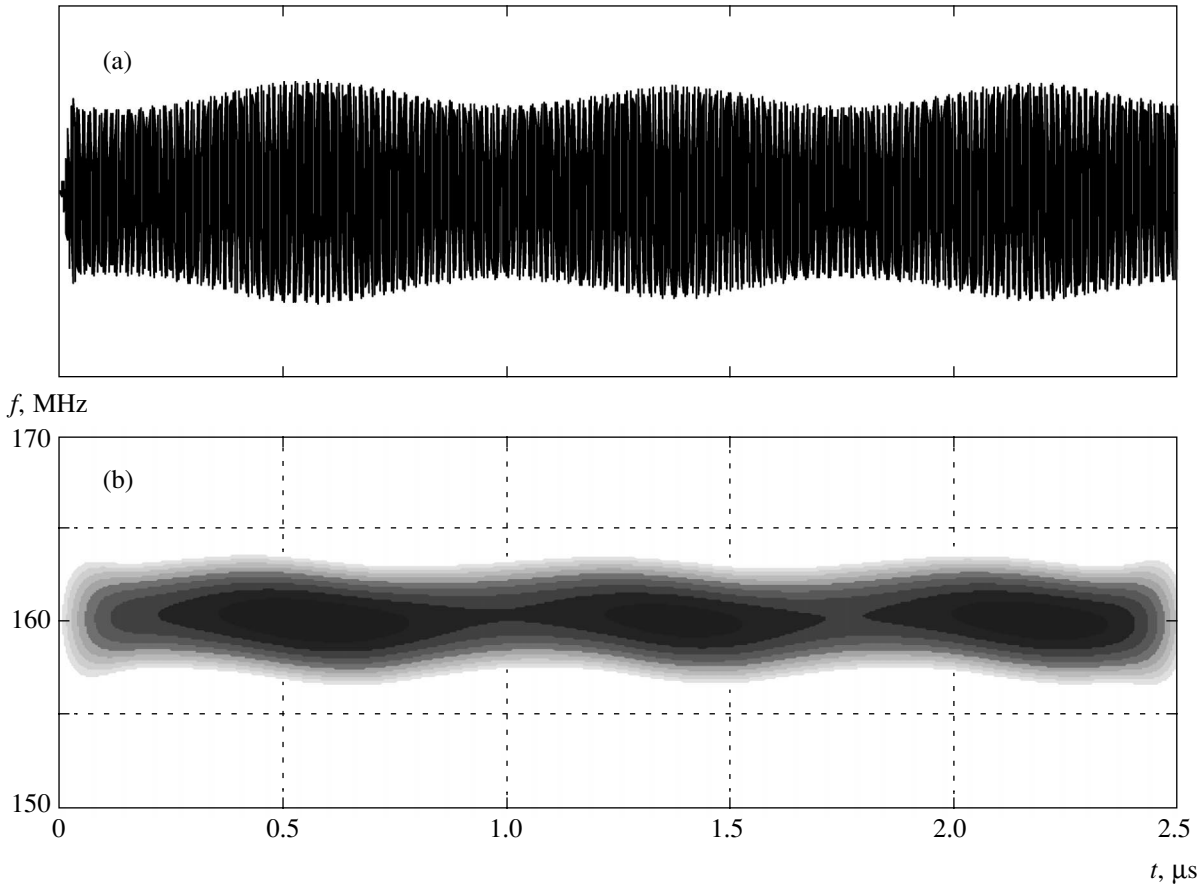


Fig. 6. An oscillogram and (b) spectrogram of the HF whistler (f_0 = 160 MHz) excited and received from the plasma at a distance $z \approx 70$ cm from the LF antenna (F = 1.2 MHz) in the region of a weak external magnetic-field perturbation ($\delta B/B_0 < 0.5\%$).

and at its minimum, when the directions of the external and loop fields were opposite.

If the antennas are separated by a large distance ($\Delta z > 1$ m) and are placed off the duct axis (Δr = 5–10 cm), then signals with a complex amplitude envelope and an intricate spectral shape can be observed (Fig. 8). Such a form of the signals can be explained by the complex structure of the whistler modes of the broad ($D_{\text{duct}} > 2\pi/k$) plasma duct. As the antenna placed off the duct axis emits whistlers, not only axisymmetric modes of the plasma duct, but also asymmetric modes with nonzero azimuthal wave numbers are probably effectively excited. The HF field recorded by the antenna at a chosen point in space is the result of interference between several modes of the plasma duct with different group velocities.

4. DISCUSSION

The experimental results can be explained by two effects. The first effect is the nonresonant parametric frequency modulation of whistlers in a plasma with a variable refractive index, whose nonstationary behavior is related to the harmonic modulation of the magnetic field.

The second effect is the compression of a FM whistler wave attributable to strong whistler group-velocity dispersion at frequencies close to the electron cyclotron frequency. The expression for the group velocity of longitudinal whistlers is

$$v_g = 2c\frac{\omega^{1/2}\omega_H^{1/2}}{\omega_p}(\omega_H - \omega)^{3/2}. \quad (3)$$

Since each point of the frequency envelope of a FM signal moves with its own group velocity, some parts of the wave are extended, while other parts are compressed. Under certain conditions, the wave energy can be localized on short time intervals where the field amplitude can be much larger than the initial amplitude [6]. The compression of FM pulses in a dispersive medium is used, in particular, to produce intense microwave pulses [16].

The propagation of a whistler wave in a plasma with a nonuniform and nonstationary magnetic field is rather difficult to analyze theoretically when the dispersion is taken into account. For this reason, we propose using a

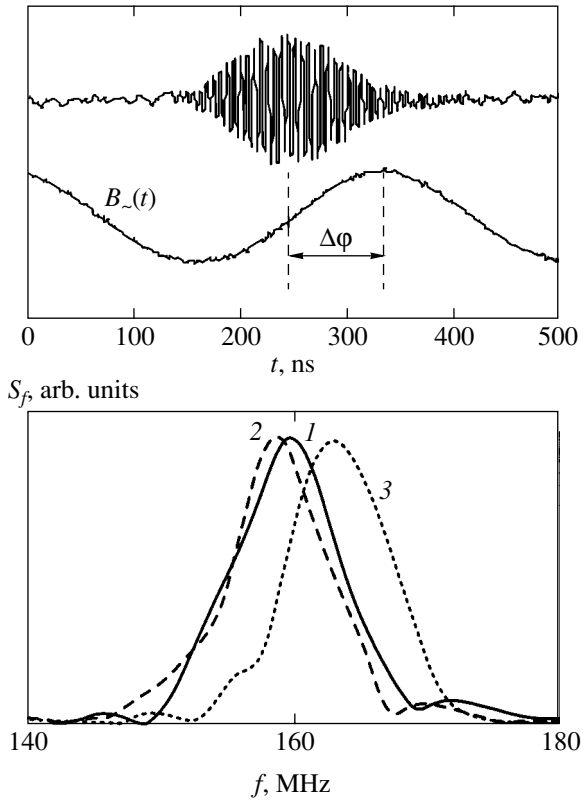


Fig. 7. Transformation of the frequency spectrum of a short pulse (f_0 = 160 MHz, $\tau \approx 150$ ns) emitted into the plasma at different phases, $\Delta\varphi = \pi/2$ (*1*), π (*2*), and 0 (*3*), of the LF (F = 1.2 MHz) magnetic-field perturbation.

simplified geometrical-optics model for theoretical estimations.

To estimate the frequency transformation effect, we disregard the whistler group-velocity dispersion. Let us assume that a longitudinal whistler with an initial frequency ω_0 propagates in a homogeneous plasma (along the z axis) with a constant (over the path) external magnetic field and represent the variable field of the inductor as a spatially uniform perturbation localized on a segment of the path with length L. The frequency transfer equation (1) for whistlers takes the form

$$\frac{\partial\omega}{\partial t} + v_g(\omega_0)\frac{\partial\omega}{\partial z} = \frac{\omega}{\omega_H}\frac{\partial\omega_H}{\partial t}. \qquad (4)$$

Let us consider the harmonic magnetic-field modulation with frequency Ω:

$$\tilde{B} = \begin{cases} 0, & z \notin [0, L], \\ \delta B \sin\Omega t, & z \in [0, L]. \end{cases} \qquad (5)$$

The solution of Eq. (4) using (5) shows that the signal frequency is modulated harmonically:

$$\omega(t) = \omega_0 + \delta\omega\sin\Omega t.$$

The maximum frequency shift

$$\delta\omega = \omega_0\frac{\delta B}{B_0}$$

is reached at the points that are at distance z_1 from the point at which the wave enters the variable magnetic field:

$$z_1 = \frac{\pi v_g}{2\Omega}N, \quad N = 1, 2, \ldots.$$

Under experimental conditions ($\omega_0/2\pi$ = 160 MHz, $\Omega/2\pi$ = 1.2 MHz, $v_g = 1.5 \times 10^8$ cm s^{-1}), the maximum frequency shift corresponding to $N = 1$ occurs at z_1 = 30 cm; i.e., a strong deviation can be observed as the wave traverses a path with a length of $L \sim z_1$. The length scale of the near field of the antenna, a segment with an intense quasi-static magnetic-field perturbation, is $\Delta z \sim D$ = 20 cm. Thus, $z_1 \sim D$ and the frequency modulation index virtually reaches its maximum after the whistler has traversed the region of the inductor, in agreement with the experimental data.

To estimate the compression of FM whistler pulses, we assume the signal frequency modulation to be fixed and at a maximum:

$$\omega(t) = \omega_0\left(1 + \frac{\delta B}{B_0}\sin\Omega t\right).$$

Taking into account the frequency dependence of the whistler group velocity (3), we can determine the distance between the point at which a FM signal enters a dispersive medium and the compression point:

$$z_2 = \frac{v_g^2}{\left|\frac{\partial v_g}{\partial\omega}\right|\left|\frac{\partial\omega}{\partial t}\right|_{\max}}. \qquad (6)$$

Since the condition

$$\frac{\partial v_g}{\partial\omega} < 0$$

is satisfied for whistlers with frequencies close to the cyclotron frequency, the elements of the FM signal with

$$\left.\frac{\partial\omega}{\partial t}\right|_{t = t_0} < 0$$

(t_0 is the time at which the segment of the wavepacket under consideration enters the dispersive medium) will

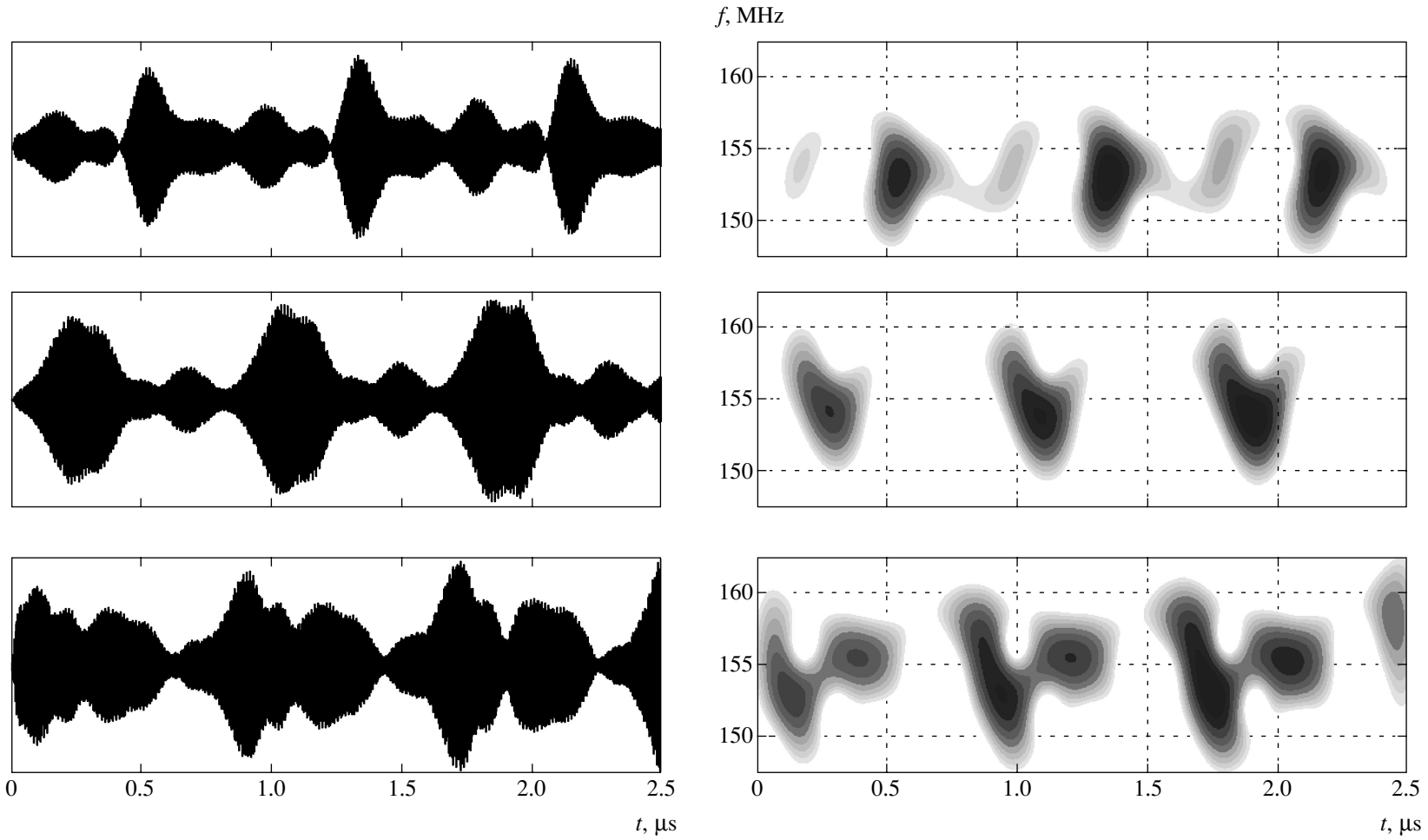


Fig. 8. Sample oscillograms and spectrograms of the HF signals (f_0 = 155 MHz) received from plasma in the presence of a magnetic-field perturbation at F = 1.2 MHz ($\delta B/B_0 \approx$ 3%). The distance between the receiving and emitting antennas is $\Delta z \approx$ 150 cm; both antennas were placed off the axis of the plasma duct, $\Delta r \approx$ 5–10 cm.

be compressed. The estimate based on formula (6) indicates that $z_2 \approx 60$ cm at $B_0 = 65$ G. Note that the field in our calculations was assumed to be uniform throughout the path of the wave. Under experimental conditions, however, the length of the segment with a uniform magnetic field at the center of the trap is only about 40 cm. We see from Fig. 4a that the signal is separated into individual packets as the wave traverses the region of the near field of the inductor. Thus, the estimate is valid on this segment of the path. Subsequently, however, the wavepackets are compressed more slowly, because the field increases in strength toward the magnetic mirror and, hence, the whistler group slowdown and frequency dispersion decrease.

The geometrical-optics approximation used in our estimations breaks down near the compression point. The minimum duration to which a pulse is compressed is determined by the full width of the FM signal frequency spectrum. As was noted above, the cyclotron damping of high-frequency spectral components causes the spectrum to narrow and prevents the effective compression of whistler waves. In addition, no significant increase in the pulse amplitude compared to the initial signal is observed due to the damping.

5. CONCLUSIONS

Our experimental results show that a nonstationary magnetic-field perturbation produces a frequency modulation of the whistler wave that propagates in a plasma. The strong whistler group-velocity dispersion at frequencies $\omega \leq \omega_H$ is responsible for the FM signal deformation; the fragmentation of a HF signal into separate wavepackets with FM filling is observed. The fact that the compression is caused by the magnetic-field modulation at an unperturbed electron density is of fundamental importance.

The peculiarities of the propagation of whistlers in a plasma with a nonstationary magnetic field can be used for diagnostic purposes. In particular, the amplitude of the magnetic-field variations ($\delta\omega/\omega_0 \sim \delta B/B_0$) can be determined from the frequency modulation of the signal that passes through a perturbed region.

In conclusion, note that the oscillograms and spectrograms of modulated whistlers resemble the recordings of the signals received from the near-Earth plasma in the ELF frequency range—structured Pc-1,2 geomagnetic pulsations or "pearls" [11]. Within the framework of existing models, it is assumed that the amplitude and frequency modulations of such signals arise directly in the generation process attributable to the growth of ion cyclotron instability. However, a clear correlation between the FM pearl repetition period and the LF geomagnetic-field modulation period is commonly observed when Pc-1,2 magnetic pulsations and lower-frequency Pc-3,4 pulsations are recorded simultaneously [12]. The similarity of the signals and full-scale experiments suggest that the pearls are formed through the parametric modulation of ion whistlers by LF pulsations of the Earth's magnetic field and their subsequent dispersive compression. In this case, the received signal contains information about the presence of a LF perturbation that is not recorded on the ground, and the Pc-3,4 pulsation amplitude can be diagnosed by the pearl frequency modulation.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, project no. 04-02-17188 and by Department of Science of the Russian Federation under the program for support of unique facilities, grant no. 701-18.

REFERENCES

1. O. A. Molchanov, *Low-Frequency Waves and Induced Radiation in the Near-Earth Plasma* (Nauka, Moscow, 1985) [in Russian].
2. Y. Tanaka, D. Lagoutte, M. Hayakawa, *et al.*, J. Geophys. Res. **92**, 7551 (1987).
3. S. Ohnami *et al.*, Geophys. Res. Lett. **20**, 739 (1993).
4. N. Sato, K. Hayashi, S. Kokubun, *et al.*, J. Atmos. Terr. Phys. **36**, 1515 (1974).
5. V. Y. Trakhtengerts and M. Hayakawa, J. Geophys. Res. **98**, 19205 (1993).
6. L. A. Ostrovskiĭ and N. S. Stepanov, Izv. Vyssh. Uchebn. Zaved., Radiofiz. **14**, 489 (1971).
7. Ya. A. Kravtsov and Yu. I. Orlov, *Geometrical Optics of Inhomogeneous Media* (Nauka, Moscow, 1980) [in Russian].
8. M. I. Bakunov and I. S. Grachev, Phys. Rev. E **65**, 036405 (2002).
9. D. K. Kalluri, V. R. Goteti, and A. M. Sessler, IEEE Trans. Plasma Sci. **21**, 70 (1993).
10. A. V. Kostrov *et al.*, Pis'ma Zh. Éksp. Teor. Fiz. **78**, 1026 (2003) [JETP Lett. **78**, 538 (2003)].
11. V. A. Troitskaya and A. V. Gul'el'mi, Usp. Fiz. Nauk **97**, 453 (1969) [Sov. Phys. Usp. **12**, 195 (1969)].
12. T. A. Plyasova-Bakounina, J. Kangas, K. Mursula, *et al.*, J. Geophys. Res. **101**, 10965 (1996).
13. M. T. Zaboronkova, A. V. Kostrov, A. V. Kudrin, *et al.*, Zh. Éksp. Teor. Fiz. **101**, 1151 (1992) [Sov. Phys. JETP **75**, 625 (1992)].
14. S. V. Egorov, A. V. Kostrov, and A. V. Tronin, Pis'ma Zh. Éksp. Teor. Fiz. **47**, 86 (1988) [JETP Lett. **47**, 102 (1988)].
15. I. G. Kondrat'ev, A. V. Kostrov, A. I. Smirnov, *et al.*, Fiz. Plazmy **28**, 977 (2002) [Plasma Phys. Rep. **28**, 900 (2002)].
16. W. M. Manheimer and B. H. Ripin, Phys. Fluids **29**, 2283 (1986).

Translated by V. Astakhov

Journal of Experimental and Theoretical Physics, Vol. 99, No. 5, 2004, pp. 987–997.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 5, 2004, pp. 1133–1145.

FLUIDS

Helicity Generation in Uniform Helical Flows

S. G. Chefranov

Oboukhov Institute of Atmospheric Physics, Russian Academy of Sciences, Moscow, 109017 Russia
Institute of Theoretical and Experimental Biophysics, Russian Academy of Sciences, Pushchino, Moscow oblast, 142292 Russia
e-mail: schefranov@mail.ru

Received July 14, 2003

Abstract—Helicity generation conditions are derived for helical flows of Joukowski type with allowance for effects due to viscosity, buoyancy, temperature nonuniformity, and solid-body rotation. The upper and lower limits are determined for the rotation-frequency interval in which helicity can be generated by viscous forces. These conditions correspond to the regime of an isolated tornado-like vortex. An exact solution to the time-independent equations of motion for inviscid incompressible flow is obtained. The solution describes a generalized Kelvin–Helmholtz vortex having the form of a localized cylindrical vortex with nontrivial stable topological vortex-core structure determined by a finite value of helicity. For linear traveling inertia waves, which must have uniform helical structure, a general representation is found that characterizes helical structures of different origin. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

It is well known that the vorticity generated in flows over obstacles can be transferred into the background stream. This process is characterized by vorticity flux [1], i.e., the average direct product of velocity $\mathbf{u}$ and vorticity $\boldsymbol{\omega} = \mathrm{curl}\,\mathbf{u}$. The trace of the tensor, or the helicity defined as $H = \boldsymbol{\omega} \cdot \mathbf{u}/2$, plays an important role in turbulence theory [2, 3] and in magnetic dynamo theory [4, 5].

Among the diversity of observed atmospheric vortices, relatively high helicity values (up to 10 m/s^2) are characteristic of tornadoes and tornado-like vortices, such as dust devils and waterspouts [6], which have essentially three-dimensional structure. Moist-convective tornadic vortices typically have a relatively coherent uniformly helical flow structure despite the huge Reynolds numbers corresponding to wind velocities up to 500 km/h inside tornadoes (see [7] and references therein). Indeed, by the Helmholtz vorticity theorem, flows of Gromeka–Beltrami type (with $|\boldsymbol{\omega} \times \mathbf{u}| = 0$[1] or Joukowski type (with $[(\boldsymbol{\omega} + 2\boldsymbol{\Omega}) \times \mathbf{u}] = 0$, where $\boldsymbol{\Omega}$ is solid-body rotation frequency) are characterized by the lowest rates of energy dissipation per unit mass corresponding to a given kinetic energy [8–11]. However, these flows are described by exact solutions to the Helmholtz vorticity equation [8, 9] and are topologically different from other time-independent flow regimes [12, 13].

Helical vortex structures similar to uniform helical flows are also observed experimentally in laboratory models of tornadic vortices at a sufficiently large distance from the rotating bottom [14].

In this paper, we apply a helicity balance equation to Joukowski-type helical flows to explore the possibility of helicity generation in the presence of buoyancy, viscous dissipation, and solid-body rotation with frequency $\boldsymbol{\Omega}(t)$.

The paper is organized as follows. In Section 2, a general balance equation for the integral helicity

$$\overline{H} = \int d^3x H$$

is derived and used as a basis for a qualitative analysis of flow regimes that admit production of $\overline{H}$ via the spin-down effect and latent heat release (cooling) in the presence of cyclonic or anticyclonic vorticity perturbations, as well as by horizontal components of vorticity perturbations over a sloped underlying surface in the presence of viscous dissipation.

In Section 3, a general uniform helical-vortex solution to the linear equations of fluid dynamics is obtained. In particular, it describes traveling inertia waves in a fluid rotating as a whole. Furthermore, an exact time-independent solution to the nonlinear equations of fluid dynamics is obtained. It has the form of a cylindrical helical vortex of radius R, which reduces to a generalization of the Helmholtz vortex line as $R \longrightarrow 0$ and is structurally stable (as are tornadic vortices) since helicity is a topological invariant. The interval of Ω in which the vortex can have a nonmonotonically varying tangential velocity u_φ is determined, which

[1] This relation is satisfied if $\boldsymbol{\omega}$ and $\mathbf{u}$ are collinear, i.e., $\boldsymbol{\omega} = k\mathbf{u}$ where k is a function of coordinates in the general case. However, k = const throughout this study, which corresponds to uniform helical flow.

implies that a contaminant can concentrate in the region of minimal u_φ.

In Section 4, it is shown that local helicity H can be generated by viscous dissipation in flow regimes with supercritical Ω. This possibility is consistent with observations of tornado-like vortices. Condition (17) is obtained for generation of integral helicity in a Joukowski-type flow by the combined effects of horizontal temperature nonuniformity and solid-body rotation with $\Omega_z \neq 0$. Moreover, the existence of two qualitatively different regimes of helicity evolution is demonstrated for a Joukowski-type flow with horizontal temperature nonuniformity and $\mathbf{\Omega} \neq 0$. One of these regimes (which corresponds to $H \longrightarrow 0$ and $t \longrightarrow \infty$) can be exactly described by a well-known solution for nonuniform solid-body rotation of a fluid ellipsoid [15]. The other regime is a steady flow of new type with a finite H in the limit of $t \longrightarrow \infty$.

2. HELICITY BALANCE EQUATION

Consider the equations of viscous incompressible flow written in the Boussinesq approximation in a coordinate system rotating with frequency $\mathbf{\Omega}(t)$ (see Eq. (A.1) in the Appendix). This noninertial coordinate system obviously corresponds to the initial stage of tornado formation, when tornado-scale cyclone rotates as a whole beneath a source cumulonimbus cloud. The localized helical perturbations analyzed in this study develop in a solid-body rotational flow.

In the Appendix (see also [16]), the following balance equation is obtained for the integral helicity

$$\bar{H} = \int_V d^3x H$$

(V is the domain of vortex flow field):

$$\frac{d\bar{H}}{dt} = -2(\dot{\mathbf{\Omega}} \cdot \mathbf{u}) + \beta g(\overline{T\omega_z}\cos\alpha + \overline{T\omega_y}\sin\alpha) + \nu\Delta\mathbf{u} \cdot \boldsymbol{\omega}, \quad (1)$$

where the overbar denotes an integral over V; $\dot{\mathbf{\Omega}} = d\mathbf{\Omega}/dt$; α is the angle between the gravitational acceleration g and the z axis (normal to the Earth surface); the y and x axes are directed northward and eastward, respectively; T is a temperature perturbation; β is the thermal expansion coefficient; $\rho = \rho_0(1 - \beta T)$ is the fluid density ($\rho_0 = \text{const}$); and ν is kinematic viscosity.

Equation (1) with $\alpha = 0$ was used in [16] to find a scaling parameter that justifies laboratory-scale modeling of atmospheric tornado-like vortices. When $\dot{\mathbf{\Omega}} = 0$, Eq. (1) yields a well-known criterion for invariance of $\bar{H}$ in inviscid homogeneous flow, because all terms on the right-hand side of (1) vanish in this case. However, if $\dot{\mathbf{\Omega}} \neq 0$ and $\dot{\mathbf{\Omega}} \cdot \bar{\mathbf{u}} < 0$, then helicity can be generated by the spin-down effect. In particular, in the case of $\dot{\mathbf{\Omega}} < 0$ (decelerating solid-body rotation), the value of $|\dot{\mathbf{\Omega}}|$ can be determined by bottom friction linearly scaling with velocity (see Appendix), in which case solid-body rotation is preserved. However, solid-body rotation is also preserved when $\dot{\mathbf{\Omega}} \neq 0$ in the (nondissipative) exact solution to the Helmholtz vorticity equation describing the "fluid" elliptic gyroscope [15] (see also the Conclusions section).

Generation of $\bar{H}$ is possible when the temperature and vorticity variations are such that the second and third terms on the right-hand side of (1) remain positive. For example, when heat is released ($T > 0$) as moisture condenses in a cyclonically rotating updraft ($\omega_z > 0$, counterclockwise rotation in the northern hemisphere), the second term of the right-hand side of (1) is responsible for $\bar{H}$ generation. Indeed, observations suggest that more than 90% of tornadoes rotate counterclockwise. Alternatively, Eq. (1) implies that $\bar{H}$ can be generated in anticyclonically rotating downdrafts ($\omega_z < 0$), in which the air mass cools down ($T < 0$).

Note that an underlying-surface slope ($\alpha \neq 0$) may also be responsible for $\bar{H}$ generation when $\omega_z = 0$ and $\overline{\omega_y T} > 0$, as in convective rolls (horizontally oriented vortices).

To complete a qualitative analysis of helicity balance equation (1), note that the sign of the viscous term in (1) may change. Indeed, it was shown in [17] that $d\bar{H}/dt|_{t=0} > 0$ under certain conditions in an appropriate turbulent flow regime because of this dissipative term. The possibility of similar evolution of local helicity in a Joukowski-type flow is examined below. However, certain properties of uniform helical vortex structures should be specified.

3. UNIFORM HELICAL VORTEX STRUCTURES

Even though Gromeka–Beltrami and Joukowski-type vortex flows have been identified under natural and experimental conditions, they are seldom mentioned in studies in classical and geophysical fluid dynamics, probably, because of low likelihood of their occurrence (see [21, p. 210]). Note, however, that even inertia waves in rotating fluids can be referred to this type of structures. This fact was noted in [4], but not even mentioned in [18–20], whereas it can be used to derive a more general representation for helical inertia waves (see below), which can be helpful in analyzing analogous helical flow structures found in other areas of physics.

3.1. Consider the linear equation for inertia waves written in a coordinate system tied to an unbounded rotating homogeneous fluid (see [18, 20] and Eq. (A.2)),

$$\frac{\partial}{\partial t}(\mathrm{curl}\mathbf{u}) = 2\Omega_z\frac{\partial \mathbf{u}}{\partial z} + \nu\Delta(\mathrm{curl}\mathbf{u}), \quad \mathrm{div}\mathbf{u} = 0, \quad (2)$$

where $\mathbf{\Omega} = (0, 0, \Omega_z)$ for simplicity. In [18], the solution to (2) corresponding to $\nu = 0$ was represented as a traveling plane wave:

$$\mathbf{u} = \mathrm{Re}\mathbf{A}\exp[i(\mathbf{k}\cdot\mathbf{x} - \omega t)], \quad \mathbf{k}\cdot\mathbf{A} = 0, \qquad \omega = 2\Omega_z k_z/k, \quad (3)$$

where

$$k = |\mathbf{k}|, \quad \mathbf{A} = \mathbf{a} + i\mathbf{b}, \quad \mathbf{a}\cdot\mathbf{k} = \mathbf{b}\cdot\mathbf{k} = \mathbf{a}\cdot\mathbf{b} = 0,$$

$$\mathbf{a}^2 = \mathbf{b}^2.$$

When $\nu \neq 0$, the solution is readily obtained by multiplying (3) by the factor $\exp(-\nu k^2 t)$ [20].

It can easily be verified (see also [4]) that waves described by (3) satisfy the relation

$$\mathrm{curl}\mathbf{u} = \pm k\mathbf{u},$$

which is characteristic of Gromeka–Beltrami flows since (3) entails

$$\mathbf{b} = \pm\frac{1}{k}[\mathbf{k}\times\mathbf{a}], \quad \mathbf{a} = \mp\frac{1}{k}[\mathbf{k}\times\mathbf{b}].$$

Here, minus corresponds to negative helicity,

$$H = -k\mathbf{u}^2/2,$$

i.e., right-handed rotation of the velocity vector [4]. Vice versa, positive helicity is associated with left-handed rotation of a velocity vector of constant magnitude. It was also noted in [4] that motion of this kind is of interest in magnetic dynamo theory. If $\omega = 0$, then (3) is an exact solution to the equations describing steady inviscid incompressible flow [8].

The solution to Eq. (2) satisfying the condition

$$\mathrm{curl}\mathbf{u} = \pm k\mathbf{u}$$

can be represented in a more general form:

$$\mathbf{u} = \exp(-\nu k^2 t)\mathrm{Re}\mathbf{A}f(\tilde{z}, x, y), \quad \tilde{z} = z \pm \frac{2\Omega_z t}{k},$$

$$f = \tilde{f}(\xi)\exp\left[-\frac{k}{A_y}(A_x\tilde{z} - A_z x)\right], \quad (4)$$

$$\xi = y + \frac{A_z\tilde{z} + A_x x}{A_y},$$

where $\tilde{f}$ is an arbitrary function. In particular, if

$$f = \exp[\pm i(k_z\tilde{z} + k_x x + k_y y)]$$

then solution (4) reduces to (3). Note also that the argument of $\tilde{f}$ at $t = 0$ equals that of an arbitrary function of the form

$$\Omega_0(k_1(x + iy) + k_2(x - iy) + k_3 z),$$

which was used in [22, 23] to describe steady helical vortex structures in a ferromagnet, because it is assumed that

$$4k_1k_2 + k_3^2 = 0.$$

Thus, solution (4) provides a generalized representation of inertia waves, which can be interpreted as Gromeka–Beltrami flows. Moreover, expression (4) also describes the general form of helical structures arising in various natural systems and phenomena. Note also that an arbitrary sign of helicity implies that the corresponding dispersion equation should be written as follows (in contrast to [18] and (3)):

$$\omega = \pm\frac{2k_z\Omega_z}{k},$$

where plus is associated with negative helicity. A similar representation of ω in (3) was given in [20], but without allowing for the correlation between the signs of ω and helicity (see also [4]).

3.2. The representations of linear uniform helical waves written out above are amenable to the superposition principle. This is not true with regard to exact solutions to the nonlinear Helmholtz equation that describe Joukowski-type flows, with

$$\mathrm{curl}\mathbf{u} + 2\Omega = k\mathbf{u}, \quad \mathbf{\Omega} \neq 0.$$

The exact solution to the time-dependent equations of inviscid incompressible fluid dynamics that describes a Joukowski-type flow can be represented as

$$\mathbf{u} = \frac{2\Omega}{k} + \tilde{\mathbf{u}}, \quad \mathrm{curl}\tilde{\mathbf{u}} = \mathrm{curl}\mathbf{u} = k\tilde{\mathbf{u}}, \quad (5)$$

$$\tilde{\mathbf{u}} = \mathrm{curlcurl}\boldsymbol{\varphi} + k\mathrm{curl}\boldsymbol{\varphi}, \quad \Delta\boldsymbol{\varphi} + k^2\boldsymbol{\varphi} = 0. \quad (6)$$

Solutions satisfying (5) and (6) with $\mathbf{\Omega} = 0$ and $\boldsymbol{\varphi} = \mathbf{x}\phi(\mathbf{x})$ were considered in [24] for a localized spherical vortex that is structurally stable (unlike Hill's spherical vortex [21]) since it is characterized by a finite helicity and a corresponding topological invariant. It was noted in [24] that helical vortex flows of this kind can be of interest with regard to magnetic plasma confinement. The flow fields generated by these spherical helical vortices outside their localization domains are identical to those of point vortex dipoles [25]. Therefore, they can be used to regularize the vortex-particle methods employed in 3D numerical simulations of vortex bursting in turbulent boundary layers. This can be done by using the fact (proved in [25]) that there exists an exact weak solution to the equations of inviscid incompressible fluid dynamics represented as a finite-dimensional Hamiltonian dynamical system describing three-

dimensional point vortex dipoles. (They can be interpreted as infinitesimal vortex rings or generalized Hill vortices with nonzero helicity.)

Here, an analogous exact solution to the fluid-dynamics equations is obtained for a cylindrically symmetric vortex of radius R characterized by nontrivial topologically invariant structure associated with nonzero value of H. As $R \longrightarrow 0$, the vortex reduces to a generalized vortex line with $H \neq 0$ [21], whereas the classical Helmholtz vortex line is obtained when $H = 0$. Cylindrical helical vortex solutions are more attractive as models of tornado-like vortices having stable complicated vortex cores, as compared to the Helmholtz vortex line [26].

In the axially symmetric case, expression (5) is the following exact time-independent solution to the Helmholtz equation for a vortex localized in a cylindrical domain of radius R:

$$u_z(r) = \left[\frac{2\Omega_z}{k} + BJ_0(kr)\right]\Theta(R-r) + v_{0z}\Theta(r-R),$$

$$\tilde{u}_\varphi(r) = BJ_1(kr)\Theta(R-r) + \left(\frac{\chi}{r} - \Omega_z r\right)\Theta(r-R), \quad (7)$$

$$\tilde{u}_r(r) = 0, \quad \Theta(p) = \begin{cases} 1, & p \geq 0, \\ 0, & p < 0, \end{cases}$$

$$\begin{aligned} \frac{2\Omega_z}{k} + BJ_0(kR) &= v_{z0}, \\ BJ_1(kR) &= \frac{\chi}{R} - \Omega_z R, \end{aligned} \quad (8)$$

where J_0 and J_1 are the first- and second-order Bessel functions, and conditions (8) ensure continuous matching of the velocity fields inside and outside the cylinder. The solution defined by (7) and (8) is obtained by analogy with the analysis performed in [24] to determine the structure of a spherical helical vortex. Unlike the vortex considered in [24], the vortex described by (7) is at rest in the laboratory frame and does not induce any velocity field analogous to that associated with a point Helmholtz vortex or a vortex filament. Outside the domain of vorticity localization (at $r > R$), u_φ is identical to the velocity field induced such a point vortex characterized by the circulation χ determined by (8). This relation holds in the laboratory frame in which the cylindrical helical vortex executes a solid-body rotation about its symmetry axis with frequency Ω_z. Note that a point Helmholtz vortex can be interpreted as the limit form vortex filament obtained as $R \longrightarrow 0$ by assuming zero axial velocity v_{z0} and uniform vorticity distribution, i.e.,

$$\omega_z = 2\Omega_z = \frac{\tilde{\chi}}{\pi R^2},$$

inside the vortex core of radius R, with $\tilde{\chi} = \text{const}$ and $\omega_z \longrightarrow \infty$. This vortex is characterized by zero helicity. The solution presented above has a nonzero helicity, and the corresponding vorticity distribution inside the vortex core is not uniform, because the vorticity components associated with **u** defined by (7) are

$$\omega_z + 2\Omega_z = ku_z, \quad \omega_\varphi = k\tilde{u}_\varphi, \quad \omega_r = 0.$$

Let us consider some characteristics of solution (7) under various particular conditions (8).

Conditions (8) with $k = \gamma_{1n}/R$ (where γ_{1n} are the zeros of J_1) imply that the circulation is

$$\chi = \Omega_z R^2,$$

i.e., differing only by a factor of 2π from $\tilde{\chi}$ for the limit form of the vortex filament obtained as $R \longrightarrow 0$ and $\Omega_z \longrightarrow \infty$ and has the circulation of the point Helmholtz vortex. In this case, the coefficient B has the specific form

$$B = B_n = \frac{v_{z0} - 2\Omega_z R/\gamma_{1n}}{J_0(\gamma_{1n})},$$

which follows from (8). The corresponding cylindrical helical vortex, inducing the vortex-filament velocity field at $r > R$, has a stable topological structure determined by the helicity value. When $n > 1$, the flow field inside the vortex core (at $r < R$) can have a very complicated structure involving both updrafts and downdrafts with opposite circulation signs analogous to the updrafts and downdrafts observed in tornadic and dust-devil flow structures [26].

Note that the solution obtained from (7) at $r \leq R$ for the flow in a bounded rotating vessel of radius R was used, for example, in [27] to describe the bottom drain flow out of the vessel.

On the other hand, there exists a solution defined at $r \geq 0$ for arbitrary k that is identical to (7) at $r \leq R$ for $\Omega = 0$ [8]. However, vorticity is not localized at $r \leq R$ in this solution, as it is in the radially nonuniform cylindrical generalization of the classical Helmholtz vortex [21] described by (7) and (8). In contrast to [8, 27], the solutions defined by (7) and (8) are obtained by matching a vortex flow localized at $r \leq R$ with a potential flow at $r > R$. Therefore, Eqs. (7) and (8) define a new exact solution describing a localized cylindrical vortex having a nontrivial topological structure whose stability, as in the case of spherical vortex [24], is due to the existence of a corresponding topological invariant, i.e., helicity [2, 3].

When $kR \neq \gamma_{1n}$ and $B \neq 0$ in (8), the sign of the circulation χ determining the velocity field at $r > R$ associated with a cylindrical helical vortex may differ from that of Ω_z (in contrast to the case of $kR = \gamma_{1n}$), because

$$\chi = \frac{R^2 J_2(kR)}{J_0(kR)}\left(\frac{v_{0z}J_1(kR)}{RJ_2(kR)} - \Omega_z\right).$$

Indeed, when kR is held constant, the sign of χ changes at supercritical rotation frequencies,

$$\Omega_z > \Omega_{cr}^0 = \frac{v_{0z}J_1(kR)}{RJ_2(kR)}, \quad J_2(p) = \frac{2J_1(p)}{p} - J_0(p). \quad (9)$$

Therefore, when (9) holds and $\Omega_{cr}^0 > 0$, nonmonotonic behavior at $r > R$ can be exhibited by the azimuthal velocity

$$\tilde{u}_\varphi = -\Omega_z r - \frac{(\Omega_z - \Omega_{cr}^0)R^2 J_2(kR)}{rJ_0(kR)}$$

as a function of the distance r from the vortex centerline, since the absolute value of $\tilde{u}_\varphi$ reaches a minimum value,

$$|\tilde{u}_\varphi| = 2(|\chi|\Omega_z)^{1/2}$$

at

$$r = r_m = \left(\frac{|\chi|}{\Omega_z}\right)^{1/2}.$$

Furthermore, the necessary condition $r_m > R$ leads to an additional upper bound for Ω_z:

$$\Omega_z < \Omega_{cr}^1 = \frac{\Omega_{cr}^0}{1 - J_0(kR)/J_2(kR)}, \quad (10)$$

where

$$0 < \frac{J_0(kR)}{J_2(kR)} < 1 \text{ at } \Omega_{cr}^0 > 0.$$

Note that the existence of a minimum of the absolute value of $\tilde{u}_\varphi$ as a function of r at $r = r_m$ in a coordinate system rotating with the fluid implies that the tracer concentration advected by the vorticity field builds up in the neighborhood of $r = r_m$. Indeed, an annular distribution of this kind was observed in [28] for suspensions in rotating vessels in a narrow range of Ω_z, in qualitative agreement with (9) and (10).

Thus, the present helical generalization of the Helmholtz vortex (parameterized by k and Ω_z) can be used to describe qualitatively different variations of the $\tilde{u}_\varphi$ field outside the vortex core (at $r > R$). If $k = \gamma_{1n}/R$ or $k \neq \gamma_{1n}/R$ and $\Omega < \Omega_{cr}^0$ or $\Omega > \Omega_{cr}^1$ in the latter case, then the variation of u_φ with r increasing from $r = R$ is similar to the dependence of $\tilde{u}_\varphi$ on r for the Helmholtz vortex (in a coordinate system rotating with frequency Ω_z). If $k \neq \gamma_{1n}/R$ and $\Omega_{cr}^1 < \Omega < \Omega_{cr}^0$, then the nonmonotonic behavior of $\tilde{u}_\varphi(r)$ indicated above may be responsible for localization of a passive scalar in the neighborhood of $r = r_m$ associated with $\min|\tilde{u}_\varphi(r)|$.

4. EVOLUTION OF HELICITY IN UNIFORM HELICAL FLOWS

Representation (7) of the velocity field at $r < R$ can be obtained by using (5) and averaging over the azimuthal angle the well-known solution for the unconfined *ABC* flow [8, 9, 12, 13],

$$\begin{aligned} \tilde{u}_x &= A\sin kz + C\cos ky, \\ \tilde{u}_y &= B\sin kx + A\cos kz, \\ \tilde{u}_z &= C\sin ky + B\cos kx. \end{aligned} \quad (11)$$

In what follows, a time-dependent generalization of (11) is derived in the context of an analysis of the possibility of helicity generation by taking into account the effects described by introducing time-dependent A, B, and C for an unsteady Joukowski-type helical flow regime. Since only the coefficient B is retained in (11) in the azimuthally symmetric case, the analysis below is developed for $A = C = 0$ and B evolves in time according to the helicity balance equation with $\nu \neq 0$, $\dot{\mathbf{\Omega}} \neq 0$, and $T \neq 0$.

4.1. Viscous Dissipation and Local Helicity Generation

The time evolution of the helicity

$$H = \frac{\boldsymbol{\omega} \cdot \mathbf{u}}{2}$$

for a flow satisfying (5) and (11) is analyzed here by using vorticity equation (A.2) with $\nu \neq 0$, $\dot{\mathbf{\Omega}} \neq 0$, and $T \neq 0$:

$$\begin{aligned} \frac{\partial H}{\partial t} &= -\frac{1}{k}\frac{d}{dt}\mathbf{\Omega}^2 - (\tilde{\mathbf{u}} \cdot \dot{\mathbf{\Omega}}) - \nu k^2(\tilde{\mathbf{u}}^2 k + \mathbf{\Omega} \cdot \tilde{\mathbf{u}}) \\ &+ \frac{\beta g}{k}(\varepsilon_{ij3}\cos\alpha + \varepsilon_{ij2}\sin\alpha)(\Omega_i + k\tilde{u}_i)\frac{\partial T}{\partial x_j}, \end{aligned} \quad (12)$$

$$H = \frac{k\tilde{\mathbf{u}}^2}{2} + \mathbf{\Omega} \cdot \tilde{\mathbf{u}}, \quad (13)$$

where the velocity field $\tilde{\mathbf{u}}$ corresponds to a flow described by (11). Equation (12) is derived without using any additional assumptions (other than those underlying starting equations (A.1)), because it is sufficient to use a time-dependent generalization of (5) for the velocity and vorticity fields. In particular, Eqs. (12) and (13) imply that local helicity (13) can be generated by allowing for viscous dissipation in (12) with $\dot{\mathbf{\Omega}} = 0$ ($\nu \neq 0$) even if $T = 0$ and $\mathbf{\Omega} \neq 0$. Indeed, if $k < 0$, then Eq. (12) predicts an increase in H provided that $H > 0$ and the sum

$$\tilde{\mathbf{u}}^2 k + \mathbf{\Omega} \cdot \tilde{\mathbf{u}}$$

is negative by virtue of the following inequalities:

$$\frac{|k|\tilde{u}}{\Omega} > \cos\Psi > \frac{|k|\tilde{u}}{2\Omega}, \quad (14)$$

where

$$\Omega = |\mathbf{\Omega}|, \quad \tilde{u} = |\tilde{\mathbf{u}}|, \quad \cos\Psi = \frac{\mathbf{\Omega}\cdot\tilde{\mathbf{u}}}{\Omega u}.$$

Inequality (14) is satisfied when Ψ has a real value only in the case of sufficiently fast (supercritical) solid-body rotation of the system,

$$\Omega > \Omega_{cr} = \frac{|k|\tilde{u}}{2}. \quad (15)$$

Inequality (15) is analogous to the criterion for dissipative–centrifugal instability obtained in [29] (see discussion of its occurrence in upper atmospheric layers in [30]).

It follows from (14) that, if $2\Omega_{cr} > \Omega > \Omega_{cr}$, there exists only a lower bound for the value of $\cos\Psi$ dictated by the physical parameters of the system (the only upper bound is determined by the inequality $\cos\Psi < 1$). Under this condition, the required effect is more likely to occur than if $\Omega > 2\Omega_{cr}$, in which case the left-hand side of (14) restricts the domain of admissible Ψ. Note that the condition for Ω in which the lower bound for frequency is half the upper one is consistent with the criterion for the formation of an isolated tornado-like vortex in the experiment described in [31]. Indeed, according to [31], this regime corresponds to the following interval of the rotational Peclet number:

$$3 < \mathrm{Pe}_\Omega < 20, \quad \mathrm{Pe}_\Omega \approx 0.63\left(\frac{\mathrm{Ra}}{\mathrm{Ta}}\right)^2,$$

where Ra is the Rayleigh number and

$$\mathrm{Ta} = \frac{4\Omega^2 D^4}{\nu^2}$$

is a dimensionless parameter associated with the solid-body rotation frequency Ω (D is the diameter of an isolated source of buoyancy). The Peclet-number interval indicated above is associated with a rotation-frequency range:

$$\Omega_1 < \Omega < \Omega_2, \quad \frac{\Omega_2}{\Omega_1} \approx 1.6.$$

Thus, the conditions for rotation frequency that correspond to the isolated intense helical vortices observed in [31] are in fair agreement with the criterion for local helicity generation resulting from dissipative instability, since

$$\frac{\Omega_{max}}{\Omega_{min}} = 2.$$

Let us compare the value of $\Omega_{min} = \Omega_{cr}$ given by (14) and (15) with the value of $\Omega_1 \approx 0.21(\varepsilon/\nu)^{1/2}$ (see [31]), where

$$\varepsilon \approx \frac{\nu}{2}\left(\left(\frac{\partial u_i}{\partial x_j}\right)^2 + \left(\frac{\partial u_j}{\partial x_i}\right)^2\right)$$

is the rate of kinetic energy dissipation per unit mass. If $\Omega_{cr} = \Omega_1$, then k is estimated as

$$k \approx \frac{0.41}{\tilde{u}}\left(\frac{\varepsilon}{\nu}\right)^{1/2}.$$

By the definition of ε, if $\tilde{u} \approx |\mathbf{u}|$, then

$$k \approx \frac{0.82|\mathrm{curl}\,\tilde{\mathbf{u}}|}{\tilde{u}}$$

provides an estimate valid for velocity field (11). This means that not only the ratio $\Omega_{max}/\Omega_{min}$ that follows from (14) and Ω_2/Ω_1 from [31] are in agreement, but so are the lower bounds Ω_{cr} and Ω_1 as well. The agreement suggests that the observations of isolated helical vortices reported in [31] can be interpreted by taking into account interaction between solid-body rotation and viscous dissipation.

When $A = C = 0$, it is obvious that $\tilde{u}$ in (14) and (15) is independent of coordinates since $\tilde{u} = B$ according to (11). The following estimate can be obtained for nonzero A, C, B:

$$\max_{\mathbf{x}}\tilde{u} \le |A| + |B| + |C| = \tilde{u}_m,$$

in which case the upper bound for Ω_{cr} is

$$\overline{\Omega}_{cr} = \frac{|k|\tilde{u}_m}{2}.$$

4.2. Generation of Integral Helicity in Nonuniform Temperature Fields

Now, consider helicity balance equation (12) in the case when temperature fluctuations determine the evolution of the integral helicity

$$\overline{H} \equiv \left(\frac{2\pi}{k}\right)^3 \int_0^{2\pi/k} dx \int_0^{2\pi/k} dy \int_0^{2\pi/k} dz H.$$

For simplicity, only constant temperature gradients are taken into account:

$$\frac{\partial T}{\partial x_j} = A_j.$$

In this case, (12) yields the following balance equation

for integral helicity:

$$\frac{d\bar{H}}{dt} = -\frac{1}{k}\frac{d}{dt}\boldsymbol{\Omega}^2 - \nu k^3\tilde{\bar{u}}^2 + \frac{\beta g}{k}(\varepsilon_{ij3}\cos\alpha + \varepsilon_{ij2}\sin\alpha)\Omega_i A_j, \quad (16)$$

where

$$\bar{H} = \frac{k\tilde{\bar{u}}^2}{2} = \frac{k(A^2 + B^2 + C^2)}{2}$$

for arbitrary A, B, and C. According to (16), integral helicity (in contrast to local helicity) cannot be generated when $k < 0$, since $\bar{H}$ has a negative value in this case and its absolute value decreases with time because of viscous dissipation.

However, Eq. (16) implies that effects due to buoyancy can be responsible for production of $\bar{H}$ when $\boldsymbol{\Omega} \neq 0$ even if the rotation frequency does not vary with time. In particular, it follows from (16) that the derivative $d\bar{H}/dt$ is positive when $\boldsymbol{\Omega} \neq 0$ at the initial moment if the following inequality holds for $\boldsymbol{\Omega} = (0, 0, \Omega_z)$:

$$\frac{\Omega_z \beta g A_x \sin\alpha}{\nu k^4 (A_0^2 + B_0^2 + C_0^2)} > 1, \quad (17)$$

where A_0, B_0, and C_0 denote the values of A, B, and C at $t = 0$, respectively. Thus, if $\alpha > 0$ and the rotation frequency is constant, then (17) implies that there exists a critical value of horizontal temperature gradient,

$$A_{\mathrm{cr}} = \frac{\nu k^4 (A_0^2 + B_0^2 + C_0^2)}{\beta g \Omega_z \sin\alpha},$$

such that the integral helicity increases with time at the initial stage of the evolution of $\bar{H}$ for $A_x > A_{\mathrm{cr}}$. Indeed, it was noted in [26] that dust devils develop only in the presence of sufficiently steep horizontal temperature gradients. Furthermore, when $\alpha \longrightarrow 0$ (the underlying surface is flat), it follows from (16) and (17) that integral helicity can be generated either if the rotation frequency varies substantially with time (see above) or if the velocity of solid-body rotation has horizontal components. The latter condition can be satisfied for convective rolls.

4.3. Regimes of Local Helicity Generation in Nonuniform Temperature Fields

Consider balance equation (12) for local helicity in the special case when $A = C = 0$. For simplicity, suppose also that only the component A_x of a constant temperature gradient is nonzero:

$$\frac{\partial T}{\partial y} = \frac{\partial T}{\partial z} = 0.$$

These conditions correspond to a time-dependent solution to the heat equation without source terms, because $\tilde{u}_x = 0$ if $A = C = 0$.

The equations describing the evolution of $B(t)$, $\Omega_z(t)$, and $\Omega_y(t)$ are derived in the Appendix. Note that $\dot{\Omega}_x = 0$ since

$$\frac{\partial T}{\partial y} = \frac{\partial T}{\partial z} = 0$$

and assume that $\Omega_x = 0$ to simplify the analysis that follows. In the Appendix, it is shown that two qualitatively different systems of evolution equations may be valid, depending on whether or not the relation

$$\Omega_z^2 + \Omega_y^2 = k^2 B^2 \quad (18)$$

for the helical-vortex parameters B, Ω_z, and Ω_y flow holds.

When (18) is not satisfied,

$$\dot{B} = -\nu k^2 B, \quad \dot{\Omega}_y = -\beta g \frac{A_x}{2}\cos\alpha, \quad \dot{\Omega}_z = \frac{\beta g A_x \sin\alpha}{2}. \quad (19)$$

This set of equations is easily solved, and (13) is used to obtain $H \longrightarrow 0$ at $t \gg 1/\nu k^2$. Equations for Ω_y and Ω_z in (19) are identical to those (written out in [15] in the Boussinesq approximation) for the angular velocities of a degenerate fluid ellipsoid (spheroid) executing a solid-body rotation in a gravity field under a constant temperature gradient. However, the value of A_x in (19) can be treated as constant at any instant if $A_y = A_z = 0$, whereas this is possible in the context of [15] only at the initial moment, because both A_y and A_z are nonzero at $t > 0$ if $\Omega_z \neq 0$, $\Omega_y \neq 0$, and $A_x \neq 0$. Next, we consider the case when relation (18) holds.

In the Appendix, the following nonlinear dynamical system is derived for $B^2 = q$, Ω_z, and Ω_y:

$$\dot{q} = -\frac{2}{3}\nu k^2 q + \frac{2\beta g A_x}{3k^2}(\Omega_z \sin\alpha - \Omega_y \cos\alpha),$$

$$\dot{\Omega}_z = -\frac{\nu k^2 \Omega_z}{3} - \frac{\beta g A_x \Omega_z}{6k^2 q}(\Omega_z \sin\alpha - \Omega_y \cos\alpha) + \frac{\beta g A_x \sin\alpha}{2}, \quad (20)$$

$$\dot{\Omega}_y = -\frac{\nu k^2 \Omega_y}{3} - \frac{\beta g A_x \Omega_y}{6k^2 q}(\Omega_z \sin\alpha - \Omega_y \cos\alpha) - \frac{\beta g A_x \cos\alpha}{2}.$$

Under condition (18), system (20) admits general solution (A.9) expressed in terms of elliptic integrals, because there exists the invariant

$$h_0 = \exp(\nu k^2 t)\left(\frac{q}{q_0}\right)^{1/2} \times\left[\frac{q}{\nu^2 k^2} - \frac{(\Omega_z \sin\alpha - \Omega_y \cos\alpha)^2}{\nu^2 k^4}\right]. \quad (21)$$

In particular, the solution to system (20) corresponding to $h_0 = 0$ is

$$B(t) = \exp\left(-\frac{\nu k^2 t}{3}\right) \times\left(B(0) + \frac{\beta g A_x}{\nu k^3}\left(\exp\left(\frac{\nu k^2 t}{3}\right) - 1\right)\right),$$
$$\Omega(t) = \Omega_z \sin\alpha - \Omega_y \cos\alpha = \exp\left(-\frac{\nu k^2 t}{3}\right) \times\left(\Omega(0) + \frac{\beta g A_x}{\nu k^3}\left(\exp\left(\frac{\nu k^2 t}{3}\right) - 1\right)\right). \quad (22)$$

As $t \longrightarrow \infty$, solution (22) describes a stable time-independent flow regime described by system (20):

$$B_0 = \frac{\beta g A_x}{\nu k^3}, \quad \Omega_{y0} = -\frac{\beta g A_x}{\nu k^2}\cos\alpha, \quad \Omega_{z0} = \frac{\beta g A_x \sin\alpha}{\nu k^2} \quad (23)$$

(which can be derived from the general solution to (A.9) as $t \longrightarrow \infty$). The time-independent integral helicity corresponding to (23),

$$\overline{H} = \frac{k}{2\pi}\int_0^{2\pi/k} dx H$$

(where H is given by (A.4) when $A = C = 0$), can be written as

$$\overline{H}_0 = \frac{(\beta g A_x)^2}{2\nu^2 k^5}. \quad (24)$$

Thus, the time-independent regime characterized by (24) corresponds to a large limit value of integral helicity when a constant temperature gradient is perpendicular to the gravity force.

Note that nonlinear equations (20) for Ω_z and Ω_y substantially differ from those obtained in [15] for the angular velocities of solid-body rotation of a degenerate fluid ellipsoid (spheroid), even though the solution to (A.9) is also expressed in terms of elliptic integrals in the general case.

It would be interesting to construct an analog of systems of hydrodynamic type not only by analyzing solid-body rotation of a fluid (as in [15]), but also by taking into account the helical structure of the Joukowski-type flows examined here.

5. CONCLUSIONS

Examples are given of uniform helical vortex flow regimes that develop via relatively simple mechanisms in a fluid rotating as a whole. In particular, a new exact representation is obtained for uniform helical vorticity fields, which generalizes both a well-known solution having the form of traveling inertia waves in a fluid rotating as a whole [18] and a helical solution used in [22, 23] to describe steady helical vortex structures in a ferromagnet. It is shown that not only the pure solid-body rotation regimes examined in [15], but also uniform helical regimes of fluid rotation, can be described by relatively simple dynamical system (20), including the limit of $t \longrightarrow \infty$ (see (23) and (24)). To obtain the dynamical system, time-dependent uniform helical vortex flow regimes are analyzed, instead of the commonly considered steady uniform helical vortex flow fields [8, 9, 12]. An analysis of the time-dependent generalization makes it possible to allow for dissipative and thermodynamic effects (reflected in the starting Eqs. (A.1)) in the evolution of local and integral helical vorticity perturbations.

The new exact solution defined by (7) and (8) describes a localized cylindrical vortex having a nontrivial stable topological structure determined by the value of helicity. This solution is qualitatively different from the spherical vortex considered in [24] and is a generalization of the Helmholtz vortex filament. It can be used to model natural tornadic vortices, which are actually characterized by nonuniform vortex structure.

The results obtained here can be used to solve magnetic dynamo problems, which are seldom analyzed by invoking the concept of uniform helical vortex flow [32, 33]. Their existence under geophysical conditions is questioned in [33, 21], probably because their occurrence even in well-known inertia waves (see above) is not taken into account.

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research, project nos. 01-05-64300 and 01-07-90211.

APPENDIX

1. The equations for a velocity field in a coordinate system rotating with frequency $\mathbf{\Omega}$ are written in the Boussinesq approximation as

$$\frac{\partial u_i}{\partial t} + u_l\frac{\partial u_i}{\partial x_l} + \varepsilon_{ijl}(2\Omega_j u_l + \dot{\Omega}_j x_l)$$

$$= -\frac{1}{\rho_0}\frac{\partial p'}{\partial x_i} + \nu\Delta u_i + \beta g T(\delta_{i3}\cos\alpha + \delta_{i2}\sin\alpha), \quad \text{(A.1)}$$

$$\mathrm{div}\mathbf{u} = 0, \quad P' = P - \frac{\rho_0}{2}(\mathbf{\Omega}^2\mathbf{x}^2 - (\mathbf{\Omega}\cdot\mathbf{x})^2).$$

The corresponding equation for vorticity $\boldsymbol{\omega} = \mathrm{curl}\mathbf{u}$ is

$$\frac{\partial(\omega_i + (n-1)\Omega_i)}{\partial t} + u_l\frac{\partial\omega_i}{\partial x_l} = (\omega_l + 2\Omega_l)\frac{\partial u_i}{\partial x_l} + \nu\Delta\omega_i + \beta g(\varepsilon_{ij3}\cos\alpha + \varepsilon_{ij2}\sin\alpha)\frac{\partial T}{\partial x_j}. \quad \text{(A.2)}$$

In (A.1) and (A.2), summation from 1 to n over repeated indices is assumed, ε_{ijl} is the Levi-Civita permutation symbol, and n is the dimensionality of the space ($n = 2, 3$).

The equation for the helicity field

$$H = \frac{\boldsymbol{\omega}\cdot\mathbf{u}}{2},$$

that corresponds to (A.1) and (A.2) for $n = 3$ is

$$2\frac{\partial H}{\partial t} + \omega_i u_l\frac{\partial u_i}{\partial x_l} + u_i u_l\frac{\partial\omega_i}{\partial x_l} + \omega_i\varepsilon_{ijl}(\dot{\Omega}_j x_l + 2\Omega_j u_l)$$

$$+ 2u_i\dot{\Omega}_i = -\frac{1}{\rho_0}\frac{\partial p'}{\partial x_i}\omega_i + (\omega_l + 2\Omega_l)u_i\frac{\partial u_i}{\partial x_l} \quad \text{(A.3)}$$

$$+ \nu(\omega_i\Delta u_i + u_i\Delta\omega_i) + \beta g(\omega_3\cos\alpha + \omega_2\sin\alpha)T$$

$$+ (\varepsilon_{ij3}\cos\alpha + \varepsilon_{ij2}\sin\alpha)u_i\frac{\partial T}{\partial x_j}\beta g.$$

Since both $\boldsymbol{\omega}$ and $\mathbf{u}$ are supposed to vanish at the boundaries of a volume V, it follows from (A.3) that the integral helicity

$$\overline{H} = \int_V d^3x H$$

is governed by balance equation (1) with

$$\omega_z = \omega_3, \quad \omega_y = \omega_2, \quad \omega_x = \omega_1.$$

When the effects due to bottom friction are taken into account, the right-hand sides of (A.1), (A.2), and (A.3) contain the additional terms

$$-2\alpha(u_i - \delta_{i3}u_3),$$

$$-2\alpha\left(\omega_i - \varepsilon_{ij3}\frac{\partial u_3}{\partial x_j}\right),$$

and

$$-2\alpha\left(2H - \omega_3 u_3 - \varepsilon_{ij3}u_i\frac{\partial u_3}{\partial x_j}\right),$$

respectively, while (1) contains the term

$$-2\alpha(\overline{H} - \overline{\omega_3 u_3}).$$

If $\Omega_z \equiv \Omega_3 \neq 0$, then

$$\alpha = \sqrt{\frac{\nu\Omega_z}{h}},$$

and if $\Omega = 0$, then

$$\alpha = \frac{\nu}{h^2},$$

where h is the fluid-layer thickness. The case of $\alpha \neq 0$ is not considered in the present study.

2. If

$$A = C = 0, \quad \frac{\partial T}{\partial y} = \frac{\partial T}{\partial z} = 0,$$

then expression (13) for H yields

$$H = \frac{kB^2}{2} + (\Omega_z\cos kx + \Omega_y\sin kx)B. \quad \text{(A.4)}$$

When (A.4) is substituted into (12), the resulting equation is satisfied for $A = C = 0$ by setting to zero the term that does not contain trigonometric functions and the groups that multiply $\sin kx$ and $\cos kx$. This leads to the following system of equations:

$$kB\dot{B} = -\frac{2\Omega_z\dot{\Omega}_z}{k} - \frac{2\Omega_y\dot{\Omega}_y}{k} - \nu k^3B^2 + \frac{\beta g}{k}\frac{\partial T}{\partial x}(\Omega_z\sin\alpha - \Omega_y\cos\alpha),$$

$$\Omega_z\dot{B} + \dot{\Omega}_z B = -\dot{\Omega}_z B - \nu k^2\Omega_z B + \beta g B\frac{\partial T}{\partial x}\sin\alpha, \quad \text{(A.5)}$$

$$\Omega_y\dot{B} + \dot{\Omega}_y B = -\dot{\Omega}_y B - \nu k^2\Omega_y B - \beta g B\frac{\partial T}{\partial x}\cos\alpha.$$

System (A.5) with

$$\frac{\partial T}{\partial x} = A_x, \quad B^2 = q,$$

can be transformed into

$$\begin{aligned} a_1\dot{\Omega}_z + b_1\dot{\Omega}_y &= c_1(\Omega_z, \Omega_y) \\ &= \frac{2\beta g A_x}{k^2}\left(\Omega_z\sin\alpha - \Omega_y\cos\alpha - \frac{k^2 q}{\Omega_z}\sin\alpha\right), \\ a_2\dot{\Omega}_z + b_2\dot{\Omega}_y &= c_2(\Omega_z, \Omega_y) \\ &= \frac{2\beta g A_x}{k^2}\left(\Omega_z\sin\alpha - \Omega_y\cos\alpha + \frac{k^2 q}{\Omega_z}\cos\alpha\right), \end{aligned} \tag{A.6}$$

where

$$a_1 = \frac{4(\Omega_z^2 - qk^2)}{\Omega_z k^2}, \quad a_2 = \frac{4\Omega_z}{k^2},$$

$$b_1 = \frac{4\Omega_y}{k^2}, \quad b_2 = \frac{4(\Omega_y^2 - k^2 q)}{\Omega_y k^2}.$$

It follows from (A.6) that

$$\dot{\Omega}_y = \frac{c_1 - a_1\dot{\Omega}_z}{b_1}$$

and

$$\dot{\Omega}_z(a_2 b_1 - a_1 b_2) = c_2 b_1 - b_2 c_1, \tag{A.7}$$

where

$$a_2 b_1 - a_1 b_2 = 16q\left(\frac{\Omega_z^2 + \Omega_y^2 - k^2 q}{k^2\Omega_z\Omega_y}\right),$$

$$c_2 b_1 - b_2 c_1 = \frac{\beta g A_x}{2}(a_2 b_1 - a_1 b_2)\sin\alpha.$$

If relation (18) holds, then system (A.6) is degenerate and Eq. (A.7) is an identity for any $\dot{\Omega}_z$ satisfying (18). If relation (18) is violated, then system (19) can be derived by combining (A.6) and (A.7).

If (18) holds, then (A.5) yields system (20), which is conveniently represented in dimensionless form:

$$\begin{aligned} &\frac{d\bar{v}}{d\tau_1} = \gamma\bar{u}, \quad \frac{d\bar{u}}{d\tau_1} = \frac{3}{4}\gamma_1 - \frac{\gamma\bar{u}^2}{4\bar{v}}, \\ &\tau_1 = 2(e^{\tau/2} - 1), \quad \tau = \frac{2}{3}t\nu k^2, \end{aligned} \tag{A.8}$$

where

$$u = \bar{u}e^{-\tau/2}, \quad v = \bar{v}e^{-\tau}, \quad u = \frac{\Omega}{\nu k^2},$$

$$v = \frac{q}{q(0)}, \quad \Omega = \Omega_z\sin\alpha - \Omega_y\cos\alpha,$$

$$\gamma_1 = \frac{\beta g A_x}{(\nu k^2)^2}, \quad \gamma = \frac{\beta g A_x}{k^2 q(0)}.$$

System (A.8) has preserves invariant (21), which can be written as

$$h_0 = \bar{v}^{3/2}\frac{\gamma_1}{\gamma} - \bar{u}^2\bar{v}^{1/2}.$$

This invariant can be used to integrate system (A.8) in quadratures to obtain

$$\begin{aligned} \gamma\tau_1 + C &= \frac{2}{h_0^{1/2}}\left(\frac{h_0}{p}\right)^{5/6}\left[\frac{(\sqrt{3}+1)(p^4 - p)^{1/2}}{(\sqrt{3}+1)p - 1}\right. \\ &\left. + 3^{1/4}E(\varphi, m) - \frac{\sqrt{3}-1}{2}3^{-1/4}F(\varphi, m)\right], \\ \varphi &= \arccos\left(-\frac{p(\sqrt{3}-1)+1}{p(\sqrt{3}+1)-1}\right), \\ m &= \sin\frac{5\pi}{12}, \end{aligned} \tag{A.9}$$

where C is an integration constant, and F and E denote the elliptic integrals of the first and second kinds, respectively:

$$E = \int_0^\varphi d\varphi\sqrt{1 - m^2\sin^2\varphi}, \quad F = \int_0^\varphi \frac{d\varphi}{\sqrt{1 - m^2\sin^2\varphi}},$$

$$p = \bar{v}^{1/2}\left(\frac{\gamma_1}{\gamma h_0}\right)^{1/3},$$

$$h_0 = \frac{(\Omega_z(0)\cos\alpha + \Omega_y(0)\sin\alpha)^2}{\nu^2 k^4}.$$

REFERENCES

1. E. A. Novikov, Izv. Akad. Nauk SSSR, Fiz. Atmos. Okeana **8**, 759 (1972).
2. H. K. Moffat, J. Fluid Mech. **35**, 117 (1969).
3. H. K. Moffat, Philos. Trans. R. Soc. London, Ser. A **333**, 321 (1990).
4. H. K. Moffat, *Magnetic Field Generation in Electrically Conducting Fluids* (Cambridge Univ. Press, Cambridge, 1978; Mir, Moscow, 1980).

5. O. G. Chkhetiani and S. S. Moiseev, Pis'ma Zh. Éksp. Teor. Fiz. **70**, 268 (1999) [JETP Lett. **70**, 277 (1999)].
6. R. Cieszelski, Izv. Ross. Akad. Nauk, Fiz. Atmos. Okeana **35**, 174 (1999) [Izv., Atmos. Ocean. Phys. **35**, 157 (1999)].
7. M. V. Kurganskiĭ, Dokl. Akad. Nauk **371**, 240 (2000).
8. I. S. Gromeka, *Some Cases of Incompressible Liquid Motion* (Akad. Nauk SSSR, Kazan, 1952), p. 116 [in Russian].
9. J. Serrin, *Mathematical Principles of Classical Fluid Mechanics* (Springer, Berlin, 1959; Nauka, Moscow, 2001), Handbuch der Physik, Vol. 81, pp. 125–368.
10. R. B. Pelz, V. Yakhot, S. A. Orszag, and E. Levich, Phys. Rev. Lett. **54**, 2505 (1985).
11. A. S. Chefranov and S. G. Chefranov, Dokl. Akad. Nauk **393**, 624 (2003) [Dokl. Phys. **48**, 696 (2003)].
12. V. I. Arnol'd, *Mathematical Methods of Classical Mechanics* (Nauka, Moscow, 1974; Springer, New York, 1978).
13. S. W. Jones, Chaos, Solitons and Fractals **4**, 929 (1994).
14. E. P. Anisimova, Yu. N. Belov, A. A. Speranskaya, and V. S. Shandin, Izv. Akad. Nauk SSSR, Fiz. Atmos. Okeana **17**, 768 (1981).
15. E. B. Gledzer, F. V. Dolzhanskiĭ, and A. M. Obukhov, *Systems of a Hydrodynamic Type and Their Application* (Nauka, Moscow, 1981) [in Russian].
16. S. G. Chefranov, Izv. Ross. Akad. Nauk, Fiz. Atmos. Okeana **39**, 760 (2003) [Izv., Atmos. Ocean. Phys. **39**, 685 (2003)].
17. J. C. Andre and M. Lesieur, J. Fluid Mech. **81**, 187 (1977).
18. L. D. Landau and E. M. Lifshitz, *Course of Theoretical Physics*, Vol. 6: *Fluid Mechanics*, 3rd ed. (Nauka, Moscow, 1986; Pergamon, New York, 1987).
19. G. J. Haltiner and F. L. Martin, *Dynamic and Physical Meteorology* (McGraw-Hill, New York, 1957; Nauka, Moscow, 1960).
20. H. P. Greenspan, *The Theory of Rotating Fluids* (Cambridge Univ. Press, New York, 1968; Gidrometizdat, Leningrad, 1975).
21. G. K. Batchelor, *An Introduction to Fluid Dynamics* (Cambridge Univ. Press, Cambridge, 1967; Mir, Moscow, 1973).
22. A. B. Borisov, Dokl. Akad. Nauk **379**, 319 (2001) [Dokl. Phys. **46**, 473 (2001)].
23. A. B. Borisov, Pis'ma Zh. Éksp. Teor. Fiz. **76**, 95 (2002) [JETP Lett. **76**, 84 (2002)].
24. S. S. Moiseev, R. Z. Sagdeev, A. V. Tur, and V. V. Yanovskiĭ, Zh. Éksp. Teor. Fiz. **83**, 215 (1982) [Sov. Phys. JETP **56**, 117 (1982)].
25. S. G. Chefranov, Zh. Éksp. Teor. Fiz. **95**, 547 (1989) [Sov. Phys. JETP **68**, 307 (1989)].
26. N. O. Renno and A. P. Ingersoll, J. Atmos. Sci. **53**, 572 (1996).
27. A. G. Yarmitskiĭ, Izv. Ross. Akad. Nauk, Mekh. Zhidk. Gaza, No. 2, 90 (2002).
28. P. J. Bren, C. A. Kruelle, and I. Rehberg, Europhys. Lett. **62**, 491 (2003).
29. S. G. Chefranov, Pis'ma Zh. Éksp. Teor. Fiz. **73**, 312 (2001) [JETP Lett. **73**, 274 (2001)].
30. S. L. Shalimov, Fiz. Plazmy **29**, 227 (2003) [Plasma Phys. Rep. **29**, 203 (2003)].
31. B. M. Boubnov and P. B. Rhines, in *Turbulent Mixing in Geophysical Flows*, Ed. by P. F. Linden and J. M. Redond (CIMNE, Barcelona, 2001).
32. C. L. Pekeris, Y. Accad, and B. Shkoller, Philos. Trans. R. Soc. London, Ser. A **275**, 425 (1973).
33. F. H. Busse, Annu. Rev. Fluid Mech. **10**, 435 (1978).

Translated by A. Betev

Journal of Experimental and Theoretical Physics, Vol. 99, No. 5, 2004, pp. 998–1004.
Translated from Zhurnal Éksperimental'noĭ i Teoreticheskoĭ Fiziki, Vol. 126, No. 5, 2004, pp. 1146–1152.

FLUIDS

Nonparametric Scaled Equation of State and Approximation of *P*–ρ–*T* Data Near the Vapor–Liquid Critical Point

P. P. Bezverkhiĭ, V. G. Martynets, and É. V. Matizen*

Nikolaev Institute of Inorganic Chemistry, Siberian Division, Russian Academy of Sciences, Novosibirsk, 630090 Russia

**e-mail: matizen@casper.che.nsk.su*

Received December 9, 2003

Abstract—We present a new scaled equation of state adequately describing the *P*–ρ–*T* data for fluids near the vapor–liquid transition point and compare this equation to the Schofield parametric equation of state. A comparative analysis of the approximation of data for He^4, C_2H_4, and H_2O in the critical region shows that both these equations correctly describe the behavior of thermodynamic functions asymptotically close to the critical point. © *2004 MAIK "Nauka/Interperiodica".*

1. INTRODUCTION

The modern scaling theory of critical-point phase transitions in liquids, or the theory of second-order phase transitions, had its beginning in the mid-1960s, when a singular character of the behavior of thermodynamic functions at the critical point was established based on the results of numerous thorough experiments. In order to explain the behavior of substances near such points, the so-called scaling hypothesis was suggested (see, e.g., [1–3]) according to which the thermodynamic potentials are generalized functions of a noninteger order with respect to the corresponding fields. Subsequently, Wilson calculated the powers of these functions (critical exponents) by solving the renormalization-group equations [4]. The approximation of experimental data in terms of these functions showed that certain corrections have to be introduced in order to reduce errors in the nonasymptotic region. The most commonly accepted were the corrections of Wegner [5] and Berestov [6], which accounted for the nonasymptotic behavior, and the Pokrovskii transformation [7], which described the asymptotic asymmetry of real fluids relative to the critical isochore.

There were many attempts to expand this description beyond the region of validity of the scaling hypothesis by matching the scaled equation to the Landau expansion [8]. Recently, this approach to the crossover was studied in [9, 10]. Alternative descriptions of the behavior of substances in the critical region were proposed as well. In particular, Martynov [11] formulated a general approach to construction of the theory of critical phenomena based on a systematic analysis of the Ornstein–Zernicke equations and the main relationships of the local statistical mechanics. However, an expression obtained in [11] for the pressure does not adequately describe the *P*–ρ–*T* data for fluids near the vapor–liquid transition point.

Up to the present, the most adequate description of the behavior of fluids in the critical region was provided by the scaling theory. In the general form, the scaled equation of state for symmetric systems such as the Ising model was proposed by Griffiths [12]:

$$h_1 = \mathrm{sgn}(A_1)|A_1|^{\delta} f(\tau/|A_1|^{1/\beta}). \qquad (1)$$

Here, h_1 is the scaling field, A_1 is the conjugated density, τ is the reduced temperature, δ and β are the critical exponents, and f is a scaling function. However, the function f entering into Eq. (1) is set only by asymptotic behavior at certain lines. The form of this function was calculated [13, 14] by method of ε-expansion in the renormalization-group theory. This form is complicated and not suited for the treatment of experimental data. For this reason, a rather large number of interpolated scaling functions were suggested. Unfortunately, these attempts did not provide a convenient equation adequately describing the critical points of fluids [15].

One of the most successful attempts of such interpolation was the parametric equation of state proposed by Schofield [16], which allowed the asymptotic behavior of a fluid to be described in the entire critical region, rather than only on the selected lines. Modifications of the Schofield equation, taking into account corrections for the nonasymptotic behavior and the asymmetry of real fluids at the critical point, significantly expanded the range of application of the scaling theory [17]. However, these parametric equations are still rather complicated and the description of experimental data in these terms is difficult. Therefore, the problem of

selecting a sufficiently simple scaling function, such that Eq. (1) could be conveniently used for the description of the behavior of a fluid in a large vicinity of the critical point, is still open.

We have used the form of the function proposed in [11] for the pressure and expressed the scaling field h_1 as

$$h_1 = \mathrm{sgn}(A_1)|A_1|^{\delta}(m + k(h_2/|A_1|^{1/\beta})^{\gamma}). \quad (2)$$

The equation of state in this form provides correct asymptotic behavior on selected lines and describes the entire asymptotic region of a symmetric system (Ising model) near the critical point. However, the simple Ising model and the isomorphous model of the lattice gas possess a special symmetry with respect to the sign of the scaling field h_1, whereas real fluids do not possess this symmetry. The passage to description of the critical point of a real fluid is provided by the Pokrovskii transformation [7].

Using the Pokrovskii transformation (in terms of the algebra of fluctuating quantities), the pressure and the chemical potential of a fluid can be represented in the form of a linear combination of quantities corresponding to the symmetric system and possessing certain scaling dimensions. This representation leads to the so-called "mixing" of the thermodynamic variables: the temperature and chemical potential of a real system can be represented as linear combinations of the temperature and chemical potential of the symmetric model system; by the same token, the density and entropy of a real system can be represented as linear combinations of the same quantities for the model system.

Restricting the consideration to the most strongly fluctuating values, we write

$$\begin{aligned} \Delta\tilde{\rho} &= A_1 + bA_2, \quad \sigma = A_2 + aA_1, \\ h_1 &= \eta + a\tau, \quad h_2 = \tau + b\eta, \end{aligned} \quad (3)$$

where h_1 and h_2 are the generalized fields corresponding to the symmetric model system; A_1 and A_2 are the conjugated generalized densities; $\Delta\tilde{\rho} = (\rho - \rho_k)/\rho_k$; $\eta = (\mu - \mu_k)(\rho_k/P_k)$; $\sigma = (s - s_k)(T_k/P_k)$; $\tau = (T - T_k)/T_k$; μ is the chemical potential; s is the entropy per unit volume; a and b are the fitting constants characterizing the degree of "mixing"; and the subscript k refers to critical values of the variables. The differential of the pressure can be written as [8]

$$dP = \rho d\mu + sdT, \quad (4)$$

$$d\pi = (1 + \Delta\tilde{\rho})d\eta + \left(\sigma + \frac{s_k T_k}{P_k}\right)d\tau. \quad (5)$$

In terms of the symmetric model, this can be rewritten as

$$d\pi = \left(\frac{1 - bM}{1 - ab} + A_1\right)dh_1 + \left(\frac{M - a}{1 - ab} + A_2\right)dh_2, \quad (6)$$

where $M \equiv s_k T_k/P_k$, $\pi = (P - P_k)/P_k$.

To the first approximation, we may set $b = 0$ in Eq. (6) because this term is responsible for singularity of the diameter of the boundary curve [7] and gives only an insignificant correction to the pressure. Then, the pressure differential is

$$d\pi = (1 + A_1)dh_1 + (M - a + A_2)dh_2. \quad (7)$$

Now let us use the above relations and transform the scaled equation of state (2). In order to determine the pressure as a function of the density and temperature near the vapor–liquid critical point, let us differentiate Eq. (2) with respect to h_1 and h_2. This yields

$$m\delta A_1^{\delta-1}\left(\frac{\partial A_1}{\partial h_1}\right)_{h_2} + kh_2^{\gamma}\left(\frac{\partial A_1}{\partial h_1}\right)_{h_2} = 1, \quad (8)$$

$$m\delta A_1^{\delta-1}\left(\frac{\partial A_1}{\partial h_2}\right)_{h_1} + kA_1\gamma h_2^{\gamma-1} + kh_2^{\gamma}\left(\frac{\partial A_1}{\partial h_2}\right)_{h_1} = 0. \quad (9)$$

Using these relations, we can determine $(\partial A_1/\partial h_2)_{h_1}$ and $(\partial A_1/\partial h_1)_{h_2}$. By definition, a differential of the thermodynamic potential of a symmetric system is the total differential:

$$d\Phi = A_1 dh_1 + A_2 dh_2. \quad (10)$$

Using this property of the total differential and Eq. (9), we obtain

$$\left(\frac{\partial A_2}{\partial h_1}\right)_{h_2} = -k\gamma h_2^{\gamma-1}A_1\left(\frac{\partial A_1}{\partial h_1}\right)_{h_2}. \quad (11)$$

Integrating Eq. (11) yields

$$A_2 = -\frac{k\gamma h_2^{\gamma-1}}{2}A_1^2 + \varphi(h_2), \quad (12)$$

where $\varphi(h_2)$ is an unknown function of the scaling field h_2. Taking into account expressions (3) and the condition $b = 0$, we eventually obtain

$$A_2 = -\frac{k\gamma\tau^{\gamma-1}}{2}\Delta\tilde{\rho}^2 + \varphi(\tau). \quad (13)$$

Substituting the values of A_1, A_2, h_1, and h_2 into Eq. (7), we obtain an expression for the total differential of the pressure:

$$d\pi = (1+\Delta\tilde{\rho})(m\delta\Delta\tilde{\rho}^{\delta-1} + k\tau^{\gamma})d\Delta\tilde{\rho} + \left[(1+\Delta\tilde{\rho})k\gamma\Delta\tilde{\rho}\tau^{\gamma-1} + M - a - \frac{k\gamma\tau^{\gamma-1}\Delta\tilde{\rho}^2}{2} + \varphi(\tau)\right]d\tau. \quad (14)$$

Integrating the first term in this equation with respect to $\Delta\tilde{\rho}$ from 0 to $\Delta\tilde{\rho}$ at τ = const, and the second term with respect to τ from 0 to τ at $\Delta\tilde{\rho} = 0$, we obtain an expression for the pressure:

$$\pi = m\Delta\tilde{\rho}^{\delta}\left(1+\frac{\delta}{1+\delta}\Delta\tilde{\rho}\right) + k\tau^{\gamma}\left(\Delta\tilde{\rho} + \frac{1}{2}\Delta\tilde{\rho}^2\right) + (M-a)\tau + \int\varphi(\tau)d\tau. \quad (15)$$

Taking into account that, on the critical isochore at $\tau \longrightarrow 0$, the heat capacity C_v behaves as $\tau^{-\alpha}$, we may conclude that the integral in expression (15) has the form of $C_0\tau^{2-\alpha}$. Thus, the final scaled equation of state has the following form:

$$\pi = (M-a)\tau + m\Delta\tilde{\rho}|\Delta\tilde{\rho}|^{\delta-1}\left(1+\frac{\delta}{1+\delta}\Delta\tilde{\rho}\right) + k\tau|\tau|^{\gamma-1}\left(\Delta\tilde{\rho} + \frac{1}{2}\Delta\tilde{\rho}^2\right) + C_0\tau|\tau|^{1-\alpha}. \quad (16)$$

Using Eqs. (3) and (13), we obtain the following expressions for the entropy and heart capacity:

$$\sigma = -\frac{k\gamma\tau^{\gamma-1}}{2}\Delta\tilde{\rho}^2 + s_0\tau^{1-\alpha} + a\Delta\tilde{\rho}, \quad (17)$$

$$C_v = \frac{TP_k}{T_k^2\rho}\left(\frac{\partial\sigma}{\partial\tau}\right)_{\rho} = \frac{P_kT}{T_k^2\rho} \times\left[-\frac{k\gamma(\gamma-1)}{2}\tau^{\gamma-2}\Delta\tilde{\rho}^2 + s_0(1-\alpha)\tau^{-\alpha}\right]. \quad (18)$$

However, experiment shows that the pressure on the critical isochore is, to within the experimental error, directly proportional to τ. Therefore, in approximating experimental data on the pressure, we may neglect the integral in Eq. (15) as well as the last term in Eq. (16). Of course, these terms cannot be ignored in determining derivatives of the pressure. It should be emphasized that Eqs. (16)–(18) describe the behavior of real fluids very close to the vapor–liquid critical point without corrections for the nonasymptotic behavior.

The isotherms of the equation of state obtained, being continuous in the phase separation region, are similar to the van der Waals isotherms. Then, using conditions on the spinodal, we can determine a curve bounding the region of absolute instability of the liquid:

$$\Delta\tilde{\rho} = \pm\left(\frac{k}{m\delta}\right)^{1/(\delta-1)}|\tau|^{\beta}. \quad (19)$$

As was noted above, the terms proportional to b in Eq. (3) reflect the "singularity of the linear diameter." Indeed, taking these terms into account, assuming that $h_1 = 0$ (a condition on the boundary curve), and using Eqs. (3) and (13), we obtain the following relations for the boundary curve:

$$\rho_L - \rho_G = 2\rho_k\left(\frac{k}{m}\right)^{1/(\delta-1)}|\tau|^{\beta}, \quad (20)$$

$$\rho_L + \rho_G = b\rho_k\left[2(2-\alpha)C_0 + k\gamma\left(\frac{k}{m}\right)^{2/(\gamma-1)}\right]|\tau|^{1-\alpha}. \quad (21)$$

These expressions show that the boundary curve behaves as $|\tau|^{\beta}$ and the diameter of this curve has a singularity of the type $|\tau|^{1-\alpha}$, in accordance with the theory [7]. This result also confirms the validity of assumptions underlying our equation of state.

Then, we compared the obtained equation of state to that of the existing scaling theory of critical phenomena. For the comparison, we used the scaled Schofield equation [16]. According to our experience, this equation most adequately reflects the thermodynamic properties of fluids. In order to obtain an expression for the pressure, we transformed the Schofield equation using the algebra of fluctuating quantities according to Pokrovskii [7]. In parametric form, expressions for the pressure, temperature, and density with allowance for the singularity of the diameter of the boundary curve are as follows [17, 18]:

$$\pi = (M-\alpha)\tau + Ar^{\gamma+\beta}(\theta-\theta^3) + Agr^{2-\alpha}(z_0 + z_2\theta^2 + z_4\theta^4),$$
$$\tau = (1-ab)^{-1}[r(1-B^2\theta^2) - bAr^{\gamma+\beta}(\theta-\theta^3)], \quad (22)$$
$$\Delta\rho = gr^{\beta}\theta + bAr^{1-\alpha}(\theta-\theta^3).$$

Here, A and g are empirical coefficients and B, z_0, z_2,

and z_4 are combinations of the critical exponents:

$$B^2 = (\gamma - 2\beta)/\gamma(1 - 2\beta),$$

$$z_0 = \gamma(\gamma - 1)/2B^2\alpha(1 - \alpha)(2 - \alpha),$$

$$z_2 = [B^2\alpha(2\gamma + 2\beta - 1) - \gamma + 2\beta]/2B^2\alpha(1 - \alpha),$$

$$z_4 = (\gamma + \beta - 3/2)/\alpha.$$

Ignoring, by analogy with Eq. (7), the terms involving b (accounting for the singularity of the boundary curve diameter and leading to a pressure correction insignificant in the context of our analysis), we obtain the Schofield equation of state for the description of real fluids:

$$\begin{aligned} \pi &= (M - a)\tau + Ar^{\gamma+\beta}(\theta - \theta^3) \\ &+ Agr^{2-\alpha}(z_0 + z_2\theta^2 + z_4\theta^4), \\ \tau &= r(1 - B^2\theta^2), \\ \Delta\rho &= gr^\beta\theta. \end{aligned} \tag{23}$$

This equation was used for comparing our nonparametric equation (16) to the relations following from the scaling theory. The comparison was based on the analysis of data for three fluids: helium-4 (He^4), water (H_2O), and ethylene (C_2H_4).

It should be emphasized that we do not introduce corrections for the nonasymptotic behavior, and our aim is only to establish the principal applicability of the scaled equation of state (16) to description of the behavior of real fluids in the critical region.

2. EXPERIMENTAL DATA

In investigations of the thermodynamic properties of fluids near the vapor–liquid transition point, there are three main factors complicating obtaining of the results which can be used for verification of the adequacy of proposed theories. The first factor is the presence of impurities that very strongly influences the critical behavior of fluids. The second factor is the effect of gravity, which accounts for the development of significant density gradients in the vertical direction. The third factor is a slow rate of attaining the state of thermodynamic equilibrium. Taking this into account, we verified Eq. (16) using our experimental results [19], which are still the most exact data obtained under the conditions of minimization of the influence of all three factors limiting the experimental accuracy.

In comparing various equations of state, there arises the problem of selecting critical parameters and exponents for the particular fluids. These quantities are determined as fitting parameters in approximation of the experimental data by one or another function and, hence, depend on the form of this function. This problem was recently discussed in [20]. In addition, the critical parameters of fluids depend on the experimental methods employed. We believe that these differences, while influencing the error of approximation, still allow the aforementioned data to be used for the comparison of the new equation of state to the existing ones.

The scatter of the critical parameters, depending on the form of the approximating function and the selected set of experimental data, may significantly exceed the experimental uncertainty. For example, the values of parameters of He^4 determined from three equations of state (with all parameters, including critical exponents, fitted over the whole body of data) are as follows.

For the equation of state (22): $T_k = 5.2028$ K; $P_k = 1711.8$ Torr (228221 Pa); $\rho_k = 69.56$ kg/m^3; $\Delta f = 0.869$ Torr (116 Pa).

For the equation of state (23): $T_k = 5.2042$ K; $P_k = 1713.9$ Torr (228501 Pa); $\rho_k = 69.75$ kg/m^3; $\Delta f = 0.902$ Torr (116 Pa).

For the Sartakov–Martynets equation of state [17, 18]), which takes into account both the asymmetry of real fluids and their nonasymptotic behavior far from the critical point [21]: $T_k = 5.1968 \pm 0.005$ K; $P_k = 1704.1 \pm 6$ Torr (227195 Pa); $\rho_k = 69.56 \pm 0.14$ kg/m^3; $\Delta f = 0.832$ Torr (111 Pa).

Here and below, Δf is the rms error of approximation; the error intervals of the values of critical parameters for the Sartakov–Martynets equation are given for a confidence level of 0.99. As can be seen, the error of approximation for the Sartakov–Martynets equation in the entire interval of densities is smaller than that for the two other equations, which is quite natural since the former equation involves the terms responsible for the asymmetry and nonasymptotic behavior. Note that the rms error of approximation rather weakly depends on the form of equations, while the critical parameters are quite sensitive to the form of functions. In this study, all approximations of the data for He^4 were obtained using the values of critical parameters corresponding to the Sartakov–Martynets equation. For He^4, the P–ρ–T data were obtained for 23 isotherms (620 points) at temperatures above and below the critical point [19]. Since the scaled equation (16) and the Schofield equation are more readily compared above the critical temperature, the approximation was obtained using the data for $T > T_k$ (340 experimental points) taken from [19].

For H_2O, we used the data of Ryvkin and Akhundov [22–24] and set the following critical parameters: $T_k = 647.3$ K; $P_k = 22.856$ MPa; $\rho_k = 307$ kg/m^3 [22–25].

For C_2H_4, we used the data of Hastings *et al.* [26] with the critical parameters $T_k = 282.3452$ K, $P_k = 5.0403$ MPa, and $\rho_k = 213.752$ kg/m^3 (7.634 mol/l).

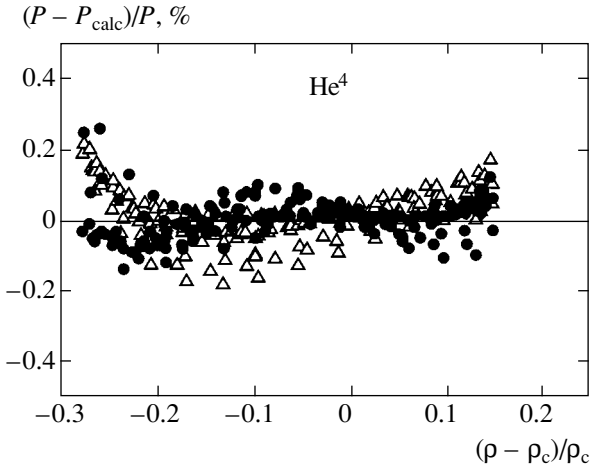


Fig. 1. A plot of the relative deviation of the experimentally measured pressure from the values calculated for He^4 at $T > T_k$ in terms of (●) Eq. (16) and (△) the Schofield equation (23).

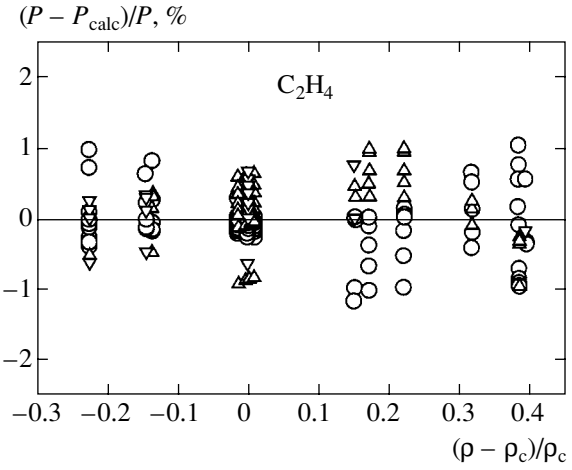


Fig. 2. A plot of the relative deviation of the experimentally measured pressure [26] from the values calculated for ethylene (C_2H_4) in terms of (o) Eq. (16) and (△, ▽) the Schofield equation (23). The interval of approximation includes all data for the single-phase region, including the results obtained in the vicinity of the boundary curve.

The rms errors of parameters of the equations of state (including critical exponents) are strongly correlated. For correct comparison of the equations of state, the critical exponents were taken equal to the values obtained within the framework of the three-dimension al Ising model [10]: $\beta = 0.3255$, $\gamma = 1.239$, and $\delta = 4.80$.

3. RESULTS OF APPROXIMATION

The comparative approximation procedure consisted in determining the minimum of the quadratic error functional (representing the sum of square deviations of the calculated values from experimental data) depending on the fitting parameters. The fitting parameters were calculated using the conventional least-squares method (for linear systems of normal equations) and the method of configurations (for nonlinear equations). The results of approximation of the *P*–ρ–*T* data for helium-4, water, and ethylene using proposed equation of state (16) and Schofield equation (23) are presented in the table (fitting parameters of the two equations, rms errors, and the interval of approximation on the density scale) and in Figs. 1–3 (relative deviations of the pressure from calculated values).

Helium-4 (He^4). Approximation of the data for He^4 by Eq. (23) was obtained in the maximum range of densities ($41.2\ kg/m^3 \le \rho \le 96.9\ kg/m^3$) where such experimental data were available. The rms deviation for Eq. (23) is $\Delta f = 6.13$ Torr (817 Pa) or 0.36%. In the same range of densities, Eq. (16) yields $\Delta f = 3.34$ Torr (445 Pa) or 0.2%. and, hence, is definitely preferred. Figure 1 and the table present the results for a narrower interval of densities. As can be seen, the approximation error and the relative scatter of data for Eqs. (16) and (23) in this narrower interval become comparable.

As was noted above, the term $C_0\tau^{2-\alpha}$ in Eq. (16) is insignificant in the approximation under consideration. Indeed, approximation of the experimental data using

Comparison of the results of approximation of the *P*–ρ–*T* data for helium-4, water, and ethylene using the proposed equation of state (16) and the Schofield equation (23)

Equation (16)				Schofield equation (23)			
	^4He	C_2H_4	H_2O		^4He	C_2H_4	H_2O
m	8.243	2.753	2.967	*A*	5.836	13.638	13.041
k	9.080	18.915	14.952	*g*	0.827	1.079	1.057
M–a	3.922	6.529	7.489	*M–a*	3.880	5.794	7.085
C_0	1.498	0.430	2.669	–	–	–	–
Δf	0.06%	0.48%	0.34%	Δf	0.08%	0.52%	0.65%
	$-0.28 \le \Delta\tilde{\rho} \le 0.15$	$-0.23 \le \Delta\tilde{\rho} \le 0.39$	$\lvert\Delta\tilde{\rho}\rvert \le 0.3$		$-0.28 \le \Delta\tilde{\rho} \le 0.15$	$-0.23 \le \Delta\tilde{\rho} \le 0.39$	$\lvert\Delta\tilde{\rho}\rvert \le 0.3$

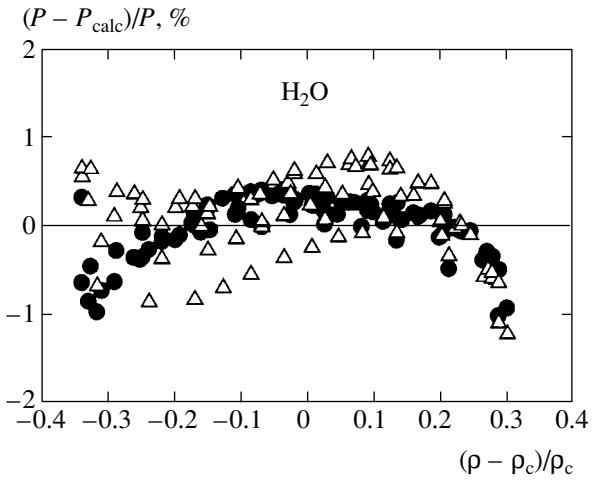


Fig. 3. A plot of the relative deviation of the experimentally measured pressure [22–24] from the values calculated for water at $T > T_k$ in terms of (●) Eq. (16) and (△) Schofield equation (23).

Eq. (16) with this term rejected showed that the rms error was the same and the fitting coefficients remained almost unchanged. The coefficient at this term is yet given in the table for the sake of complete data presentation.

Ethylene (C_2H_4). Figure 2 shows the relative deviations of the pressure from calculated values for C_2H_4 [26]. In this system, the rms error of data approximation in terms of Eq. (16) is somewhat greater than that for the Schofield equation.

Water (H_2O). Figure 3 shows the results of approximation of the experimental data [22–24] for water at $T > T_k$ in terms of Eqs. (16) and (23). In this case, the proposed equation is again somewhat more preferred than the Schofield model.

An analysis of the results of approximation showed that Eq. (16) in a narrow interval of densities ($0.8 \le \rho/\rho_k < 1.2$) describes the experimental data even better than the modified scaled Schofield equation. The average scatter of data in this interval is comparable with the experimental uncertainty. As can be seen from Figs. 1–3, the scatter of data for both equations grows with increasing distance from the critical point. This behavior is quite natural because the scaled equation (as well as the scaling theory as such) is asymptotic. In order to improve the description of experimental data, it is necessary to introduce nonasymptotic corrections and provide for matching of the scaled equation to an equation describing the regular part of thermodynamic functions, for example, to the virial equation of state. In recent years, there have been attempts (see [9, 10]) to combine the scaled equation with the Landau expansion for second-order phase transitions [8]. The approximation using such combined equations naturally provides for a decrease in error, since additional fitting parameters are introduced. For example, the scatter of data for water fell to within 0.08% and for ethylene, to within 0.03%. In this study, neither the Landau expansion nor any other corrections were used because the aim was to establish the principal applicability of the proposed scaled equation of state to describe the asymptotic behavior of real fluids in the critical region.

4. CONCLUSIONS

We have presented a new scaled equation of state obtained using the scaling field function for the pressure suggested by Martynov [11]. It was shown that the proposed equation correctly describes the asymptotic behavior of fluids near the vapor–liquid critical point. In the asymptotic vicinity of the critical point, the new equation approximates the experimental dependences of the pressure on the temperature and density with greater precision than the modified parametric Schofield equation. At the same time, proposed equation (16) is simpler than parametric Schofield equation (23), more convenient for application in practice, and gives simple expressions for the binodal and spinodal curves.

REFERENCES

1. A. Z. Patashinskiĭ and V. L. Pokrovskiĭ, Zh. Éksp. Teor. Fiz. **50**, 439 (1966) [Sov. Phys. JETP **23**, 292 (1966)].
2. L. P. Kadanoff, Physics (Long Island City, N.Y.) **2**, 263 (1966).
3. F. Dyson, E. Montroll, M. Katz, and M. Fischer, in *Stability and Phase Transitions* (Mir, Moscow, 1973) [in Russian].
4. K. G. Wilson and J. Kogut, Phys. Rep. **12C**, 75 (1975).
5. F. J. Wegner, Phys. Rev. B **5**, 4529 (1972).
6. A. T. Berestov, Zh. Éksp. Teor. Fiz. **72**, 348 (1977) [Sov. Phys. JETP **45**, 184 (1977)].
7. A. Z. Patashinskiĭ and V. L. Pokrovskiĭ, *Fluctuation Theory of Phase Transitions*, 2nd ed. (Nauka, Moscow, 1982; Pergamon Press, Oxford, 1979).
8. L. D. Landau and E. M. Lifshitz, *Statistical Physics*, 3rd ed. (Nauka, Moscow, 1976; Addison-Wesley, Reading, MA, 1970).
9. Z. Y. Chen, P. C. Albright, and J. V. Sengers, Phys. Rev. A **41**, 3161 (1990).
10. V. A. Agayan, M. A. Anisimov, and J. V. Sengers, Phys. Rev. E **64**, 026125 (2001).
11. G. A. Martynov, Dokl. Akad. Nauk **378**, 173 (2001) [Dokl. Phys. **46**, 310 (2001)].
12. R. B. Griffiths, Phys. Rev. **158**, 176 (1967).
13. G. M. Avdeeva and A. A. Migdal, Pis'ma Zh. Tekh. Fiz. **16**, 253 (1972) [Sov. Tech. Phys. Lett. **16**, 178 (1972)].
14. E. Brezin, D. J. Wallace, and K. G. Wilson, Phys. Rev. B **7**, 232 (1973).
15. M. Vicentini-Missoni, J. M. H. Levelt-Sengers, and M. S. Green, Phys. Rev. Lett. **22**, 389 (1969); M. Barmatz, P. C. Hohenberg, and A. Kornblit, Phys. Rev. B **12**, 1947 (1975); O. B. Verbeke *et al.*, J. Phys. Chem. **73**,

4076 (1969); R. D. Goodwin, J. Res. Natl. Bur. Stand., Sect. A **79**, 84 (1974).

16. P. Schofield, Phys. Rev. Lett. **22**, 606 (1969); P. Schofield, G. D. Litster, and G. T. Ho, Phys. Rev. Lett. **23**, 1098 (1969).
17. A. G. Sartakov and V. G. Martynets, Izv. Sib. Otd. Akad. Nauk SSSR, Ser. Khim. Nauk, No. 7 (3), 14 (1982).
18. P. P. Bezverkhiĭ, V. G. Martynets, É. V. Matizen, and V. F. Kukarin, Teplofiz. Vys. Temp. **26**, 700 (1988).
19. V. F. Kukarin, V. G. Martynets, É. V. Matizen, and A. G. Sartakov, Fiz. Nizk. Temp. **6**, 549 (1980) [Sov. J. Low Temp. Phys. **6**, 263 (1980)].
20. V. A. Rabinovich and Yu. E. Sheludyak, Zh. Fiz. Khim. **77**, 1708 (2003).
21. V. F. Kukarin, V. G. Martynets, É. V. Matizen, and A. G. Sartakov, Fiz. Nizk. Temp. **7**, 1501 (1981) [Sov. J. Low Temp. Phys. **7**, 725 (1981)].
22. S. L. Rivkin and T. S. Akhundov, Teploénergetika (Moscow), No. 10, 66 (1963).
23. S. L. Rivkin, T. S. Akhundov, E. A. Kremenevskaya, and N. N. Asadullaeva, Teploénergetika (Moscow), No. 4, 59 (1966).
24. S. L. Rivkin and T. S. Akhundov, Teploénergetika (Moscow), No. 1, 57 (1962).
25. N. S. Osborn, H. F. Stimson, and D. C. Ginnings, J. Res. Natl. Bur. Stand. **18**, 389 (1937); **23**, 261 (1939).
26. J. R. Hastings, J. M. H. Levelt Sengers, and F. W. Balfour, J. Chem. Thermodyn. **12**, 1009 (1980).

Translated by P. Pozdeev